

JMP 9 Highlights

Bernd Heinen
SAS Institute GmbH
In der Neckarhelle 162
69118 Heidelberg
Bernd.heinen@jmp.com

Zusammenfassung

Seit Oktober 2010 ist die neue Version JMP 9 erhältlich. Gleichzeitig beginnt auch eine neue Produktstrategie, mit der die Leistungen von JMP anwendungsorientiert oder technisch differenziert werden und so den Bedürfnissen der unterschiedlichen Nutzergruppen näher kommen. Dieser Vortrag bezieht sich auf JMP 9, Abweichungen zu anderen Produkten werden nur kurz angerissen. Mit der Degradation Plattform und den vollständig überarbeiteten Neuronalen Netzen fallen die wichtigsten Änderungen gleich ins Auge. Aber auch die Möglichkeit, geografische Karten als Hintergrund in Grafiken einzublenden und statistische Kennzahlen mit regionalem Bezug darzustellen, unterstreicht das Ziel, die visuelle Datenanalyse weiter voranzutreiben. Schnittstellen, über die EXCEL und R in die Funktionalität von JMP eingebunden werden, erweitern die analytische Tiefe und den Umfang der Anwendungsbereiche.

Schlüsselwörter: Degradation, Neuronale Netze, Karten, JMP 9

1 Auf den ersten Blick

JMP 9 erscheint mit einer neuen Oberfläche (s. Abbildung 1). Das Startfenster, jede Datentabelle und die Auswertungen erscheinen in separaten Fenstern. Icons in den Fußleisten der Fenster sowie die Tastenkombination Strg+Tab erlauben schnelles Navigieren zwischen den einzelnen Objekten.

2 Plattformen

Die statistischen Analysemethoden werden in JMP thematisch zu Gruppen zusammengefasst, die „Plattformen“ genannt werden. In der Version 9 sind zwei neue Plattformen enthalten: Degradation und Neuronale Netze.

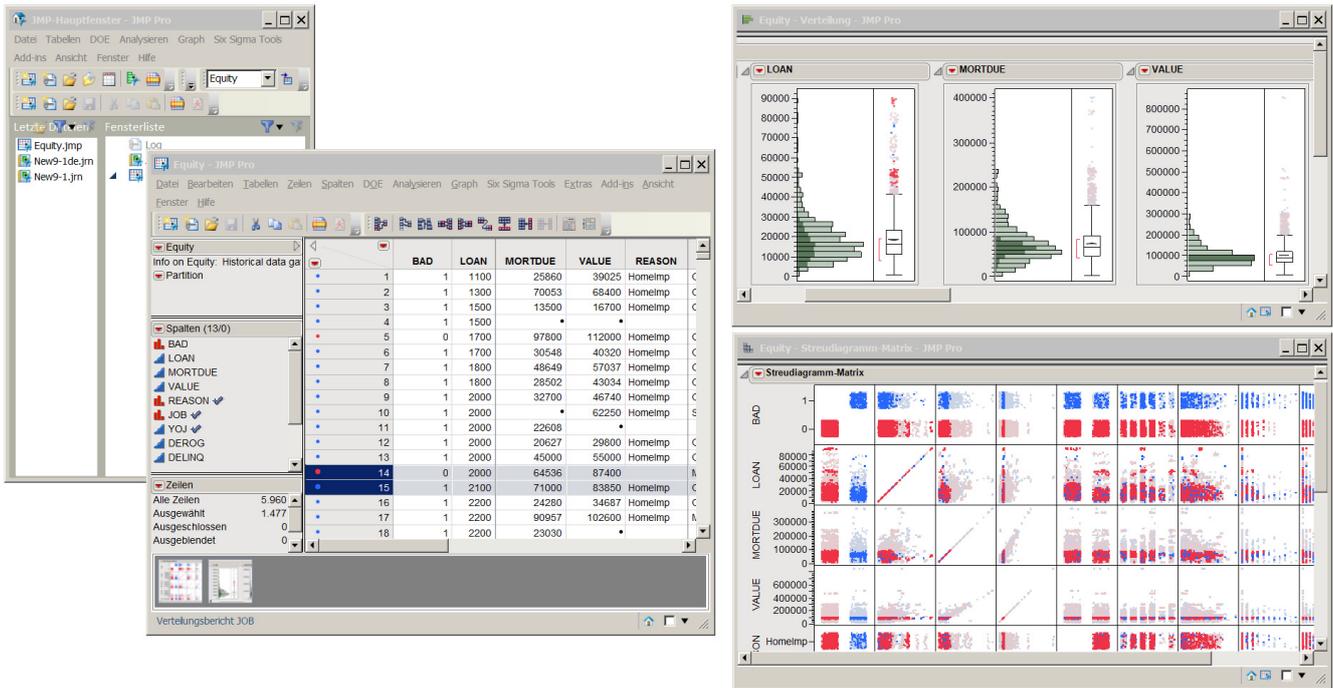


Abbildung 1: Neue Fensteroberfläche

2.1 Degradation

Bei der Analyse von Lebensdauern ist es oft so, dass das Zielereignis nicht abrupt eintritt, eventuell gar nicht beobachtet wird, aber eine messbare Funktion sich über die Zeit

