



Munich Personal RePEc Archive

**Classifying Hedge Funds using k-means  
Clustering of Self-Organizing Maps: a  
return-based analysis of misclassification  
and the problem of style creep**

Deetz, Marcus and Poddig, Thorsten and Varmaz, Armin

Universität Bremen

25 August 2009

Online at <https://mpra.ub.uni-muenchen.de/16939/>  
MPRA Paper No. 16939, posted 26 Aug 2009 06:20 UTC

# Klassifizierung von Hedge-Fonds durch das k-means Clustering von Self-Organizing Maps: eine renditebasierte Analyse zur Selbsteinstufungsgüte und Stiländerungsproblematik \*

Deetz, Marcus<sup>†</sup>    Poddig, Thorsten<sup>‡</sup>    Varmaz, Armin<sup>§</sup>

**Zusammenfassung:** Mit der Implementierung des 2-stufigen Ansatzes nach Vesanto & Alhoniemi (2000) erweitert der vorliegende Artikel das in der Hedge-Fonds Literatur zur Klassifikation mit Self-Organizing Maps üblicherweise gewählte Vorgehen der visuellen Auswertung der Kohonen-Karten und stellt damit ein automatisiertes Verfahren vor, welches eine konsistente Zusammenfassung benachbarter Output-Units und damit eine objektive Klassifizierung garantiert. Die empirische Anwendung dieses Verfahrens resultiert in einer Reduktion der durch die Datenbank vorgegebenen Strategiegruppen. Diese Taxonomie ist robust in Hinblick auf unterschiedliche Zeiträume, Topologien und Datensample. Weiterführende Untersuchungen zur Konsistenz der Eigenangaben der Hedge-Fonds-Manager zeigen ein erhebliches Maß an Fehlklassifikation auf. Darüber hinaus können nicht bekannt gegebene Stiländerungen im Untersuchungszeitraum festgestellt werden. Dabei ist die ex post Wahrscheinlichkeit eines „Style Creep“ von Hedge-Fonds mit falscher im Vergleich zu Fonds mit richtiger Selbsteinstufung deutlich höher.

**Keywords:** Self-Organizing Maps, Clustering, Klassifizierung, Hedge-Fonds, Style Creep

**J.E.L.:** C45, C81, G11, G23

---

\* Neben eigenen Routinen werden in der vorliegenden Untersuchung die frei verfügbaren Algorithmen der SOM-Toolbox 2.0 (<http://www.cis.hut.fi/projects/somtoolbox/>) verwendet.

<sup>†</sup> Marcus Deetz ist Doktorand am Lehrstuhl für ABWL, insbes. Finanzwirtschaft der Universität Bremen. Corresponding Author: Email: mdeetz@uni-bremen.de

<sup>‡</sup> Prof. Dr. Thorsten Poddig ist Inhaber des Lehrstuhls für ABWL, insbes. Finanzwirtschaft der Universität Bremen.

<sup>§</sup> Dr. Armin Varmaz ist Habilitand am Lehrstuhl für ABWL, insbes. Finanzwirtschaft der Universität Bremen.

# 1 Einführung

Insbesondere in den letzten beiden Dekaden haben sich Hedge-Fonds rasant entwickelt und als Anlageinstrument auf den internationalen Finanzmärkten etabliert. So hat sich seit Beginn dieses Jahrtausends, nach Angaben vom BarclayHedge, die Anzahl der Hedge-Fonds, von anfänglich 6.000 auf über 13.300 im Januar 2007, mehr als verdoppelt.<sup>1</sup> Im selben Zeitraum versechsfachte sich das verwaltete Volumen von 323 auf 1.860 Mrd. USD. Dieser Erfolg wird oft mit der Tatsache begründet, dass diese neue Assetklasse weniger restriktiven gesetzlichen Rahmenbedingungen und Regulierungen unterliegt als traditionelle Investmentfonds. Damit steht Hedge-Fonds ein viel breiteres Spektrum an Finanzanlageinstrumenten und Strategiemöglichkeiten zur Verfügung, das sogar Gewinnmöglichkeiten in fallenden Marktphasen gestattet.<sup>2</sup> Neben diesem attraktiven Renditegenerierungspotenzial sehen Investoren den Nutzen dieser Assetklasse insbesondere in der Verbesserung der Rendite-Risikostruktur ihrer aus traditionellen Anlagen bestehenden Portfolios.<sup>3</sup> Die Freiheiten in der Kapitalanlage führen zu einer sehr großen Vielfalt an verschiedenen Hedge-Fonds-Strategien und -Stilen und machen Hedge-Fonds zu einer äußerst heterogenen Assetklasse, der der Makel der Intransparenz anhaftet.<sup>4</sup> Kenntnisse über den Anlagestil eines Fonds sind jedoch aus unterschiedlichen Gründen sehr wichtig. Baghai-Wadji et al. (2006) nennen in diesen Zusammenhang insbesondere die Verbesserung der Investitionsentscheidung, die Vermeidung von unangemessener Diversifikation bei der Dachfondkonstruktion und die Nützlichkeit bei der Entwicklung von Performancemessungs- sowie Risikomanagementmodellen. Vor diesem Hintergrund kann ein konsistentes Klassifizierungssystem notwendige Informationen bereitstellen.

Im Gegensatz zu traditionellen Investmentfonds, die nach den Assetklassen unterteilt werden, in die sie vorwiegend investiert sind, existiert für Hedge-Fonds keine einheitliche Einteilung zur Klassifizierung.<sup>5</sup> Zwar existieren viele Ansätze, um Hedge-Fonds nach ihrer Anlagestrategie zu klassifizieren. Diese sind jedoch durch die Datenbankanbieter geprägt und in der Regel weder vergleichbar noch in sich konsistent.<sup>6</sup> Empirische Untersuchungen verschiedener Hedge-Fonds-Datenbanken zeigen regelmäßig, dass ein und derselbe Hedge Fond in verschiedenen Datenbanken unterschiedlich geführt wird und die verschiedenen Datenbanken nur wenige Gemeinsamkeiten aufweisen.<sup>7</sup> Die

---

<sup>1</sup> Stand 02.06.08.

<sup>2</sup> Vgl. z.B. Seiler (2008), S. 13 und Schäfer (2004), S. 463.

<sup>3</sup> Vgl. Kaiser (2004), S. 190 f., Ineichen (2003), S. 140 f. und Signer (2005), S. 211 ff.

<sup>4</sup> Vgl. Schäfer (2004), S. 465; Bretzler & Rudolph (2004), S. 20.

<sup>5</sup> Vgl. z.B. Ohlms (2006), Bookstaber (2003) und Agarwal & Naik (2000).

<sup>6</sup> In ihren Untersuchungen stellen Indijc & Heen (2003) und Eling (2006) fest, dass sich sowohl in der Wissenschaft und der Praxis die Systematisierung nach CreditSuisse First Boston/Tremont durchzusetzen scheint.

<sup>7</sup> Siehe z.B. die Untersuchung von Das & Das (2005) oder Eling (2006).

Unterschiede äußern sich in den Definitionen der Anlagestile und in der Anzahl der Hedge-Fonds-Kategorien, in denen sich die Hedge-Fonds-Manager in der Regel selbst klassifizieren müssen. Dabei besteht das Problem, dass angesichts der Heterogenität der Hedge-Fonds-Strategien und -Stile auf der einen Seite und der Heterogenität der Datenbankanbieter auf der anderen Seite, es zu Fehleinstufungen kommen kann. Darüber hinaus besteht die Gefahr des so genannten „Style Creep“, wonach Fonds-Manager gezielt Fehleingaben tätigen bzw. Änderungen ihres Anlagestiles nicht anzeigen, um beispielsweise im Vergleich zur „Peer Group“ die eigene Performance zu schönen.<sup>8</sup> Die Untersuchung von Baghai-Wadji et al. (2006) zeigt, dass dieses Phänomen in der Hedge-Fonds-Branche verbreitet ist. Dabei stellen die Autoren insbesondere heraus, dass die Wahrscheinlichkeit einer Stiländerung von Hedge-Fonds mit falscher Selbsteinstufung im Vergleich zu Fonds mit richtiger Selbsteinstufung doppelt so hoch ist.<sup>9</sup>

Die Konsequenzen dieser gewollten oder ungewollten Klassifizierungsfehler sind weitreichend, da sie zu Verzerrungen in der Analyse von Hedge-Fonds-Strategien und deren Performance führen. Fernerbeinhalten sie das Gefahrenpotenzial, falsche Investitionsentscheidungen, beispielsweise zur Konstruktion von Dachfonds oder zur Beimischung eines bestehenden Portfolios, zu treffen. Das & Das (2005) stellen hierzu fest, dass verschiedene empirische Analysen unterschiedliche Ergebnisse aufgrund der gewählten Datenbank produzieren. Vor diesem Hintergrund besteht der Bedarf eines Verfahrens, das die Probleme in Zusammenhang mit den inkonsistenten und oft nicht vergleichbaren Strategieklassen der Datenbankanbieter umgeht und eine objektive Klassifikation ermöglicht. In verschiedenen empirischen Untersuchungen wurde daher der Versuch unternommen, basierend auf den beobachteten Renditen von Hedge-Fonds, spezifische Charakteristiken zu identifizieren, um Strategien zu extrahieren und das Hedge-Fonds Universum objektiv zu klassifizieren. In ihrer Studie zeigen Fung & Hsieh (1997) jedoch, dass das von Sharpe (1992) entwickelte Verfahren der Stilregression in der ursprünglichen Form für Hedge-Fonds nicht anwendbar ist und schlagen zur Lösung die Erweiterung dieses Ansatzes um sogenannte Asset-based Style (ABS-) Faktoren vor.<sup>10</sup> Derartige Faktoren versuchen die primäre Quelle des systematischen Risikos eines Hedge-Fonds, die zugrunde gelegte Strategie, nach der Kauf- und Verkaufsentscheidungen getroffen werden, abzubilden. In einer späteren Untersuchung zeigen Fung & Hsieh (2004) beispielsweise, dass die Rendite-Risiko-Charakteristik von Hedge-Fonds, die eine Long/Short-Equity Strategie verfol-

---

<sup>8</sup> Vgl. Brown & Goetzmann (1997), Baghai-Wadji et al. (2006).

<sup>9</sup> Vgl. Baghai-Wadji et al. (2006), S. 172.