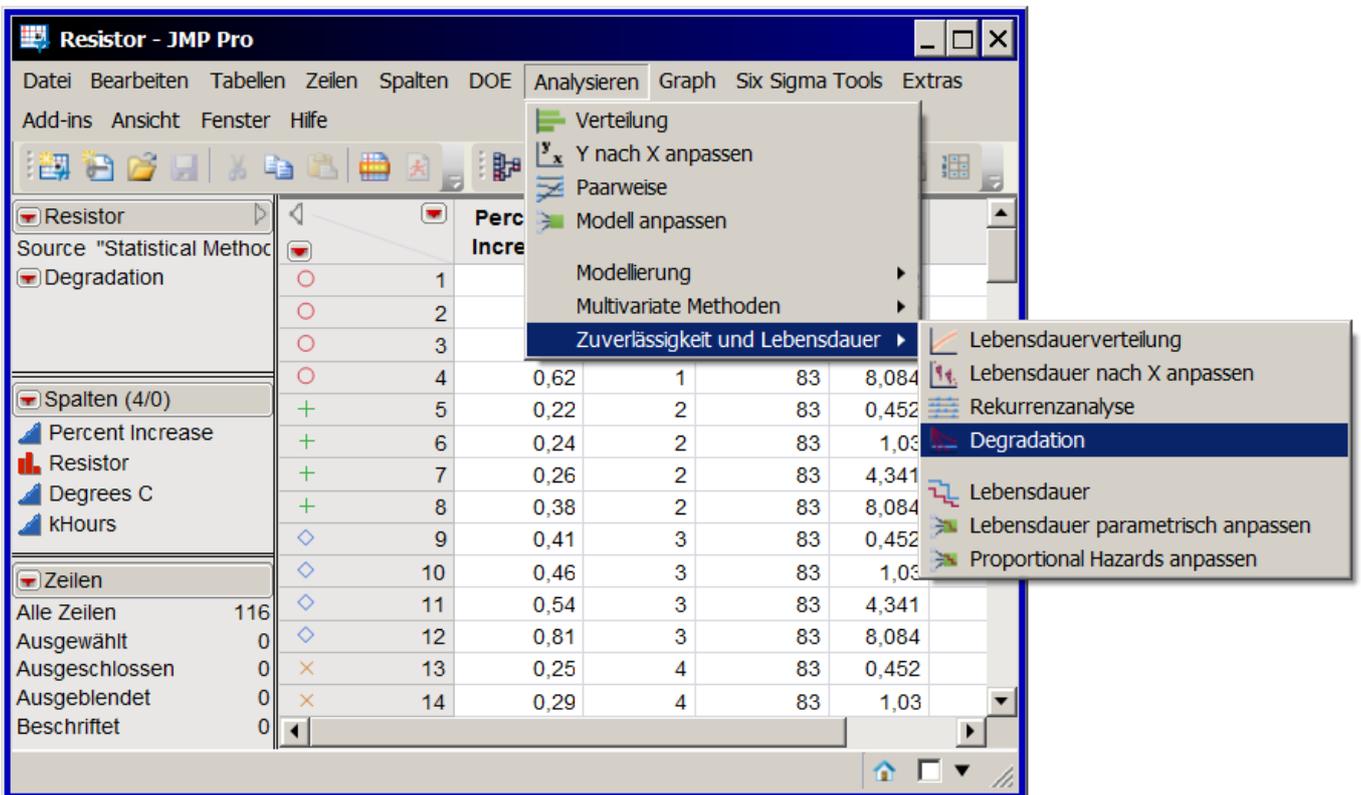


Abbildung 2: Aufruf der Degradation Plattform

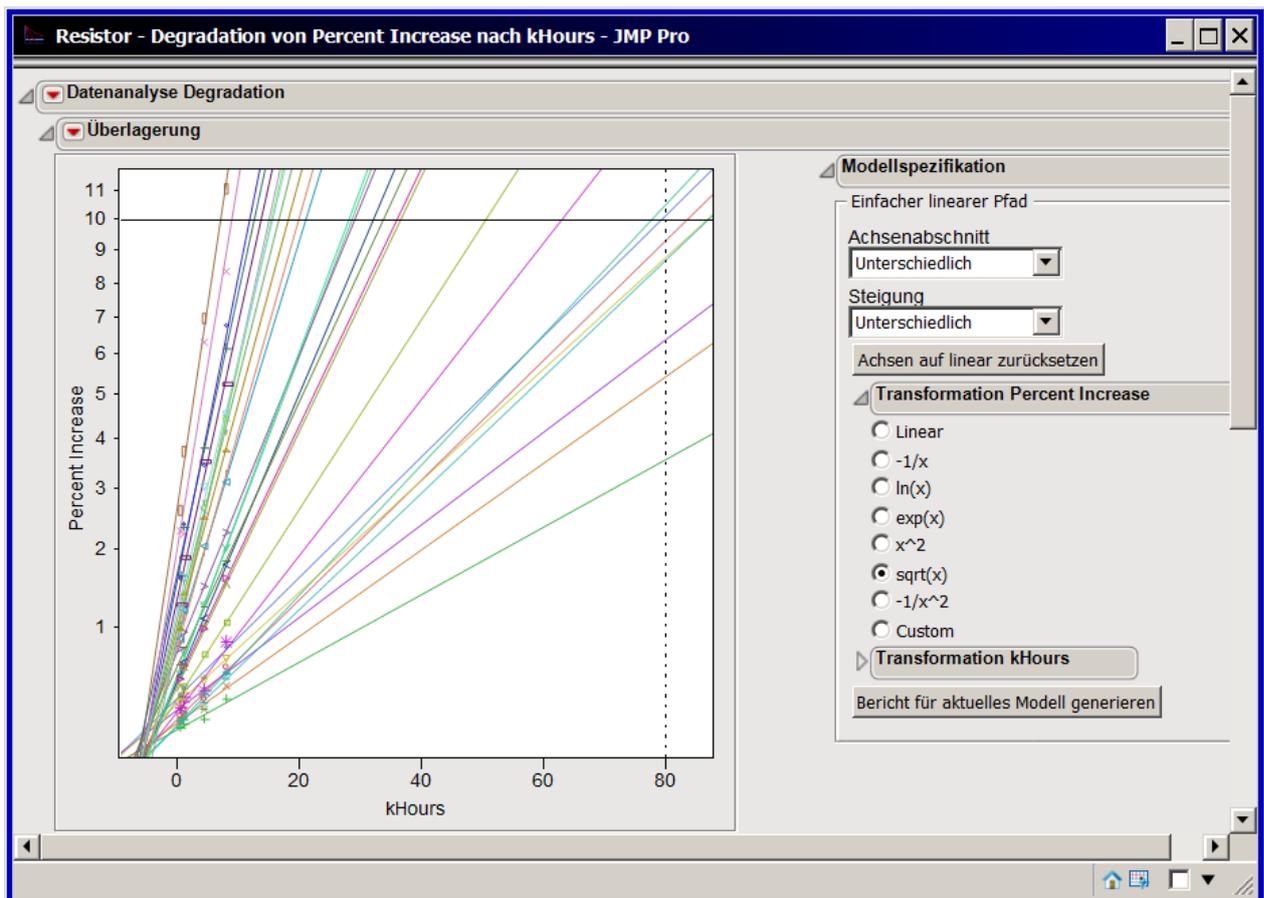
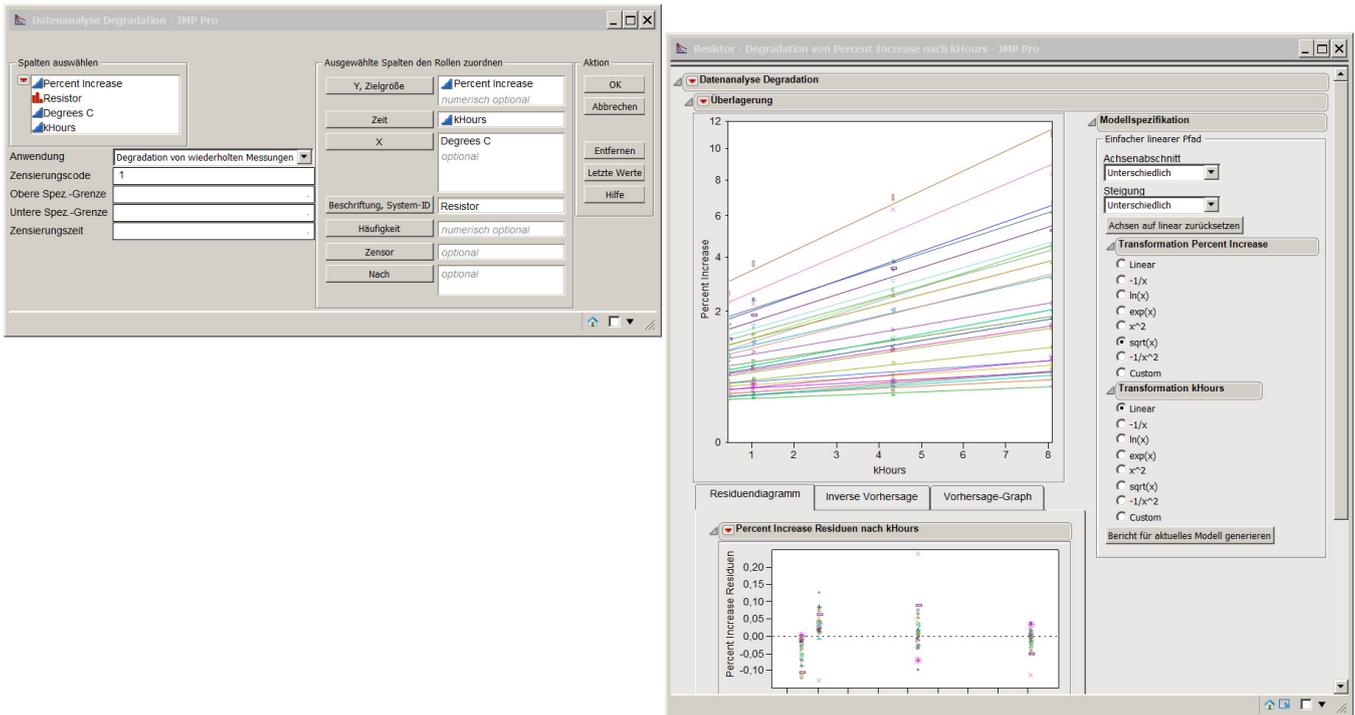
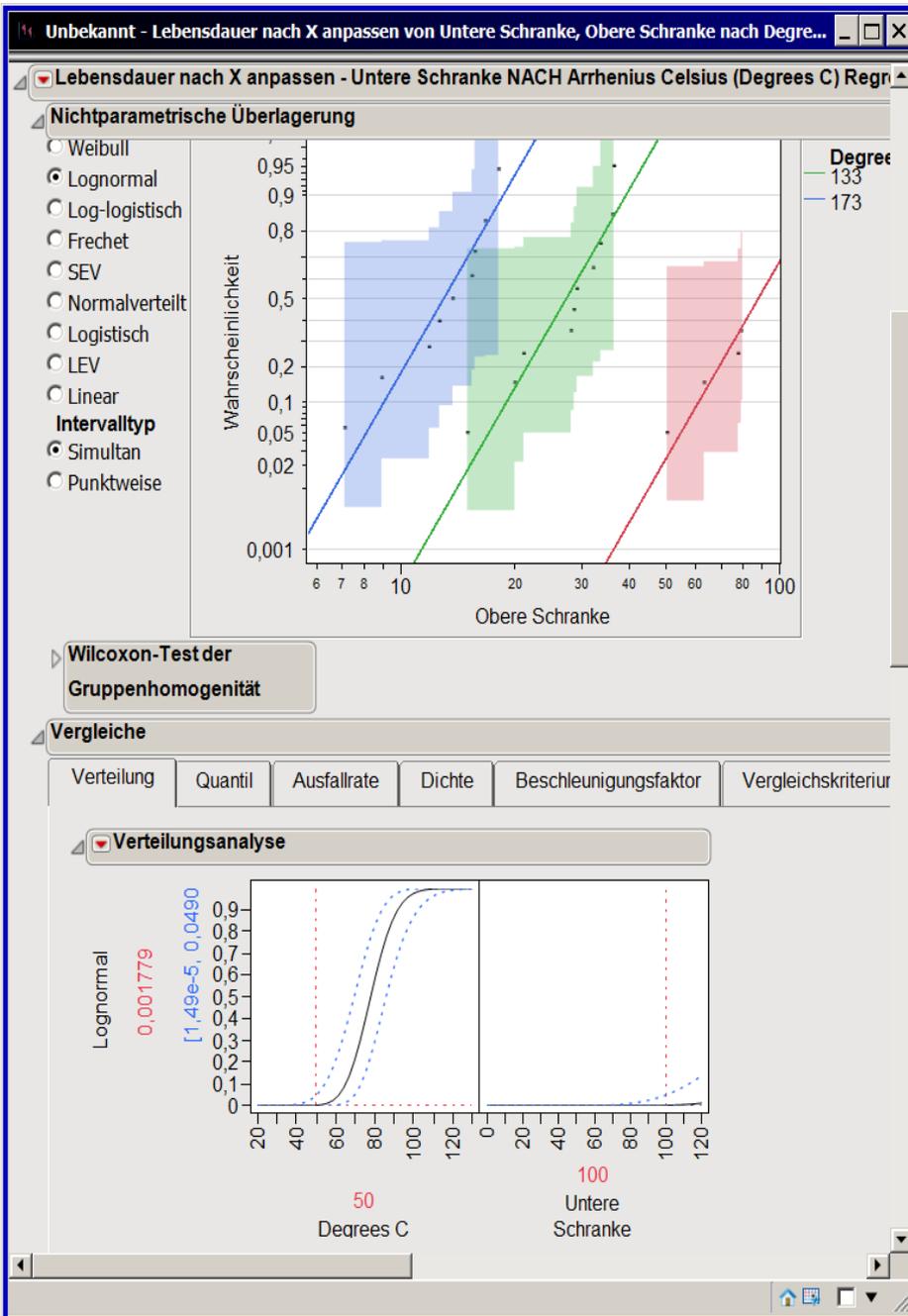