<sup>10</sup> In ihrer Untersuchungen zeigen die Autoren, dass es aufgrund der Verfolgung dynamischer Handelsstrategien (das Eingehen von Long- und Short-Positionen) in einer Standardregression zu einer Durchschnittsbildung kommen kann.

gen, durch einen ABS-Faktor beschrieben werden kann, der sich aus einer Kaufoption in einem marktbreiten Aktienindex und einer Kaufoption in dem Spread „Small Caps minus Large Caps“ zusammensetzt.<sup>11</sup> Der hier kurz skizzierte ABS-Faktor zählt zu den sog. Spread-Faktoren. Daneben lassen sich zwei weitere Kategorien von ABS-Faktoren ausmachen, die Optionsstrategie- und die Indexfaktoren. Bekannte in der Literatur identifizierte ABS-Faktoren werden in Tabelle 1 aufgeführt.<sup>12</sup>

Strategie	Optionsstrategie-Faktoren	Index-Faktoren	Spread-Faktoren	Untersuchung
Trend Following (Statistical Arbitrage, CTAs)	Long Lookback Straddle			Fung & Hsieh (2001)
Merger Arbitrage	Long risikolose Anleihen plus Short Index Put Option	Risk Arbitrage Index nach Mitchell/Pulvino		Mitchell & Pulvino (2001)
Convergence Trading (Fixed Income Arbitrage)	Short Lookback Straddle		Credit Yield Spread	Fung & Hsieh (2002)
Long/Short Equity			Spread Small Caps minus Large Caps plus Long Equity Index	Fung & Hsieh (2004)

Tabelle 1: Asset-based Style Faktoren

Die Einführung der ABS-Faktoren hat zwar zu einer signifikanten Verbesserung der Stilregression geführt, jedoch bleibt kritisch anzumerken, dass es sich bei diesen Faktoren oftmals um synthetisch, also nicht direkt am Markt beobachtbare Größen handelt, deren Bezug zu den verfolgten Strategien nicht immer ersichtlich ist. Baghai-Wadji et al. (2006) kommen zu dem Schluss, dass die Konstruktion derartiger Faktoren im Rahmen einer Stilanalyse nicht zur Klassifizierung geeignet ist. Generell kritisieren Maillet & Rousset (2001) die Anwendung der linearen (renditebasierten) Stilanalyse insbesondere vor dem Hintergrund, dass die dynamischen Handelsstrategien von Hedge-Fonds oft zu nichtlinearen Rendite Risiko Strukturen führen. In beiden Untersuchungen werden die Self-Organizing Maps (SOMs) als überlegende Klassifikationsmethode heraus gearbeitet, nach-

<sup>11</sup> Hinter dieser Konstruktion steht zum einen die Beobachtung, dass die Renditedifferenz zwischen Aktien mit einer niedrigen Marktkapitalisierung (Small Caps) und Aktien mit einer hohen Marktkapitalisierung (Large Caps) einen statistisch signifikanten Erklärungsbeitrag des systematischen Risikos von Hedge-Fonds mit der benannten Long/Short-Equity Strategie leistet und zum anderem die Feststellung, dass Hedge-Fonds dieser Strategie systematisch einen Long Bias in Aktien und Small Caps und einen Short Bias in Large Caps aufweisen. Vgl. Fung & Hsieh (2004), S. 3 ff.

<sup>12</sup> Tabelle entnommen aus Poddig & Viebig (2006).

dem u.a. Untersuchungen von Ultsch & Vetter (1994), Mangiameli et al. (1996) und Moutarde & Ultsch (2005) deren Superiorität gegenüber traditionellen Clustering-Methoden aus der Statistik demonstriert haben.<sup>13</sup> Die Verwendung von SOMs wird ebenfalls in den Untersuchungen von Deboeck (1998) und Das & Das (2005) angeregt. Im Ergebnis kommen die genannten Studien, im Vergleich zur Vorgabe durch die (verschiedenen) Datenanbieter, zu einer geringeren Anzahl von Hedge-Fonds-Klassen und decken ein hohes Maß an Fehlklassifikation auf.

Motiviert durch diese Erkenntnisse behandelt der vorliegende Artikel mit den Self-Organizing Maps einen Ansatz, der bereits in verschiedenen Untersuchungen seine überlegende Fähigkeit zur objektiven Klassifikation gezeigt hat. Der Beitrag dieses Artikels liegt in der Erweiterung des üblicherweise in der Hedge-Fonds Literatur gewählten Vorgehens. So liefert eine trainierte SOM-Karte (Kohonenschicht) in der Regel noch keine exakte Darstellung der entdeckten Strukturen in den Inputdaten.<sup>14</sup> Zur Überwindung dieses Problems wird das 2-stufige Verfahren nach Vesanto & Alhoniemi (2000) implementiert. Hiernach werden zunächst die SOMs trainiert und in einem zweiten Schritt unter Verwendung des K-means Clustering Algorithmus Partitionierungen abgeleitet. Auf Grundlage der so ermittelten Taxonomie wird der Frage nach der Selbsteinstufungsgüte der Hedge-Fonds-Manager nachgegangen sowie das Ausmaß der beschriebenen „Style Creep“ Problematik untersucht.

Die Ergebnisse der empirischen Untersuchung lassen sich wie folgt zusammenfassen: Wie in vergleichbaren Studien kommt es zu einer Reduktion der durch die Datenbank vorgegebenen Strategiegruppen. Dabei kann im 10-jährigen Untersuchungszeitraum für die 2789 einbezogenen Hedge-Fonds, die originär in 11 Anlagestile organisiert sind, eine 6-Klassen Partitionierung festgestellt werden. Diese Taxonomie ist robust in Hinblick auf unterschiedliche Zeiträume, Topologien und Datensample. Im Rahmen der weiterführenden Analysen zur Selbsteinstufungsgüte und „Style Creep“ Problematik zeigt sich, dass die Eigenangaben überwiegend inkonsistent sind und für 17.37% der Fonds eine Stiländerung festgestellt werden kann. Dabei ist die ex post Wahrscheinlichkeit eines „Style Creep“ von Hedge-Fonds mit falscher im Vergleich zu Hedge-Fonds mit richtiger Einstufung mit 17.71% zu 10.27% deutlich höher.

Der Artikel ist wie folgt gegliedert. Im nächsten Abschnitt werden die Self-Organizing Maps eingeführt sowie die Vorteile des Clustern der Kohonenschicht mit dem k-means Algorithmus als Interpretationshilfe aufgezeigt. Der dritte Abschnitt präsentiert die Daten sowie den Aufbau der

---

<sup>13</sup> Vgl. z.B. Seiler (2008) oder Barès et al. (2001) zur Anwendung von traditionellen Clustering-Methoden in Zusammenhang mit Hedge-Fonds.

<sup>14</sup> Vgl. Poddig & Sidorovitch (2001), S. 391 f.

empirischen Untersuchung und beschreibt die Implementierung des 2-stufigen Ansatzes nach Vesanto & Alhoniemi (2000). Die empirischen Ergebnisse werden im Abschnitt 4 diskutiert. Der fünfte Abschnitt beschließt den Artikel mit der schlussfolgernden Darstellung der wichtigsten Ergebnisse.

## 2 Methodologischer Rahmen

### 2.1 Self-Organizing Maps

Die Self-Organizing Maps wurden von Kohonen in der 1980er Jahren entwickelt.<sup>15</sup> Sie gehören zu den unüberwachten Lernverfahren (Neuronale Netze) und haben damit den Vorteil, ohne a priori Informationen, unbekannte Strukturen in dem zu untersuchenden Datensatz aufzudecken. Die SOMs besitzen dabei die Fähigkeit hochdimensionale Merkmalsräume, unter Wahrung der inhärenten Struktur der eingegebenen Originaldaten (Topologieerhalt), in einen Outputraum niedrigerer Dimension abzubilden.

Die Netzwerkarchitektur eines SOM besteht nur aus einer Input- und einer Output-Schicht. Es handelt sich dabei um ein zweischichtiges „feed forward“ Netzwerk, wobei die Input-Schicht zum Einlesen externer Informationen in das Netzwerk dient. Die Anzahl der Input-Units wird durch die Dimension des abzubildenden Merkmalsraums determiniert. Im Vergleich zu anderen Netzwerktypen weisen SOMs einige Besonderheiten auf. So sind zunächst die Units der Input-Schicht vollständig mit den Units der Output-Schicht verbunden. Darüber hinaus, und darin liegt die eigentliche Besonderheit, besteht für die Units auf der Output-Schicht (Kohonenschicht) eine fest definierte Ordnungsrelation, die die räumliche Nachbarschaft untereinander anzeigt. Durch diese fixierten Nachbarschaftbeziehungen der Output-Units wird der Topologieerhalt bei der dimensionsreduzierenden Abbildung des Inputraums ermöglicht. Ferner wird dadurch eine Metrik implementiert, welche es erlaubt, Abstände zwischen den Output-Units zu quantifizieren. Dabei sind unterschiedliche Anordnungen der Output-Units (z.B. Kettenstrukturen sowie zwei- bzw. dreidimensionale Gitterstrukturen) denkbar. In der Regel werden jedoch zweidimensionale Gitterstrukturen (sogenannte Maps) verwendet. In der Mitte der Abbildung 1 wird eine Gitterstruktur, wie sie typischerweise in den Untersuchungen zur Klassifikation von (Hedge-) Fonds implementiert wird, exemplarisch dargestellt.<sup>16</sup>

SOMs werden in einem iterativen Prozess trainiert. Dabei wird in jedem Schritt  $t$  ein Trainingsvektor  $\mathbf{x}(t)$  aus den Inputdaten gemäß einer Wahrscheinlichkeitsfunktion ausgewählt und die

---

<sup>15</sup> Vgl. Kohonen (1981), Kohonen (1982a), Kohonen (1982b) und Kohonen (1982c).

<sup>16</sup> Vgl. Baghai-Wadji et al. (2006), Das & Das (2005), Maillet & Rousset (2001) und Deboeck (1998).

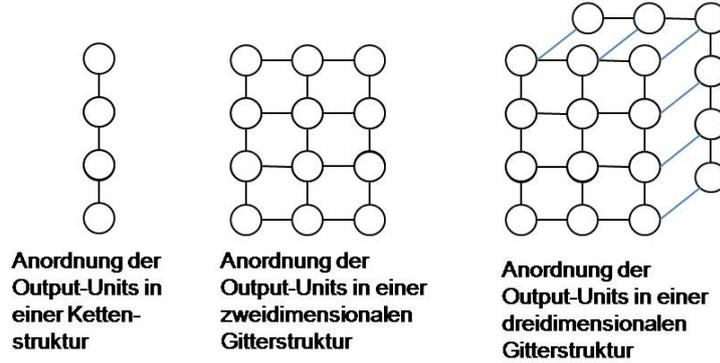


Abbildung 1: Exemplarische Anordnungen der Output-Units

Distanzen  $d_i$  zu den Gewichtsvektoren (auch Prototypvektoren)  $\mathbf{w}_i(t) = [w_{i1}, \dots, w_{in}]$ , wobei  $n$  der Dimension des Inputvektors entspricht, aller Output-Units auf der Kohonenschicht bestimmt.<sup>17</sup> Zur Berechnung der Distanz wird in der Literatur das Euklidische Maß  $ed$  vorgeschlagen:<sup>18</sup>

$$ed_i = \|x(t) - w_i(t)\|_2 := \sqrt{\sum_{j=1}^n (x_j(t) - w_{ij}(t))^2} \quad (1)$$

Das Ziel besteht in der Ermittlung der best matching unit  $\mathbf{w}_{\text{BMU}}$  (BMU), welche die geringste Distanz zum Trainingsvektor  $\mathbf{x}(t)$  aufweist:

$$\|\mathbf{x}(t) - \mathbf{w}_{\text{BMU}}(t)\| = \min \{\|\mathbf{x}(t) - \mathbf{w}_i(t)\|\} \quad (2)$$

Nachdem die BMU, welche auch als Winning-Unit bezeichnet wird, ermittelt ist, beginnt der eigentliche Lernprozess: die Gewichte der BMU sowie die Gewichte der benachbarten Units innerhalb eines bestimmten Radius  $\sigma$  werden dahingehend modifiziert, dass sie sich nach Gleichung (3) in die Richtung des gewählten Inputvektors  $x(t)$  anpassen.

$$\mathbf{w}_i(\text{neu}) = \mathbf{w}_i(\text{alt}) + \alpha(t)h_{\text{BMU}_i}(t) [\mathbf{x}(t) - \mathbf{w}(t)] \quad \text{mit} \quad (3)$$

$\alpha(t)$ : Lernrate,  $0 \leq \alpha(t) \leq 1$

$h_{\text{BMU}_i}(t)$ : Nachbarschaftsfunktion

Bei diesem Prozess „lernt“ die BMU am stärksten. Welche der verbleibenden Output-Units sich mit welcher Intensität verändern, wird durch die Nachbarschaftsfunktion  $h_{\text{BMU}_i}(t)$  bestimmt.