Abbildung 3: Anwendung und Ergebnisse Degradation Analyse
 Der Modelldialog (oben), die angepassten Geraden nach einer Wurzeltransformation und die Residuen (rechts). Durch Spezifikation des Zielereignisses (10% Erhöhung des Widerstands) und eines Zensierungszeitpunktes (80.000 h) werden die Prognosefunktionen für diesen Bereich angezeigt (unten).

verändert. In dieser Plattform (für den Aufruf siehe Abbildung 2) sind nun alle Funktionen zusammengefasst, die zur Analyse und Modellbildung in dieser Situation erforderlich sind.



Hierzu ein Beispiel: Ein elektrischer Widerstand wird als defekt bezeichnet, wenn sein Widerstand 10% vom Sollwert abweicht. Da man nicht abwarten kann, bis alle zu testenden Widerstände tatsächlich so weit von der Norm abweichen, berechnet man ein Prognosemodell für die Lebensdauer (Abb. 3).

Indem die Schätzungen für die Ausfallrate und den Zensierungszeitpunkt in eine Tabelle gespeichert werden, können über die Regression für Überlebenskurven eine vertiefte Diagnose und der bekannte JMP Profiler ausgeführt werden, der es erlaubt, verschiedene Prognosekombinationen durchzuspielen (Abb. 4).