<sup>17</sup> Output-Units werden durch die Gewichtsvektoren  $\mathbf{w}_i(t)$  repräsentiert.

<sup>18</sup> Vgl. Poddig & Sidorovitch (2001), S. 385; Kohonen (1997), S. 130.

Dabei sind sowohl der Abstand zwischen der BMU und den benachbarten Output-Units  $U_i$  als auch der oben genannte Radius  $\sigma$  maßgeblich. Als geeignete Nachbarschaftsfunktion nennen Poddig & Sidorovitch (2001) die Gauß'sche Glockenkurve, welche ihr Maximum und ihr Symmetriezentrum an der BMU hat:

$$h_{BMU_i} = \exp \left[ -\frac{\|BMU - U_i\|^2}{2\sigma^2(t)} \right] \quad (4)$$

Sowohl für den Radius  $\sigma(t)$  als auch für die Lernrate  $\alpha(t)$  wird in der Literatur eine zeitlich monoton fallende Funktion vorgeschlagen. Dadurch kommt es anfänglich zunächst zu einer schnellen und groben Entfaltung des Netzes, welches im weiteren Ablauf immer weiter verfeinert werden kann.<sup>19</sup>

Nachdem die Trainingsphase abgeschlossen ist, wobei in der Literatur zwischen 10.000 und 100.000 Lernschritte empfohlen werden, haben sich die Output-Units über den (Hyper-) Raum verteilt.<sup>20</sup> Letztendlich hat eine Segmentierung des Inputraums in einzelne Einzugsgebiete (Voronoi-Elemente) jeweiliger Output-Units stattgefunden. Dabei stellen die Gewichtsvektoren der Output-Units die Mittelpunkte dieser Voronoi-Elemente dar und können als Clusterzentroide ihrer Einzugsgebiete aufgefasst werden.<sup>21</sup> Die Bestimmung der Gewichte eines SOM beruht auf der Minimierung folgender Fehlerfunktion:

$$E = \sum_{k=1}^N \sum_{i=1}^M h_{BMU_i} \|\mathbf{x}_k - \mathbf{w}_i\|^2 \quad (5)$$

Diese Minimierung erfolgt im Rahmen des iterativen Verfahrens nach Gleichung (3), welches auch als online oder Muster-pro-Muster Gewichtsanpassungsprozedur bekannt ist und als eine Form des stochastischen Lernens aufgefasst werden kann. Die Bestimmung des aktuellen Gradienten nach jedem präsentierten Trainingsvektor führt zu einer „Zitterbewegung“ um den wahren Gradienten und vermeidet somit lokale Minima.<sup>22</sup> Sollte sich ein SOM in einem lokalen Minimum befinden, ist der Gradient über alle Trainingsmuster gleich Null. Dieses muss jedoch nicht für die Gradienten der einzelnen Trainingsmuster gelten. Aufgrund der Lernregel nach Gleichung (3), in der die Einzelgradienten bestimmt werden und die Gewichts Anpassung unmittelbar erfolgt, kann das Netz aus einem lokalem Minimum heraus geführt werden.<sup>23</sup>

<sup>19</sup> Vgl. Kerling (1997), S. 419, Kohonen (1997), S. 87 f.; Vesanto & Alhoniemi (2000), S. 589.

<sup>20</sup> Vgl. Kohonen (1997), S. 88; Poddig & Sidorovitch (2001), S. 387.

<sup>21</sup> Vgl. Kerling (1997), S. 420.

<sup>22</sup> Vgl. Zimmermann (1994), S. 42.

<sup>23</sup> Neben der hier vorgestellten Muster-pro-Muster Gewichts Anpassungsprozedur hat Kohonen (1997) auch eine batch-Version dieses Algorithmus entwickelt. Dieser benötigt keine Lernrate und ist durch eine deutlich schnellere

Bei näherer Betrachtung der im Rahmen des iterativen Lernfahrens nach (3) zu minimierenden Fehlerfunktion lassen sich große Ähnlichkeiten zum k-means Clustering feststellen. Ist der Radius  $\sigma$  Null, so nimmt die Nachbarschaftsfunktion  $h_{BMU_i}$  den Wert Eins für die BMU und Null für alle verbliebenen Output-Units an. Damit reduziert sich die Fehlerfunktion (5) zur Fehlerfunktion des k-means Algorithmus:<sup>24</sup>

$$E = \sum_{k=1}^C \sum_{x \in Q_k} \|\mathbf{x}_k - \mathbf{c}_k\|^2 \quad (6)$$

wobei  $C$  die Anzahl der Cluster  $Q_i = 1, \dots, C$  und  $c_k$  das Zentrum des Clusters  $k$  angibt. Weist die Nachbarschaftsfunktion hingegen einen Wert größer Null auf, wird die topologische Nachbarschaft der BMU berücksichtigt. Aus Gleichung (5) folgt dann, dass die Prototypvektoren sich nicht im Clusterzentrum ihrer Voronoi-Elemente befinden, sondern ein mit der Nachbarschaftsfunktion gewichtetes Mittel bilden.

Die Vorteile der SOMs gegenüber einer herkömmlichen k-means Clusteranalyse liegen in der bereits angesprochenen Fähigkeit, hochdimensionale Merkmalsräume, unter Wahrung der inhärenten Struktur der eingegebenen Originaldaten (Topologieerhalt), in einen Outputraum niedrigerer Dimension abzubilden (Dimensionsreduktion). Durch den Topologieerhalt bleiben die Beziehungen im ursprünglichen Merkmalsraum in der (dimensionsreduzierten) Abbildung erhalten. Damit entsprechen benachbarte Units auf der Output-Schicht ähnlichen (Daten-)Clustern im originären Merkmalsraum. Diese Eigenschaft widerspricht dem Grundgedanken der Clusteranalyse, wonach die Variation zwischen den verschiedenen Clustern möglichst groß sein soll, erlaubt hingegen aber das Gruppieren benachbarter Output-Units und damit das Zusammenfassen ähnlicher Datengruppen des Merkmalsraums. Letztendlich kann über eine derartige Zusammenfassung eine Form von hierarchischer Ordnung in den aufgedeckten Strukturen eingebracht werden. Ein weiterer wesentlicher Vorteil besteht in der Approximation der Wahrscheinlichkeitsdichte des ursprünglichen Merkmalsraums. Im Gegensatz zur Clusteranalyse, in der eine schwache Datenhäufung genauso durch ein einzelnes Cluster abgebildet wird wie eine starke Häufung in einem anderen Unterraum, erfolgt bei den SOMs eine differenzierte Behandlung, indem häufig vertretenen Unterräumen mehr Units auf der Output-Schicht zugeordnet werden. Wichtige Datenbereiche können so differenziert abgebildet werden.

Letztendlich handelt es sich bei dem SOM-Algorithmus um ein partitionierendes Verfahren. Allerdings führt der angesprochene Topologieerhalt zu einer Kontinuität in der Abbildung, wodurch

---

Konvergenz gekennzeichnet. Vgl. Kohonen (1997), S. 135 ff. Insbesondere vor dem Hintergrund lokaler Minima wird in weiteren Verlauf dieser Arbeit der (adaptive) Muster-pro-Muster Algorithmus verwendet.  
<sup>24</sup> Zum k-means Clustering siehe z.B. MacQueen (1967), S. 281 ff.

keine eindeutige Anzahl an Cluster nach Abschluss der Trainingsphase angegeben werden kann. Eine exakte Bestimmung von Clustergrenzen erfordert weitere Verfahren.

## 2.2 Clustern von Self-Organizing Maps

Typischerweise erfolgt die dimensionsreduzierte Abbildung eines SOMs auf einer zweidimensionalen Karte. Eine Möglichkeit zur Bestimmung der Clusteranzahl besteht daher in der Visualisierung dieser trainierten Kohonenschicht („Clustering via Visualization“). Der einfachste Ansatz besteht im sogenannten „Labeling“ der trainierten Karte. Hierbei werden die Trainingsdaten dem SOM nochmals präsentiert und die jeweiligen BMUs entsprechend den zugehörigen Inputdaten gekennzeichnet. Ein fortgeschrittener Ansatz stellt die sogenannte „Unified Distance Matrix“ (U-Matrix) dar. Die U-Matrix fügt auf der Kohonenschicht als dritte Dimension die Euklidischen Distanzen zwischen jeder Output-Unit und ihren direkten Nachbarn hinzu. Damit wird die manuell vorzunehmende Clusterbildung erleichtert.<sup>25</sup>

Letztendlich können die Visualisierungstechniken nur eine Hilfestellung für die im Anschluss „manuell“ vorzunehmende Clusterbildung geben. Dieses ist ein mühsamer Prozess, der keine konsistente Zusammenfassung benachbarter Output-Units garantiert. Zur Überwindung dieser Schwäche regen Vesanto & Alhoniemi (2000) eine „automatisierte“ Vorgehensweise an, welche in Abbildung 2 dargestellt ist.

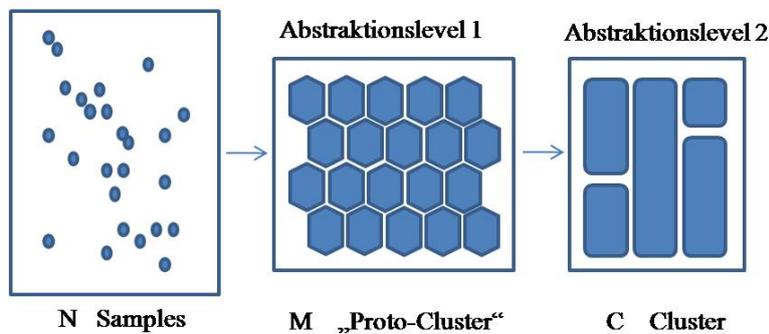


Abbildung 2: 2-stufiger Ansatz nach Vesanto & Alhoniemi (2000)

In der ersten Stufe erfolgt mit Hilfe der Self-Organizing Maps eine dimensionsreduzierte Abbildung des originären Merkmalsraums unter Wahrung der inhärenten Struktur der eingegebenen

<sup>25</sup> Ausführliche Darstellungen zur U-Matrix finden sich z.B. bei Ultsch (1993). Für weitere in der Literatur diskutierte Visualisierungstechniken wie z.B. die P-, und die U\*-Matrix siehe Ultsch (2003). Für das Adaptive Coordinates Verfahren siehe Merkl & Rauber (1997) und die dort genannten Quellen.

Originaldaten. Das nach Abschluss der Trainingsphase entstandene Set an Prototypvektoren (Abstraktionslevel 1) übersteigt in der Regel die Anzahl der erwarteten Cluster. Die Prototypvektoren können jedoch als „Proto-Cluster“ aufgefasst werden, welche in einem zweiten Schritt zu den eigentlichen Cluster (Abstraktionslevel 2) zusammengefasst werden können. Dazu nutzen Vesanto & Alhoniemi (2000) das k-means Clustering Verfahren, welches sich durch folgende fünf Schritte skizzieren lässt:<sup>26</sup>

1. Bestimmung der Clusteranzahl
2. Zufällige Initialisierung der Clusterzentren
3. Berechnung der Partitionierung/Klassifizierung
4. Berechnung der neuen Clusterzentren
5. Wiederhole Schritte 3 bis 5, bis die Partitionierung unverändert bleibt oder der Algorithmus konvergiert ist

Die zu minimierende Fehlerfunktion, wonach der Algorithmus die Clusterzentren für  $C$  Cluster derart bestimmt, dass die Summe der Streuungsquadrate innerhalb der Cluster minimal über alle Cluster wird, wurde bereits in Gleichung (6) aufgeführt.

Ist die „wahre“ Anzahl an Clustern, wie in der vorliegenden Untersuchung a-priori unbekannt, so kann der oben vorgestellte Algorithmus für ein Set von unterschiedlichen Clusteranzahlen wiederholt werden. Die „beste“ Partitionierung wird dann im Anschluss unter Verwendung eines Validitätsindex bestimmt. Im Zusammenhang mit SOMs nutzen Vesanto & Alhoniemi (2000) den Davies-Bouldin Index (DBI), welcher die Kompaktheit innerhalb der Cluster in Beziehung zur Separiertheit zwischen den Clustern setzt.<sup>27</sup> Mit  $S_c(Q_k)$  als mittlerer Abstand (gemittelt über alle Objekte  $N_k$ ) im Cluster  $Q_k$  zum Clustermittelpunkt  $c_k$  und  $d_{ce}(Q_k, Q_l)$  als Abstand zwischen den Cluster  $Q_k$  und  $Q_l$  ergibt sich der *DBI* wie folgt:

$$DBI(C) = \frac{1}{C} \sum_{k=1}^C \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \quad \text{mit} \quad (7)$$

$$S_c = \frac{\sum_i \|x_i - c_k\|}{N_k}$$

$$d_{ce} = \|c_k - c_l\|$$

<sup>26</sup> Siehe MacQueen (1967), S. 281 ff.

<sup>27</sup> Siehe Davies & Bouldin (1979). Für alternative Validitätsindices siehe Bezdek (1998), oder Milligan & Cooper (1985).

Zur Berechnung der Kompaktheit und Separiertheit verwenden Vesanto & Alhoniemi (2000) die centroid distance ( $S_c$ ) bzw. centroid linkage ( $d_{ce}$ ) und begründen die Wahl dieser Distanzmaße mit dem Problem steigender Rechenkomplexität alternativer Maße, wie beispielsweise die average distance ( $S_a$ ) bzw. das average linkage ( $D_a$ ), insbesondere bei größeren Datenmengen.<sup>28</sup>

Den Vorteil ihres zweistufigen Ansatzes gegenüber einem direkten Clustern des Datensatzes sehen die Autoren neben der Verringerung des Rechenaufwands in der Reduktion des Rauschens. Die Prototypvektoren („Proto-Cluster“) sind lokale Durchschnitte der Daten und daher im Vergleich zu den Originaldaten weniger sensitiv gegenüber zufälligen Veränderungen. Letztendlich kommt es zu einer Verringerung des negativen Einflusses von Ausreißern auf das im zweiten Schritt durchzuführende Clustering.<sup>29</sup>

Durch die Überwindung der oben angesprochenen Schwächen der Visualisierungstechniken ermöglicht dieses zweistufiger Verfahren letztendlich eine objektive Klassifizierung der originären Daten bei gleichzeitiger Wahrung der genannten Vorteile der SOMs gegenüber traditionellen Verfahren. Die Abbildung 3 zeigt exemplarisch eine trainierte SOM-Karte und verdeutlicht anhand des „Labelings“, des U-Matrix- und des k-means Clustering Verfahrens die Bestimmung der Clusteranzahl.

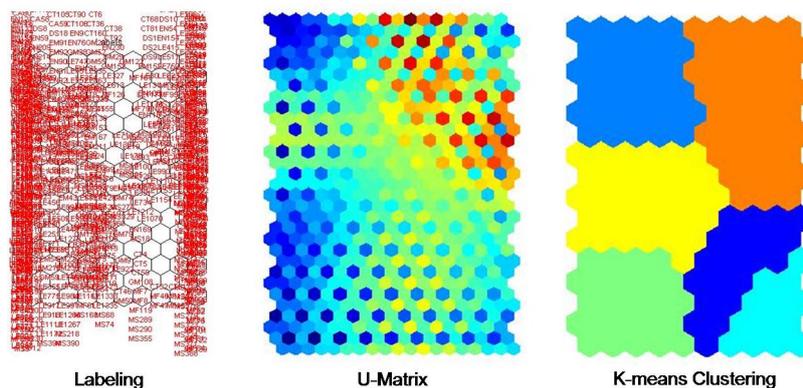


Abbildung 3: Auswertung trainierte SOM-Karte

Wie aus der Abbildung 3 hervorgeht ermöglicht nur das k-means Clustering eine eindeutige Bestimmung der Clusteranzahl. Es zeigt sich ferner, dass insbesondere bei größeren Datenmengen der „Labeling“-Ansatz zur Ermittlung der Cluster ungeeignet ist und das U-Matrix Verfahren einen großen Interpretationsspielraum lässt.

<sup>28</sup> Vgl. Vesanto & Alhoniemi (2000), S. 587 f.

<sup>29</sup> Siehe Vesanto & Alhoniemi (2000), S. 588. Zur Rauschreduktion im Zusammenhang mit SOM Vgl. Kerling (1997), S. 420 f.

## 3 Empirische Umsetzung des 2-stufigen Klassifikationsansatzes

### 3.1 Untersuchungsgegenstand und Daten

In der vorliegenden Untersuchung soll ein systematischer Ansatz zur Klassifikation von (Single) Hedge-Fonds abgeleitet werden. Die empirischen Studien von Mailet & Rousset (2001), Das & Das (2005) oder Baghai-Wadji et al. (2006) haben gezeigt, dass SOMs das geeignete Werkzeug zur Umsetzung bieten. Darüber hinaus zeigen die Untersuchungen von Ultsch & Vetter (1994), Mangiameli et al. (1996) oder Moutarde & Ultsch (2005), dass die SOMs im Vergleich zu den traditionellen Klassifizierungsansätzen verrauschte Datensätze besser klassifizieren können und dazu weniger zusätzliche Informationen benötigen. In den bisherigen Studien zur Klassifizierung von Hedge-Fonds mit SOMs erfolgte die Analyse unter Verwendung der oben kritisierten Visualisierungstechniken, welche Interpretationsspielräume bei der Clusterbildung lassen und damit keine konsistente Zusammenfassung benachbarter Output-Units garantieren. Mit der Anwendung des 2-stufigen Ansatzes nach Vesanto & Alhoniemi (2000) erweitert die vorliegende Untersuchung das bis dahin gewählte Vorgehen in der Hedge-Fonds Literatur und stellt darüber hinaus eine objektive, konsistente Klassifizierung, die frei von Interpretationsspielräumen ist, sicher.

Das so entwickelte objektive und systematische Klassifikationsinstrumentarium soll zur Beantwortung der Frage nach der Qualität und der Stabilität der Eigenangabe von Hedge-Fonds über ihren Anlagestil genutzt werden. In Anlehnung an die Studie von Baghai-Wadji et al. (2006) wird zur Messung der Qualität und Stabilität die ermittelte Klassifikation der Einteilung des Datenbankanbieters gegenübergestellt.

Für die empirische Analyse stehen monatliche Renditen aus der Lipper-TASS Datenbank zur Verfügung. Der Datensatz umfasst 11 Strategien von 2863 Hedge-Fonds im Zeitraum vom 31.01.1999 bis zum 31.12.2008. Ackermann et al. (1999) folgend, werden nur Fonds mit einer Mindesthistorie von 24 Monaten in die Analyse aufgenommen, um die Stabilität der Berechnungen zu gewährleisten. Damit verbleiben 2789 Hedge-Fonds im Datensatz. Für die empirische Untersuchung werden 2 Datensample gebildet. Das erste Sample umfasst 1025 Fonds im Zeitraum vom 31.01.2001 bis zum 31.12.2008 und zeichnet sich dadurch aus, dass alle Hedge-Fonds die gleiche Renditehistorie von 84 Monaten aufweisen. Zur Überprüfung der Stabilität der Klassifizierung, wird dieses Sample in 2 Sub-Samples von jeweils 42 Monaten unterteilt. Das zweite Datensample enthält alle 2789 Fonds über den gesamten Untersuchungszeitraum und wird für einen weiteren Stabilitätscheck genutzt. Darüber hinaus dient es zur Untersuchung des SOM-Algorithmus im Umgang mit „Missing Va-

lues“. In der Tabelle 2 werden die zur Verfügung stehende Anzahl an Hedge-Fonds je Datensample angegeben.

Codierung	Anlagestrategie gemäß Eigenangabe	Anzahl der Hedge Fonds	
		Datensample 1	Datensample 2
		Sub-Samples 1 und 2	
CT	Commodity Trading	111	163
CA	Convertible Arbitrage	48	86
DS	Dedicated Short Bias	8	17
EM	Emerging Markets	80	229
EN	Equity Market Neutral	70	220
GM	Global Macro	39	146
LO	Long Only	2	12
LE	Long Short Equity	444	1287
MF	Managed Futures	109	236
MS	Multi Strategies	112	386
OA	Options Arbitrage	2	7
Gesamt		1025	2789

Tabelle 2: Überblick der zur Verfügung stehenden Anlagestile

Im Gegensatz zu traditionellen Clusterverfahren ermöglichen Self-Organizing Maps Hedge-Fonds mit einer unterschiedlichen Renditehistorie zusammenzufassen.<sup>30</sup> Die Renditehistorien der einzelnen Fonds in diesem Sample sind unterschiedlich lang, umfassen aber mindestens 24 Beobachtungen. Die Tabelle 7 im Appendix fasst die deskriptiven Statistiken der Clusterzentroide des ersten Datensamples zusammen.

### 3.2 Implementierung des 2-stufigen Ansatzes

Zur Umsetzung des vorgestellten zweistufigen Klassifizierungsansatzes müssen zunächst in der ersten Stufe die modellspezifischen „Freiheitsgrade“ der SOMs konkretisiert werden. Wie bei jedem Neuronalen Netz zählt hierzu die Spezifikation der Netzwerkarchitektur. In Abschnitt 2.1 werden SOMs als zweischichtige „feed forward“ Netze mit einer festen Anordnung der Units auf der Output-Schicht beschrieben. Aufgrund dieser besonderen Architektur beschränkt sich die Netzwerkspezifikation auf die Festlegung der Anzahl der Output-Units sowie deren Anordnung auf der Output-Schicht.<sup>31</sup> Basierend auf die in der Einführung genannten Untersuchungen, insbesondere der Studie von Baghai-Wadji et al. (2006), die auf einem ähnlich großen Datensample beruht, werden in der vorliegenden Studie verschiedene  $10 \times 10^{S1,S2}$ ,  $15 \times 15^{S1,S2}$ ,  $20 \times 20^{S2}$  und  $25 \times 25^{S2}$  sowie

<sup>30</sup> Vgl. Baghai-Wadji et al. (2006), S. 162.

<sup>31</sup> Für eine tieferegehende Diskussion vgl. z. B. Blackmore & Miikkulainen (1993), S. 450 ff. oder Fritzsche (1994), S. 1441 ff.

verschiedene auf einer von Vesanto et al. (2000) dokumentierten Heuristik ermittelte  $16 \times 10^{S1}$ ,  $20 \times 8^{S1}$  und  $25 \times 11^{S2}$  Topologien getestet.<sup>32</sup>

Weitere Freiheitsgrade bestehen in der Wahl der Lernrate  $\alpha$ , der Nachbarschaftsfunktion  $h_{BMU}$ , des Nachbarschaftsradius  $\sigma$  sowie der Anzahl der Lernschritte  $t$ . Verfahren, die eine optimale Bestimmung dieser Parameter ermöglichen, sind nicht bekannt. Daher wird die Spezifizierung auf Grundlage empirisch gewonnener Erkenntnisse vorgenommen. In Anlehnung an die Studie von Baghai-Wadji et al. (2006) erfolgt die Initialisierung der Lernrate mit 0.06 und des Nachbarschaftsradius mit 11. Kohonen (1997) folgend wird der Radius der Nachbarschaft als linear abnehmende Funktion und die Anpassung der Lernrate auf Grundlage einer invers monoton abnehmenden Funktion jeweils in Abhängigkeit der Lernschritte  $t$  bestimmt. Die Anzahl der Lernschritte  $t$  ergibt sich nach Kohonen (1997) als das 500-fache der Output-Units.<sup>33</sup> Die Gauß'sche Glockenkurve wird aus den in Abschnitt 2.1 bereits aufgezeigten Eigenschaften als Nachbarschaftsfunktion gewählt.