Abbildung 4: Profiler für Degradation Modelle

2.2 Neuronale Netze

Mit neuronalen Netzen wird man weniger die Struktur und Abhängigkeit von Variablen studieren, als vielmehr Prognosemodelle entwickeln. In einer Studie wurden Pilze in essbare und nicht essbare eingeteilt und anhand von 22 Faktoren beschrieben. Eine schrittweise Partitionierung zeigt, dass nach drei Teilungsschritten eine fast vollständige Unterscheidung in essbare und nicht essbare Pilze erfolgt (Abb. 5).

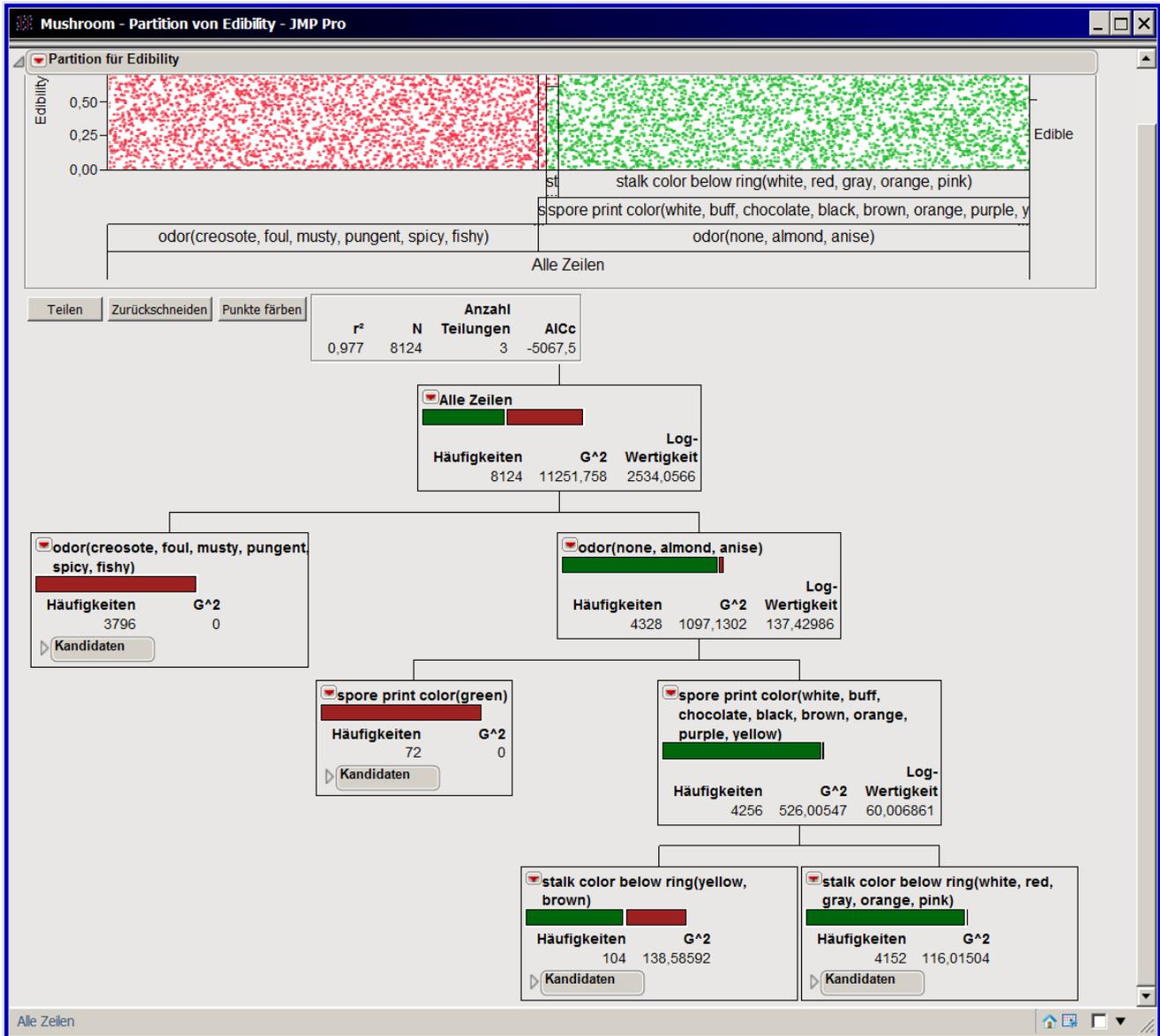
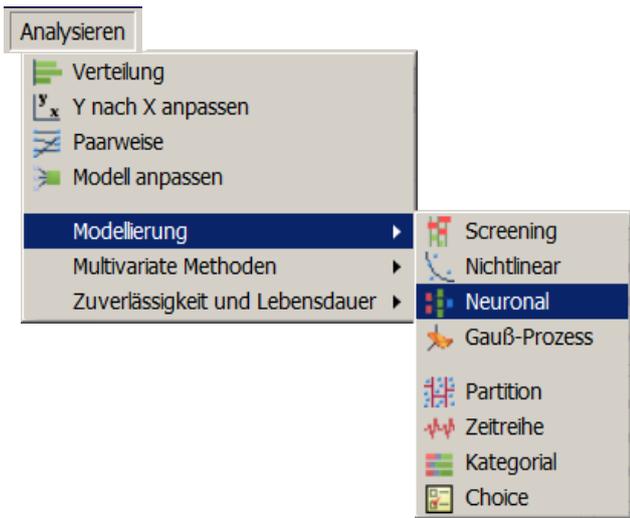


Abbildung 5: Partition



Für die zukünftige Klassifizierung von Pilzen soll ein Neuronales Netz angepasst werden.