Neben den hier aufgeführten Freiheitsgraden erfordert die stochastische Natur des Lernprozesses, welcher zum einen von der Reihenfolge der gewählten Trainingsvektoren als auch von der Anfangsinitialisierung der Output-Units abhängig ist, eine mehrfache Wiederholung des Trainings der oben aufgeführten Topologien mit unterschiedlichen Parametrisierungen und Initialisierungen. Zur anschließenden Identifizierung der „adäquaten“ SOM-Realisation werden in der Literatur der mittlere Quantisierungsfehler (mQuant) und der Topologiefehler (TopoErr) diskutiert. Der mittlere Quantisierungsfehler, welcher an die Fehlerfunktion der Self-Organizing Maps (Gl. 5) anknüpft, ergibt sich aus der mittleren Euklidischen Distanz zwischen allen Trainingsvektoren mit deren best matching units.<sup>34</sup> Der Topologiefehler beurteilt die Qualität einer SOM-Realisation über die Kontinuität der Abbildung vom Inputraum auf die trainierte SOM-Karte. Nach Kaski & Lagus (1996) lässt sich dieses Maß als Prozentsatz der Inputvektoren, deren best matching units nicht mit deren second best matching units benachbart sind, operationalisieren.<sup>35</sup>

In der zweiten Stufe des Klassifizierungsansatzes erfolgt das Clustering der in der ersten Stufe identifizierten „adäquaten“ SOM-Karte zur objektiven Bestimmung der Cluster Grenzen. Wie in Abschnitt 2.2 ausgeführt, nutzen Vesanto & Alhoniemi (2000) den k-means Algorithmus, welcher auch in der vorliegenden Untersuchung zur Anwendung kommt. Aufgrund der Sensitivität dieses Algorithmus gegenüber der Anfangsinitialisierung wird dieses Verfahren 1000-mal mit zufälligen Startwerten durchgeführt. Dabei wird in jedem dieser 1000 Durchgänge ein Set von unterschiedli-

---

<sup>32</sup> <sup>S1</sup> Datensample 1; <sup>S2</sup> Datensample 2.

<sup>33</sup> Vgl. Kohonen (1997), S. 87 f.

<sup>34</sup> Vgl. Kohonen (1997), S. 121.

<sup>35</sup> Vgl. Kaski & Lagus (1996), S. 810 ff.

chen Clusteranzahlen getestet und unter Verwendung des Davies-Bouldin Indices die Clusteranzahl zur „besten“ Partitionierung ermittelt.<sup>36</sup> Letztendlich ergibt sich auf diesem Wege eine Verteilung der Clusteranzahlen. Zur Bestimmung der finalen Partitionierung wird das k-means Clusterverfahren erneut auf die identifizierte SOM-Karte mit dem Unterschied angewendet, dass der Modalwert der ermittelten Verteilung der Clusteranzahlen als „wahrer“ Wert vorgegeben ist. Aus den oben genannten Grund wird das Verfahren 1000-mal durchgeführt und in jedem Durchgang die Summe der mittleren Euklidischen Distanzen über alle Cluster ( $mEUKLID$ ) als Gütekriterium bestimmt:

$$mEUKLID = \frac{1}{C} \sum_{k=1}^C \left( \frac{\sqrt{\sum_i (x_i - c_k)^2}}{N_k} \right) \quad (8)$$

Der Durchlauf mit der geringsten  $mEUKLID$  wird in der vorliegenden Untersuchung als die finale Partitionierung der in der ersten Stufe ermittelten „adäquaten“ SOM-Karte identifiziert.

## 4 Ergebnisse

Wie in der Studie von Baghai-Wadji et al. (2006) führen die unterschiedlichen Topologien zu robusten Partitionierungen der Untersuchungssample 1 und 2. Dabei zeichnen sich jedoch die heuristisch bestimmten Topologien durch eine größere Stabilität der Verteilung der Clusteranzahlen beziehungsweise durch höhere Fallzahlen für den Modalwert aus. Im weiteren Verlauf erfolgt daher die Ergebnisdiskussion auf Grundlage der heuristisch bestimmten Topologien.

In Übereinstimmung mit den Untersuchungen von Das & Das (2005) und Baghai-Wadji et al. (2006), in denen eine Reduktion der ursprünglichen Clusteranzahl auf 6 festgestellt wird, führt die Anwendung des vorgestellten zweistufigen Klassifizierungsansatzes auf das Datensample 1 ebenfalls zu einer Reduktion von originär 11 auf 6 Strategiegruppen.<sup>37</sup> Die damit einhergehende Verringerung der Summe der mittleren Euklidischen Distanzen über alle Cluster ( $mEUKLID$ ) von originär 8.29 auf 7.01 als auch die Reduktion des Davies-Bouldin Index von 5.48 auf 3.17 zeigen eine Verbesserung der Partitionierung des Datensamples.

<sup>36</sup> Siehe Abschnitt 2.2 für weitere Ausführungen zum Davies-Bouldin Index. Die maximal zu testende Clusteranzahl beträgt nach Vesanto et al. (2000):  $\sqrt{\text{Anzahl Fonds}} = \sqrt{1025} \doteq 32$ .

<sup>37</sup> Wie in der vorliegenden Untersuchung werden bei Baghai-Wadji et al. (2006) 2 verschiedene Datensample mit Hedge-Fonds gleicher sowie unterschiedlicher Renditehistorie gebildet. Während sich für das Sample mit gleicher Historie eine Reduktion von 14 auf 6 Cluster ergibt, verringert sich die Strategiegruppen beim Sample mit unterschiedlicher Historie von 15 auf 9. In der Untersuchung von Das & Das (2005) ergibt sich eine Reduktion von originär 11 auf 6 Gruppen.

Datensample 1: 31.01.2001 bis 31.12.2008													
Cluster	Bezeichnung	CT	CA	DS	EM	EN	GM	LO	LE	MF	MS	OA	Gesamt
C11:	<b>DS</b> & GM & MS	12.61	27.08	<b>75.00</b>	17.50	25.71	<b>25.64</b>	0.00	26.35	17.43	<b>35.71</b>	0.00	251
C12:	CA & <b>EN</b>	24.32	<b>39.58</b>	25.00	16.25	<b>55.71</b>	17.95	0.00	13.06	15.60	23.21	0.00	208
C13:	<b>CT</b> & <b>MF</b>	<b>55.86</b>	2.08	0.00	0.00	0.00	12.82	0.00	0.00	<b>61.47</b>	3.57	50.00	140
C14:	<b>EM</b>	1.80	6.25	0.00	<b>55.00</b>	5.71	17.95	100.00	26.58	2.75	14.29	0.00	199
C15:	<b>LE</b>	5.41	22.92	0.00	3.75	10.00	12.82	0.00	<b>27.48</b>	2.75	14.29	50.00	174
C16:	Sonstige	0.00	2.08	0.00	7.50	2.86	12.82	0.00	6.53	0.00	8.93	0.00	53
Summe (%)		100	100	100	100	100	100	100	100	100	100	100	
Gesamt <sup>1</sup>		111	48	8	80	70	39	2	444	109	112	2	1025

Darstellung der Ergebnisse in Anlehnung an die Untersuchung von Baghai-Wadji et al. (2006). Eigene Berechnung.

Die Ergebnisse sind in Prozent angegeben.

<sup>1</sup> Gesamtzahl der Hedge Fonds in der jeweiligen Kategorie.

Tabelle 3: Vergleich des Klassifizierungsansatzes (Zeilen) mit der originären Datenbankgruppierung (Spalten) des Datensample 1

Die Tabelle 3 stellt die finale Partitionierung des zweistufigen Klassifizierungsansatzes der originären Gruppierung der Datenbank gegenüber. Zur Analyse der Konsistenz der Selbsteinstufungen der Hedge-Fonds-Manager erfolgt in Anlehnung an Fung & Hsieh (1997), Brown & Goetzmann (2003) sowie Baghai-Wadji et al. (2006) die Bezeichnung der Cluster gemäß der vorherrschenden selbstdeklarierten Strategien der originären Klassifizierung in einer (ermittelten) Gruppe. Wie aus der Tabelle 3 hervorgeht, ist die Selbsteinstufung der Hedge-Fonds im untersuchten Datensample in Frage zu stellen. Wird zur Beurteilung der Selbsteinstufungsgüte die Quote, wie in der Studie von Baghai-Wadji et al. (2006) auf 65% gesetzt, sind lediglich die DS-Fonds weitgehend konsistent mit der selbstdeklarierten Strategie.<sup>38</sup> Die Fehleinstufungsquoten von CA-, GM-, LE- und MS-Fonds liegen zwischen 60 und 75%. Die Fonds dieser Strategiegruppen verteilen sich über die ermittelte SOM-Karte und scheinen daher für eine nützliche Selbsteinstufung ungeeignet. Zu stabileren Einstufungen kommen die Manager der CT-, EM-, EN- und MF-Fonds. Hier liegt die Fehlklassifikation zwischen 35 und 50%.

Die Benennung der Cluster auf Basis der vorherrschenden selbstdeklarierten Strategien und die daraus gewonnenen Erkenntnisse sind kritisch zu hinterfragen, da die Fonds mit sehr hoher Fehleinstufungsquote die Interpretation der Konsistenzanalyse beeinflussen können. Als Beispiel sei hier das Cluster 1 angeführt, welches aufgrund der enthaltenen vorherrschenden selbstdeklarierten Strategien der originären Klassifizierung als DS&GM&MS bezeichnet wird. Die Auswertung der Zusammensetzung des Clusters 1 zeigt jedoch, dass im Innerclustervergleich, die GM-Fonds zur Bezeichnung des Clusters nicht herangezogen worden wären. Die hervorgehobenen Komponenten der Clusterbezeichnungen in der Tabelle 3 zeigen jene Strategiegruppen, die sowohl vorherrschend innerhalb der originären als auch in der ermittelten Klassifizierung sind. Danach können DS, EN, EM und LE als einzelne charakteristische Hedge-Fonds-Stile sowie CT und MF als eine (zusammengefasste) Strategie identifiziert werden. Neben diesem Ergebnis lassen sich Aussagen bezüglich der Ähnlichkeit der untersuchten Strategien ableiten. Die Zusammenfassung von CT und MF zu einem Cluster zeigt, dass sich gemäß dem vorgestellten 2-stufigen Klassifizierungsansatz diese Strategien nicht unterscheiden und es sich letztendlich um denselben Anlagestil handelt. Das scheint aus ökonomischer Sicht plausibel, da die Handelsstrategie der MF (von sogenannten Commodity Trading Advisors) fast ausschließlich über börsengehandelte Derivate, wie Futures, abgebildet wird und Commodities in der Regel über Futurekontrakte gehandelt werden.

---

<sup>38</sup> Die LO- und OA-Strategien werden aufgrund ihres Umfanges von je 2 Fonds in der Analyse nicht mit einbezogen.

Trotz der genannten Kritik an der beschriebenen Vorgehensweise der Konsistenzanalyse bleibt letztendlich festzuhalten, dass aufgrund der Verteilung der Fonds über die ermittelte Partitionierung die Selbsteinstufung durch die Hedge-Fonds-Manager überwiegend inkonsistent ist.

Die Folgen dieser gewollten oder ungewollten Klassifizierungsfehler sind weitreichend, da sie zu Verzerrungen in der Analyse von Hedge-Fonds-Strategien und deren Performance führen sowie das Gefahrenpotenzial beinhalten, falsche Investitionsentscheidungen, beispielsweise zur Konstruktion von Dachfonds oder zur Beimischung eines bestehenden Portfolios, zu treffen. Die Tabelle 8 im Appendix fasst die deskriptiven Statistiken der Zentroide der ermittelten 6 Cluster zusammen.

Unabhängig von der Konsistenzanalyse soll in einem weiteren Schritt die Stabilität der ermittelten 6-Gruppen Klassifizierung geprüft werden. Dazu wird der 2-stufige Ansatz auf die beiden in Abschnitt 3.1 benannten Sub-Samples angewendet. Die Tabelle 4 stellt die finalen Partitionierungen des zweistufigen Klassifizierungsansatzes der beiden Sub-Samples der originären Gruppierung der Datenbank gegenüber. Dabei gilt es anzumerken, dass die Klassifizierung des Datensample 1 (Tabelle 3) auf einer 16X10 und die beiden nachfolgend aufgeführten Partitionierungen der Sub-Samples je mit einer 20X8 Topologie bestimmt wurden.

Im Ergebnis kommt es auch hier für jedes der beiden Sub-Sample zu einer Reduktion der Klassenzahl auf 6 Cluster mit einer im Vergleich zur Partitionierung des Sample 1 sehr ähnlichen Zusammensetzung der Gruppen. Die Gegenüberstellung, der in den Tabellen 3 und 4 dargestellten vorherrschenden selbstdeklarierten Strategien in den ermittelten Klassen der drei Partitionierungen, zeigt ein hohes Maß an Übereinstimmung. Damit können zum einen die Erkenntnisse der oben diskutierten Konsistenzanalyse übernommen und zum anderen die durch den vorgestellten 2-stufigen Klassifizierungsansatz aufgedeckte Struktur von 6 Clustern als stabil im gesamten Untersuchungszeitraum angenommen werden. Diese Feststellung stimmt mit den Ergebnissen von Baghai-Wadji et al. (2006) überein, die ebenfalls zu einer im Zeitablauf stabilen Klassifizierung und Stilkonsistenz bzw. -inkonsistenz der untersuchten Hedge-Fonds-Gruppen kommen. Die Autoren weisen jedoch darauf hin, dass auf Grundlage dieses Ergebnisses nur bedingt Erkenntnisse in Bezug auf den sogenannten „Style Creep“ der Fonds getroffen werden können. Hierzu muss jeder Hedge Fond isoliert betrachtet und analysiert werden, ob sich die auf Basis des 2-stufigen Ansatzes ermittelte Clusterzuordnung zwischen den beiden Sub-Samples geändert hat.

Subsample 1: 31.01.2001 bis 30.06.2005													
Cluster	Bezeichnung	CT	CA	DS	EM	EN	GM	LO	LE	MF	MS	OA	Gesamt
Cl1:	<b>DS</b> & GM & MS	13.51	29.17	<b>75.00</b>	16.25	30.00	<b>30.77</b>	0.00	24.77	15.60	<b>35.71</b>	0.00	248
Cl2:	CA & <b>EN</b>	24.32	<b>41.67</b>	25.00	15.00	<b>55.71</b>	23.08	0.00	9.23	17.43	24.11	50.00	197
Cl3:	<b>CT</b> & <b>MF</b>	<b>54.05</b>	2.08	0.00	1.25	0.00	15.38	0.00	0.45	<b>61.47</b>	3.57	50.00	142
Cl4:	<b>EM</b>	1.80	6.25	0.00	<b>50.00</b>	5.71	12.82	50.00	23.20	3.67	13.39	0.00	177
Cl5:	<b>LE</b>	3.60	4.17	0.00	7.50	5.71	12.82	50.00	<b>28.15</b>	0.00	6.25	0.00	154
Cl6:	Sonstige	2.70	16.67	0.00	10.00	2.86	5.13	0.00	14.19	1.83	16.96	0.00	107
Summe (%)		100	100	100	100	100	100	100	100	100	100	100	
Gesamt <sup>1</sup>		111	48	8	80	70	39	2	444	109	112	2	1025

Subsample 2: 31.07.2005 bis 31.12.2008													
Cluster	Bezeichnung	CT	CA	DS	EM	EN	GM	LO	LE	MF	MS	OA	Gesamt
Cl1:	<b>DS</b> & GM & MS	15.32	25.00	<b>62.50</b>	17.50	27.14	<b>30.77</b>	0.00	24.10	18.35	<b>35.71</b>	0.00	246
Cl2:	CA & <b>EN</b>	24.32	<b>37.50</b>	37.50	13.75	<b>54.29</b>	20.51	0.00	12.39	15.60	16.96	0.00	196
Cl3:	<b>CT</b> & <b>MF</b>	<b>52.25</b>	2.08	0.00	0.00	0.00	15.38	0.00	0.90	<b>58.72</b>	2.68	50.00	137
Cl4:	<b>EM</b>	3.60	4.17	0.00	<b>50.00</b>	5.71	15.38	50.00	22.75	5.50	21.43	0.00	188
Cl5:	<b>LE</b>	0.90	20.83	0.00	7.50	7.14	5.13	50.00	<b>28.83</b>	0.92	8.93	50.00	165
Cl6:	Sonstige	3.60	10.42	0.00	11.25	5.71	12.82	0.00	11.04	0.92	14.29	0.00	93
Summe (%)		100	100	100	100	100	100	100	100	100	100	100	
Gesamt <sup>1</sup>		111	48	8	80	70	39	2	444	109	112	2	1025

Darstellung der Ergebnisse in Anlehnung an die Untersuchung von Baghai-Wadji et al. (2006). Eigene Berechnung.

Die Ergebnisse sind in Prozent angegeben.

<sup>1</sup> Gesamtzahl der Hedge Fonds in der jeweiligen Kategorie.

Tabelle 4: Vergleich des Klassifizierungsansatzes (Zeilen) mit der originären Datenbankgruppierung (Spalten) der beiden Sub-Sample

Diesem Vorgehen folgend fasst Tabelle 5 die Ergebnisse der Style Creep Analyse auf Basis der beiden Sub-Sample zusammen. Wie aus dem oberen Drittel der Tabelle 5 hervorgeht, ist, trotz der relativ hohen prozentualen Übereinstimmungen der Partitionierungen, das Phänomen des Style Creep in der Hedge-Fonds-Branche existent.

Prozentuale Übereinstimmung der Clusterzusammensetzung der Sub Sample						
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Anteil der Übereinstimmung, in %	87.50	79.70	85.21	85.21	83.12	67.29

Style Creep nach originärer Klassifizierung						
	CT	CA	DS	EM	EN	GM
Style Creep, absolut <sup>1</sup>	21	12	1	9	8	18
Style Creep, in % <sup>2</sup>	18.92	25.00	12.50	11.25	11.43	46.15
Gesamt	111	48	8	80	70	39

Style Creep nach originärer Klassifizierung (forts.)						
	LO	LE	MF	MS	OA	Gesamt
Style Creep, absolut <sup>1</sup>	0	66	14	28	1	178
Style Creep, in % <sup>2</sup>	0.00	14.86	12.84	25.00	50.00	17.37
Gesamt	2	444	109	112	2	1025

Style Creep nach SOM-basierter Klassifizierung						
	DS, GM, MS	CA, EN	CT, MF	EM	LE	
Style Creep, absolut <sup>1</sup>	47	20	35	9	66	
Style Creep in % <sup>2</sup>	29.56	16.95	15.91	11.25	14.86	
Style Creep bei richtiger Selbsteinstufung, absolut <sup>3</sup>	9	5	13	3	12	
Style Creep bei richtiger Selbsteinstufung, in % <sup>4</sup>	15.52	8.47	10.24	7.50	9.60	
Anzahl der richtiger Selbsteinstufungen <sup>5</sup>	58	59	127	40	125	
Gesamt	159	118	220	80	444	

Darstellung der Ergebnisse in Anlehnung an die Untersuchung von Baghai-Wadji et al. (2006). Eigene Berechnungen.

<sup>1</sup> Anzahl der Fonds je Kategorie mit Style Creep.

<sup>2</sup> Anteil der Fonds (an allen Fonds in der jeweiligen Kategorie) mit Style Creep.

<sup>3</sup> Anzahl der Fonds mit Style Creep, welche im Untersuchungszeitraum 2001-2005 richtig klassifiziert waren.

<sup>4</sup> Anteil der Fonds (an allen richtig eingestufteten Fonds der jeweiligen Kategorie) mit Style Creep, welche im Untersuchungszeitraum 2001-2005 richtig klassifiziert waren.

<sup>5</sup> Anzahl der Fonds mit richtiger Selbsteinstufung im Zeitraum 2001-2005.

<sup>6</sup> Gesamtzahl der Fonds in der jeweiligen Kategorie.

Tabelle 5: Style Creep

Bei einem Vergleich, der in der Mitte der Tabelle 5 wiedergegebenen Aufschlüsselung dieser Stiländerungen nach der originären Klassifizierung mit den in der Tabelle 4 dokumentierten Ergebnissen der durchgeführten Konsistenzanalyse, deutet sich bereits an, dass die Fonds mit den höchsten Fehleinstufungsquoten auch den höchsten Style Creep aufweisen. Diese Beobachtung wird durch eine tiefergehende Analyse, in der die Stiländerungen differenziert nach Hedge-Fonds mit richtiger und mit falscher Selbsteinstufung unterteilt werden, bestätigt. Im unteren Teil der Tabelle 5 sind diese Ergebnisse zusammengefasst. Offensichtlich kommen im Vergleich zu Fonds mit

richtiger Selbsteinstufung, Stiländerungen bei Fonds mit falscher Selbsteinstufung deutlich häufiger vor. Baghai-Wadji et al. (2006) kommen zu einem ähnlichen Ergebnis, wonach die (ex post) Wahrscheinlichkeit eines Style Creep für Hedge-Fonds mit falscher Selbsteinstufung im Vergleich zu Hedge-Fonds mit richtiger Einstufung mit 23.3% zu 11.7% doppelt so hoch ist. In der vorliegenden Untersuchung beträgt dieses Verhältnis 17.7% zu 10.3%.

Letztendlich bleibt festzuhalten, dass im Beobachtungszeitraum zwar für 17.4% der Hedge-Fonds ein Style Creep festgestellt werden kann, dieses Phänomen jedoch in der Gegenüberstellung mit dem Investmentfondsektor, wo vergleichbare Untersuchungen höhere Stiländerungsquoten von über 55% aufweisen, nicht so stark verbreitet ist.<sup>39</sup> Vor dem Hintergrund dieser Resultate scheint die festgestellte 6-Klassen Partitionierung des Datensamples 1 robust im Beobachtungszeitraum zu sein. Diese Beobachtung wird durch einen weiteren Robustheitstest bekräftigt.