Im Plattformdialog (Aufruf Abb. 6, Abb. 7) werden die Variablen und ihre Funktionen zugewiesen, wobei eine Validierungsvariable nur in JMP Pro zugewiesen werden kann: Der Folgedialog bietet dann die Möglichkeit, ein Neuronales Netz zu spezifizieren und berechnen zu lassen.

Abbildung 6: Aufruf Neuronale Netze

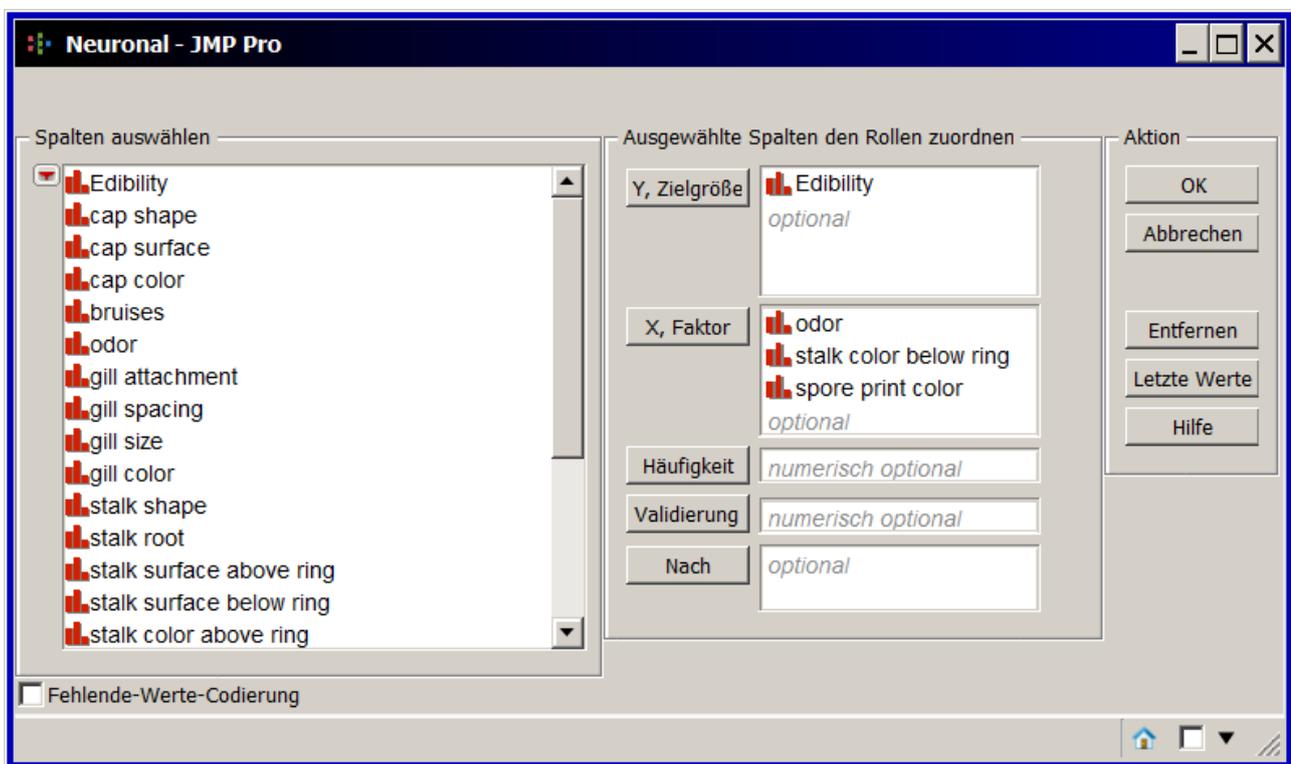


Abbildung 7: Dialog Neuronale Plattform

Hierbei unterscheiden sich JMP und JMP Pro: in der Basisversion ist nur eine Schicht von Neuronen modellierbar, in JMP Pro stehen zwei Schichten zur Verfügung. Beide Programme bieten drei Linkfunktionen an, die Zahl der Neuronen pro Schicht ist beliebig. Darüber hinaus bietet JMP Pro die Technik des Boosting zur Modellbildung an, bei der das spezifizierte Modell iterativ auf die Residuen der jeweils vorher geschätzten Modelle angewandt wird.

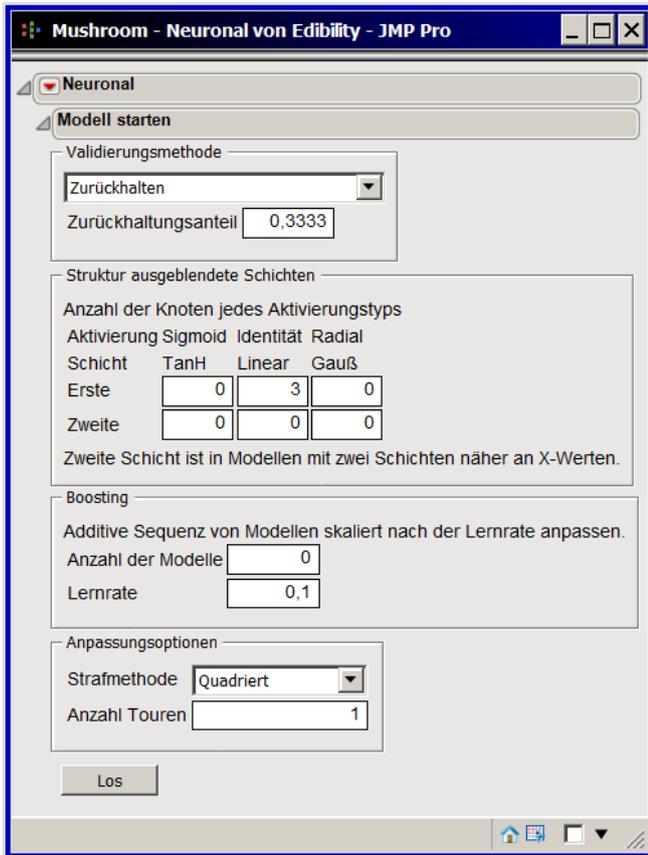


Abbildung 8: Modellbildung Neuronale Netze

Neuronale Netze tendieren zum „Overfitting“, d.h. sie passen sich perfekt an den aktuellen Datensatz an, haben aber eine geringe prognostische Güte, wenn sie mit neuen Daten berechnet werden. Um dies zu verhindern, wird ein Teil der Daten zur Validierung der Prognosestärke zurückbehalten. Ihr Anteil kann beliebig oberhalb von 10% gewählt werden. Die übrigen Sätze bilden den Trainingsdatensatz, an dem das Modell entwickelt wird. Mit Hilfe des Validierungsdatensatzes wird berechnet, wann der Anpassungsprozess endet.

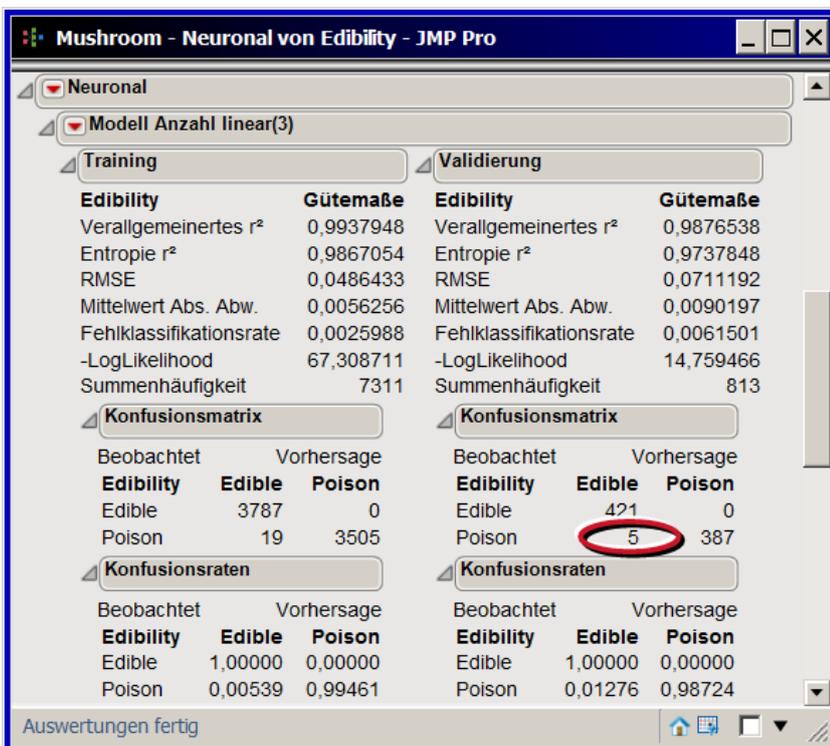


Abbildung 9: Anpassungsgüte Neuronaler Netze

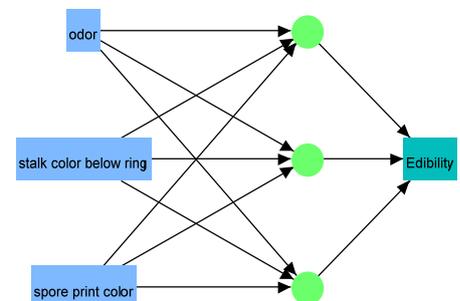


Abbildung 10: Neuronales Netz

Der Ergebnisbericht zeigt verschiedene Gütemaße für beide Datensätze und auch die Konfusionsmatrizen, die hier zeigen, dass für das Prognosemodell immerhin noch 1,3% der Pilze als essbar eingestuft werden, obwohl sie giftig sind (Abb. 9). Weitere Ergebnisdarstellungen lassen sich über den „Hot Spot“, das kleine rote Dreieck in der Titelzeile des Reports, auswählen; unter anderem eine Grafik, die das gewählte Neuronale Netz darstellt (Abb. 10). Der Modellierungsdialog bleibt im oberen Teil des Berichts ständig zur Verfügung, so dass man mehrere Modelle durchrechnen und innerhalb desselben Reports vergleichen kann. Wie bei allen anderen Analyseplattformen kann man die Schätzer wieder in den Datensatz speichern, wo sie dann für weitere Analysen zur Verfügung stehen.