Wie im Abschnitt 3.1 beschrieben wird das 2-stufige Klassifizierungsverfahren zur Überprüfung der Stabilität der gewonnenen Ergebnisse auf das Datensample 2, welches sich durch eine umfassendere Datenbasis von 2789 Hedge-Fonds, einen längeren Untersuchungszeitraum (31.12.199 bis 31.12.2008) sowie unterschiedlich lange Renditehistorien der Fonds auszeichnet, angewendet. In diesem Zusammenhang zeigt sich ein wesentlicher Vorteil des SOM-Algorithmus gegenüber der Clusteranalyse, welche vollständige Daten benötigt, der einen Umgang mit „Missing Values“ in den Inputvektoren und damit die Klassifizierung von Fonds mit unterschiedlicher Renditehistorie ermöglicht.<sup>40</sup> Die Tabelle 6 fasst die Ergebnisse des 2-stufigen Klassifizierungsansatzes (auf Basis einer heuristisch bestimmten 25X11 Topologie) zusammen. Wie die auf einer 16X10 Topologie basierende Klassifizierung des Sample 1 bzw. die auf 20X8 Topologien ermittelten Gruppierungen der Sub-Sample führt auch die auf einer 25X11 Topologie beruhende Partitionierung zu 6 Klassen. Ferner zeigt ein Vergleich zwischen den Partitionierungen der Datensample 1 und 2, dass sowohl die vorherrschenden selbstdeklarierten Strategien der originären Klassifizierung in den ermittelten Gruppen als auch deren Zusammensetzung weitgehend übereinstimmen. Nur bei den GM- und LE-Fonds führen die (geringfügigen) Abweichungen zu einer Verschiebung der vorherrschenden selbstdeklarierten Anlagestile in der berechneten Klassifizierung, was vor dem Hintergrund der Ergebnisse der Konsistenzanalyse mit Fehleinstufungsquoten von bis zu 75% dieser Strategieruppen nicht weiter verwundert.

---

<sup>39</sup> Vgl. Kim et al. (2000), S. 319 und Indro et al. (1998), S. .

<sup>40</sup> Vgl. Kohonen et al. (1995), S. 7 f. und Vesanto et al. (2000), S. 7, sowie die dort angegebene Quelle.

Datensample 2: 31.01.1999 bis 31.12.2008												
Cluster	CT	CA	DS	EM	EN	GM	LO	LE	MF	MS	OA	Gesamt
Cl1	15.09	34.88	<b>62.94</b>	17.47	33.55	27.51	33.33	<b>29.14</b>	19.92	<b>38.86</b>	28.57	797
Cl2	30.06	<b>43.00</b>	37.06	19.65	<b>47.73</b>	<b>27.56</b>	8.33	15.10	17.46	25.83	28.57	621
Cl3	<b>47.49</b>	1.16	0.00	0.87	0.00	12.33	0.00	1.63	<b>57.12</b>	2.59	42.86	267
Cl4	2.45	2.33	0.00	<b>47.60</b>	6.55	10.85	33.33	25.96	1.27	11.99	0.00	533
Cl5	4.29	15.14	0.00	3.93	10.82	10.96	16.67	20.28	2.12	10.62	0.00	378
Cl6	0.62	3.49	0.00	10.48	1.36	10.79	8.33	7.89	2.12	10.10	0.00	193
Summe (%)	100	100	100	100	100	100	100	100	100	100	100	
Gesamt <sup>1</sup>	163	86	17	229	220	146	12	1287	236	386	7	2789

Darstellung der Ergebnisse in Anlehnung an die Untersuchung von Baghai-Wadji et al. (2006). Eigene Berechnung.

Die Ergebnisse sind in Prozent angegeben.

<sup>1</sup> Gesamtzahl der Hedge Fonds in der jeweiligen Kategorie.

Tabelle 6: Vergleich des Klassifizierungsansatzes (Zeilen) mit der originären Datenbankgruppierung (Spalten) des Datensample 2

## 5 Schlussbetrachtung und Ausblick

Gegenwärtig existiert keine allgemein anerkannte Klassifizierung von Hedge-Fonds. Die bestehenden Ansätze sind durch die Datenbankanbieter geprägt und sind, so die Ergebnisse verschiedener empirischer Untersuchungen, weder konsistent noch vergleichbar und weisen ein hohes Maß an Fehlklassifikation auf.

Mit den Self-Organizing Maps behandelt der vorliegende Beitrag eine Methode, welche die Probleme der traditionellen regressionsbasierten Klassifikationsverfahren, die sich im Zusammenhang mit den oft nichtlinearen Rendite-Risikostrukturen von Hedge-Fonds ergeben können, überkommt und eine objektive Taxonomie ermöglicht. Dabei erweitert die Untersuchung mit der Implementierung des 2-stufigen Ansatzes von Vesanto & Alhoniemi (2000) das in der Hedge-Fonds Literatur üblicherweise gewählte problematische Vorgehen der visuellen Auswertung der SOM-Karten, die nicht immer eine exakte Darstellung der entdeckten Strukturen liefern und weitere Interpretationshilfen nötig machen. Darüber hinaus tragen die Anwendungen der Kohonen Regel zur Bestimmung der Lernzyklenanzahl als auch die von Vesanto et al. (2000) dokumentierte Heuristik zur Ermittlung der Topologie zu einer weiteren Objektivierung des Klassifikationsansatzes bei. Nach aktuellem Kenntnisstand existiert keine weitere Veröffentlichung, welche den 2-stufigen Ansatz zur Klassifikation von Hedge-Fonds untersucht und unter Verwendung der genannten Regeln und Heuristiken eine objektive Klassifikation implementiert.

In der empirischen Untersuchung hat die Anwendung des 2-stufigen SOM-basierten Klassifikationsverfahrens im Untersuchungszeitraum vom 31.01.1999 bis zum 31.12.2008 6 Hedge-Fonds-Kategorien ergeben. Diese Reduktion von ursprünglich 11 Klassen der Datenbank sowie die Zusammensetzung der ermittelten Gruppen ist robust in Bezug auf unterschiedlich getestete Topologien und Datensample, die sich hinsichtlich der Untersuchungsperiode sowie der Anzahl an Hedge-Fonds unterscheiden. Auf Grundlage dieser Ergebnisse werden weiterführende Analysen zur Selbsteinstufungsgüte und zur Stiländerungsproblematik durchgeführt. Dabei zeigt sich, dass die Eigenangaben der Hedge-Fonds-Manager überwiegend inkonsistent sind und für 17.37% der untersuchten Hedge-Fonds eine Stiländerung festgestellt werden kann, wobei die ex post Wahrscheinlichkeit eines Style Creep von Hedge-Fonds mit falscher Selbsteinstufung im Vergleich zu Hedge-Fonds mit richtiger Selbsteinstufung mit 17.71% zu 10.27% deutlich höher ist.

Die unter Verwendung des vorgestellten 2-stufigen SOM-basierten Klassifikationsansatzes gewonnenen Erkenntnisse über die „wahre“ Zugehörigkeit der einzelnen Hedge-Fonds sowie deren Anlageverhalten sind aus verschiedenen Gründen wichtig. Sie können beispielsweise zur Vermeidung

dung nicht diversifizierter Portfolios beitragen, was insbesondere bei der Konstruktion von Dachfonds, wo eine Risikostreuung durch Stilmix erzielt wird, von Bedeutung ist. Ferner ermöglichen sie die Konstruktion von Benchmarks, die im Rahmen einer Stilanalyse z.B. für die Performance Attribution oder das Risikomanagement genutzt werden können. Bereits Maillet & Rousset (2001) weisen auf die Möglichkeit hin, unter Verwendung des Kohonen Algorithmus zur Selektion der Fonds Benchmarks bzw. Indices zu extrahieren.<sup>41</sup>

In diesem Zusammenhang können die vorliegenden Ergebnisse dazu beitragen, diversifizierte Portfolios zu bilden. Ferner wird mit dem 2-stufigen Klassifizierungsansatz eine Technik vorgestellt und implementiert, welche zur automatisierten Extraktion von Benchmarks und Stilindices genutzt werden kann.

Die Zuverlässigkeit der empirischen Ergebnisse ist vom verwendeten Datensatz abhängig. In der vorliegenden Untersuchung stehen nur überlebende Hedge-Fonds zur Verfügung. Daher können die gewonnenen Erkenntnisse durch den so genannten „survivorship bias“ durchsetzt sein, was jedoch keinen Einfluß auf die ermittelte 6-Klassen Taxonomie hat.<sup>42</sup> Allerdings können keine Aussagen bezüglich des Verhaltens von „dead“ Hedge-Fonds auf Grundlage der vorliegenden Ergebnisse getroffen werden. Dieses wäre im Rahmen einer weiteren Analyse auf Grundlage einer umfassenderen Datenbasis zu untersuchen.

## Literatur

- Ackermann, C., McEnally, R. & Ravenscraft, D. (1999), ‘The performance of hedge funds: Risk, return, and incentives’, *Journal of Finance* **54**, 833–874.
- Agarwal, V. & Naik, N. (2000), ‘On taking the ‘alternative’ route: risks, rewards style and performance persistence of hedge funds’, *Journal of Alternative Investments* **2**, 6–23.
- Amin, G. & Kat, H. (2003), ‘Welcome to the dark side: Hedge fund attrition and survivorship bias 1994–2001’, *Journal of Alternative Investments* **6**(2), 57–73.
- Baghai-Wadji, R., El-Berry, R., Klocker, S. & Schwaiger, M. (2006), ‘Changing investment styles: Style creep and style gaming in the hedge fund industry’, *Journal of Intelligent Systems in Accounting, Finance and Management* **14**, 157–177.

---

<sup>41</sup> Vgl. Maillet & Rousset (2001), S. 13.

<sup>42</sup> Zum „survivorship bias“ siehe z.B. Amin & Kat (2003).

- Barès, P., R., G. & Gyger, S. (2001), Style consistence and survival probability in the hedge fonds industry, Technical report, Swiss Institute of Banking.
- Bezdek, J. (1998), 'Some new indexes of cluster validity', *IEEE Transactions on Systems, Man and Cybernetics B* **28**, 301–315.
- Blackmore, J. & Miikkulainen, R. (1993), Incremental grid growing: Encoding high-dimensional structure into a two-dimensional feature map, in 'Proceedings of ICNN'93; IEEE International Conference on Neural Networks', Vol. I, Piscataway, NJ, pp. 450–455.
- Bookstaber, R. (2003), 'Hedge fund existential', *Financial Analysts Journal* **59**(5), 19–23.
- Bretzler, M. & Rudolph, D. (2004), *Hedge Fonds und Alternative Investments*, 1 edn, Frankfurt School Verlag, Frankfurt am Main.
- Brown, S. & Goetzmann, W. (1997), 'Mutual fund styles', *Journal of Financial Economics* **43**(3), 373–399.
- Brown, S. & Goetzmann, W. (2003), 'Hedge funds with style', *Journal of Portfolio Management* **29**, 101–112.
- Das, N. & Das, R. (2005), Hedge fund classification technique using self organizing feature map neural network, Technical report, Financial Management Association International.
- Davies, D. & Bouldin, D. (1979), 'A cluster separation measure', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-1**, 224–227.
- Deboeck, G. (1998), Picking mutual funds with self-organizing maps, in G. Deboeck & T. Kohonen, eds, 'Visual Explorations in Finance with Self-Organizing Maps', Springer, London, pp. 39–58.
- Eling, M. (2006), 'Hedgefonds-Strategien - Systematisierung und Überblick', *BankArchiv (ÖBA)* **54**(8), 543–555.
- Fritzke, B. (1994), 'Growing cell structures - a self organizing network for unsupervised and supervised learning', *Neural Network* **7**, 1441–1460.
- Fung, W. & Hsieh, D. (1997), 'Empirical characteristics of dynamic trading strategies: The case of hedge funds', *Review of Financial Studies* **10**(2), 275–302.
- Fung, W. & Hsieh, D. (2001), 'The risk in hedge fund strategies: Theory and evidence from trend followers', *Review of Financial Studies* **14**(2), 313–341.

- Fung, W. & Hsieh, D. (2002), 'Risk in fixed-income hedge fund styles', *Journal of Fixed Income* pp. 1–22.
- Fung, W. & Hsieh, D. (2004), 'Extracting portable alphas from equity long/short hedge funds', *Journal of Investment Management* **2**(4), 1–19.
- Indijc, D. & Heen, A. (2003), 'Aima survey of hedge fund classification practice', *AIMA Journal*.
- Indro, C., Jiang, C., Hu, M. & Wayne, L. (1998), 'Mutual fund performance: A question of style', *Journal of Investing* **7**(2), 46–53.
- Ineichen, A. (2003), *Absolute Returns: The Risk and Opportunities of Hedge Fund Investing*, Wiley, New York.
- Kaiser, D. (2004), *Hedgefonds: Entmystifizierung einer Anlageklasse; Strukturen - Chancen - Risiken*, 1 edn, Gabler, Wiesbaden.
- Kaski, S. & Lagus, K. (1996), Comparing self-organizing maps, *in* von der Malsburg, C. and von Seelen, W. and Vorbrügge, J. and Sendhoff, B., ed., 'Proceedings of ICANN'96, International Conference on Artificial Neural Networks', pp. 809–814.
- Kerling, M. (1997), *Moderne Konzepte in der Finanzanalyse. Markthypothesen, Renditegenerierungsprozesse und Modellierungswerkzeuge*, PhD thesis, University of Bremen, Uhlenbruch.
- Kim, M., Shukla, R. & Tomas, M. (2000), 'Mutual fund objective misclassification', *Journal of Economics and Business* **52**, 309–323.
- Kohonen, T. (1981), Automatic formations of topological maps in a self-organizing system, *in* E. Oja & O. Simula, eds, 'Proceedings of the 2nd Scandinavian Conference on Image Analysis', pp. 214–220.
- Kohonen, T. (1982a), Clustering, taxonomy and topological maps of patterns, *in* 'Proceedings of the sixth International Conference on Pattern Recognition', Silver Spring, MD (IEEE Computer Society), pp. 114–118.
- Kohonen, T. (1982b), 'Self-organized formation of topologically correct feature maps', *Biological Cybernetics* **43**, 59–69.
- Kohonen, T. (1982c), A simple paradigm for the self-organisation of structured feature maps, competition and cooperation in neural nets, *in* S. Amari & M. Arbib, eds, 'Lecture Notes in Biomathematics', Vol. 45, Berlin, pp. 248–266.

- Kohonen, T. (1997), *Self-Organizing Maps*, Springer, Berlin.
- Kohonen, T., Hyninen, J., Kangas, J. & Laaksonen, J. (1995), Som pak the self-organizing map program package, Technical report, Helsinki Univeristy of Technology.
- MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, in 'Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability', Vol. 1, pp. 281–297.
- Maillet, B. & Rousset, P. (2001), Classifying hedge funuds with kohonen maps: A first attempt, Technical report, Social Science Research Network (SSRN).
- Mangiameli, P., Chen, S. & West, D. (1996), 'A comparison of som neural network and hierarchical clustering methods', *European Journal of Operational Research* **93**, 402–417.
- Merkel, D. & Rauber, A. (1997), Alternative ways for cluster visualization in self-organizing maps. in: Proceedings of wsom'97, workshop on self-organizing maps, Technical report, Helsinki University of Technology.
- Milligan, G. & Cooper, M. (1985), 'An examination of procedures for determining the number of clusters on a data set', *Psychometrika* **50**(2), 159–179.
- Mitchell, M. & Pulvino, T. (2001), 'Characterisitcs of risk and return in risk arbitrage', *Journal of Finance* **56**(6), 2135–2175.
- Moutarde, F. & Ultsch, A. (2005), U\*f clustering: A new performant "cluster-mining" method based on segmentation of self-organizing maps, in '5th workshop on self-organizing maps', pp. 25–32.
- Ohlms, C. (2006), Aktives Investmentportfolio-Management: Optimierung von Portfolios aus derivatebasierten dynamischen Investmentstrategien, PhD thesis, Technische Universität Darmstadt, Wiesbaden.
- Poddig, T. & Sidorovitch, I. (2001), Künstlich Neuronale Netze: Überblick, Einsatzmöglichkeiten und Anwendungsprobleme, in H. Hippner, U. Küsters, M. Meyer & K. Wilde, eds, 'Handbuch Data Mining im Marketing', Vieweg, pp. 363–402.
- Poddig, T. & Viebig, J. (2006), 'Hedgefonds-Strategien und Asset-based Style-Faktoren', *Kredit und Kapital* **2**, 281–316.

- Schäfer, D. (2004), Hedge-Fonds - eine gute Anlageform?, Wochenbericht 32, Deutsches Institut für Wirtschaftsforschung.
- Seiler, K. (2008), Phasenmodelle und Investmentstilanalyse von Hedge- und Investmentfonds, PhD thesis, Universität Bremen.
- Sharpe, W. (1992), 'Asset allocation: Management style and performance measurement', *Journal of Portfolio Management* **18**(2), 7–19.
- Signer, A. (2005), Generieren Hedge Funds einen Mehrwert?, in H. Dichtl, J. Kleeberg & C. Schlenger, eds, 'Handbuch Hedge Funds; Chancen, Risiken und Einsatz in der Asset Allocation', Uhlenbruch, Bad Soden/Ts., pp. 211–232.
- Ultsch, A. (1993), Self-organizing neural networks for visualization and classification, in O. Opitz, B. Lausen & R. Klar, eds, 'Information and Classification: Concepts, Methods and Applications', Springer, Berlin, pp. 307–313.
- Ultsch, A. (2003), U\*-matrix: A tool to visualize clusters in high dimensional data, Technical Report 36, University of Marburg.
- Ultsch, A. & Vetter, C. (1994), Self-organizing-maps versus statistical clustering methods: A benchmark, Report 0994, FG Neuroinformatik und Künstliche Intelligenz, Universität Marburg.
- Vesanto, J. & Alhoniemi, E. (2000), 'Clustering of self-organizing map', *IEEE Transactions on Neural Networks* **11**(3), 586–600.
- Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankangas, J. (2000), Som toolbox for matlab 5, Technical Report Report A57, Helsinki University of Technology.
- Zimmermann, H. (1994), Neuronale Netze als Entscheidungskalkül, in H. Rehkugler & H. Zimmermann, eds, 'Neuronale Netze in der Ökonomie', Frankfurt School Verlag, pp. 1–87.

## Appendix

Cluster	Mean	Std	Kurt.	Schiefe	JB-	White-	ADF-Test	Autokorrelation			Ljung-Box-Q-Test		
	%	%			Test	Test	I(d)	Lag1	Lag6	Lag12	Q1	Q6	Q12
CT	0.194	2.535	5.823	0.781	36.426**	3.843**	1	0.110	0.074	-0.064	1.061	5.916	11.222
CA	-0.186	1.960	2.893	-0.191	0.550	5.026**	1	0.063	0.066	0.035	0.349	1.553	4.313
DS	-0.693	3.185	2.504	0.216	1.512	0.847	1	0.403	0.048	-0.015	14.160**	18.109**	23.095*
EM	0.476	3.512	2.991	-0.690	6.665*	1.952	1	0.286	0.182	-0.092	7.129**	24.851**	29.559**
EN	0.002	1.948	4.794	-0.602	16.346**	1.730	1	0.089	0.087	0.038	0.684	3.311	7.438
GM	0.254	2.292	5.007	0.520	17.891**	1.502	1	0.017	0.088	-0.139	0.026	1.584	7.214
LE	0.043	2.270	2.880	-0.497	3.505	0.925	1	0.109	0.147	-0.041	1.034	7.756	9.255
LO	-0.273	5.429	6.818	-0.027	51.037**	0.210	1	0.226	0.020	-0.126	4.453*	8.380	17.140
MF	0.362	2.757	5.259	0.647	23.709**	3.889**	1	0.149	0.038	-0.005	1.934	12.041	16.499
MS	0.024	1.753	4.011	-0.756	11.579**	0.712	1	0.054	0.196	-0.041	0.255	12.038	13.424
OA	0.013	3.469	3.126	0.511	3.716	0.537	1	0.079	-0.020	-0.096	0.540	14.095	18.975

Korrelationsmatrix											
	CT	CA	DS	EM	EN	GM	LE	LO	MF	MS	
CA	0.399										
DS	0.226	-0.039									
EM	0.145	0.672	-0.253								
EN	0.627	0.795	0.259	0.471							
GM	0.722	0.716	0.078	0.551	0.785						
LE	0.390	0.840	-0.135	0.853	0.752	0.772					
LO	-0.157	0.449	-0.280	0.688	0.168	0.259	0.635				
MF	0.937	0.299	0.149	0.113	0.503	0.653	0.312	-0.200			
MS	0.458	0.834	-0.105	0.853	0.765	0.794	0.957	0.558	0.383		
OA	0.659	0.322	0.084	0.145	0.358	0.521	0.277	-0.219	0.722	0.335	

\* zeigt die Singnifikanz auf dem 1%- und \*\* auf dem 5% Niveau an; Jarque-Bera-Statistik (JB) testet die Nullhypothese (H0), dass die Daten normalverteilt sind; White-Statistik testet die H0, dass die Daten homoskedastisch sind; Augmented Dickey Fuller (ADF) Statistik testet die H0, dass die Daten nicht-stationär sind; I(d) gibt den Integrationsgrad an; Ljung-Box Q Statistik testet die H0, dass keine Autokorrelation bis zur Ordnung k vorhanden ist

Tabelle 7: Deskriptive Statistik der Clusterzentroide der originären Klassifizierung

Cluster	Mean	Std	Kurt.	Schiefe	JB-	White-	ADF-Test	Autokorrelation			Ljung-Box-Q-Test		
	%	%			Test	Test	I(d)	Lag1	Lag6	Lag12	Q1	Q6	Q12
C11	0.230	0.681	3.275	-0.202	0.834	2.068	1	0.676	0.303	-0.113	39.794**	142.099**	146.786**
C12	-0.043	2.540	4.794	-0.024	11.269**	0.260	1	0.088	0.035	-0.004	0.673	2.919	6.497
C13	0.445	3.861	4.139	0.409	6.885*	3.702	1	0.082	-0.026	-0.005	0.592	9.216	17.422
C14	0.090	4.170	2.806	-0.717	7.331*	1.240	1	0.242	0.110	-0.084	5.100*	16.333*	20.557
C15	-0.204	3.003	2.996	-0.359	1.802	1.450	1	0.068	0.062	-0.048	0.397	1.551	2.540
C16	0.423	2.552	4.267	-0.626	11.098**	0.582	1	-0.005	0.051	-0.020	0.002	2.078	3.917

Korrelationsmatrix					
	C11	C12	C13	C14	C15
C12	0.125				
C13	0.051	0.356			
C14	0.361	0.325	0.057		
C15	0.211	0.820	0.244	0.748	
C16	0.417	0.678	0.461	0.720	0.764

\* zeigt die Singnifikanz auf dem 1%- und \*\* auf dem 5% Niveau an; Jarque-Bera-Statistik (JB) testet die Nullhypothese (H0), dass die Daten normalverteilt sind; White-Statistik testet die H0, dass die Daten homoskedastisch sind; Augmented Dickey Fuller (ADF) Statistik testet die H0, dass die Daten nicht-stationär sind; I(d) gibt den Integrationsgrad an; Ljung-Box Q Statistik testet die H0, dass keine Autokorrelation bis zur Ordnung k vorhanden ist

Tabelle 8: Deskriptive Statistik der ermittelten Clusterzentroide