2.3 Grafiken

Der Graph Builder ist mit seinen interaktiven Gestaltungsmöglichkeiten um zwei Bereiche erweitert worden: die Farbsteuerung und das Formfeld.

Wenn eine Variable zur Farbsteuerung eingesetzt wird, dann werden die Datenpunkte entsprechend den Ausprägungen dieser Variable farbig dargestellt.

Das Formfeld ermöglicht, einen räumlichen Bezug herzustellen. Wird hier z.B. eine Variable mit Namen der Bundesländer hergezogen, dann erscheinen die Zielvariablen (Abb. 11) mit entsprechender geografischer Darstellung. Auch in dieser Darstellung bleibt die Linkfunktion von JMP erhalten. Ein ausgewählter Datensatz (Bayern 2005) wird in der Datentabelle und allen anderen Darstellungen zu der Tabelle markiert.

Weiterhin steht jetzt auch die Möglichkeit zur Verfügung, mit dem Graph Builder Konturdiagramme zu erstellen, mit beliebiger Zahl der Konturstufen und freier Farbwahl (Abb. 12). Wird eine Häufigkeitsvariable angegeben, werden alle verfügbaren Statistiken ebenfalls entsprechend dieser Häufigkeiten gewichtet. Für den Mittelwert stehen verschiedene Fehlerbalken (Standardabweichung, Standardfehler, Bereich, Konfidenzintervall) zur Verfügung.

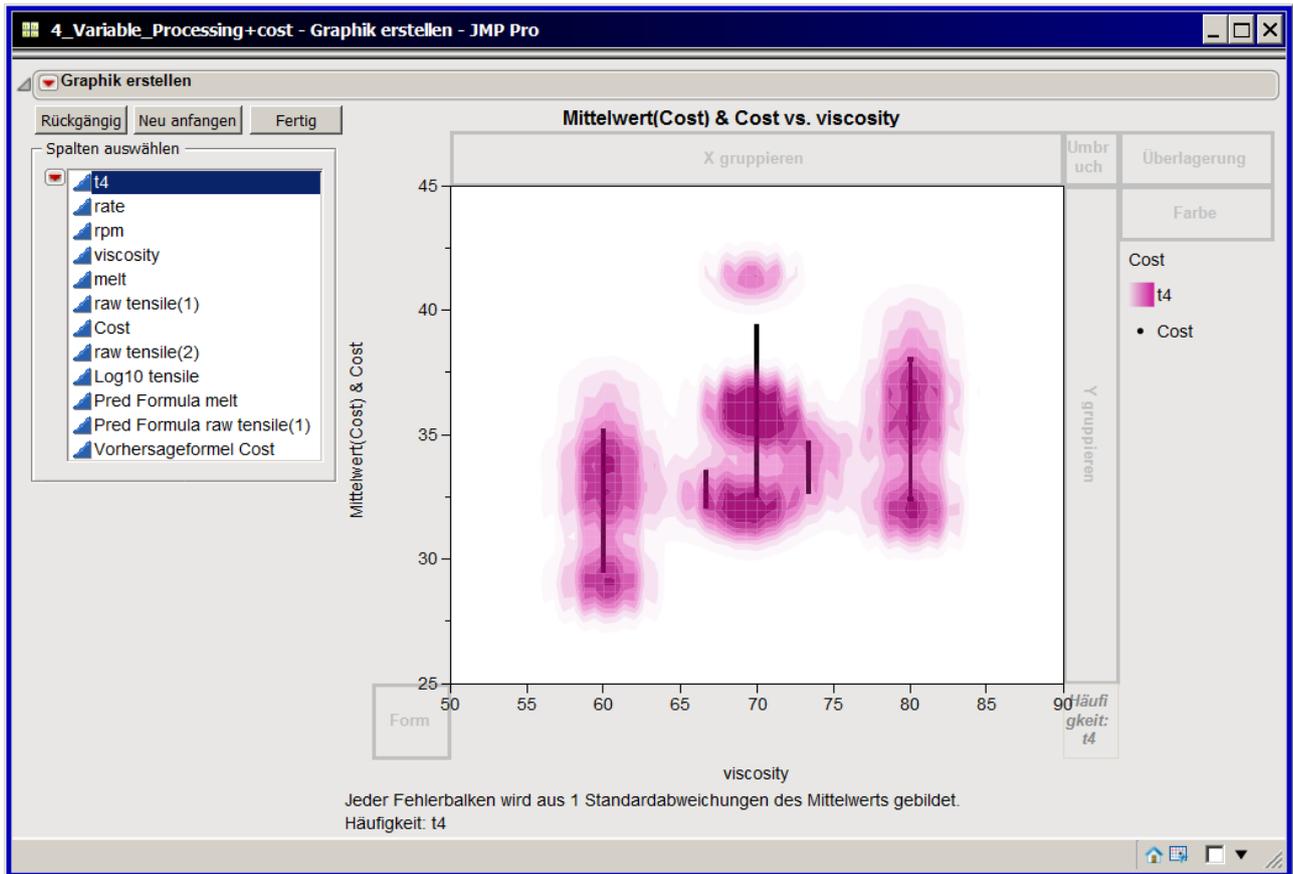


Abbildung 11: Konturplot mit 10 Stufen und Bereichsmarkierung

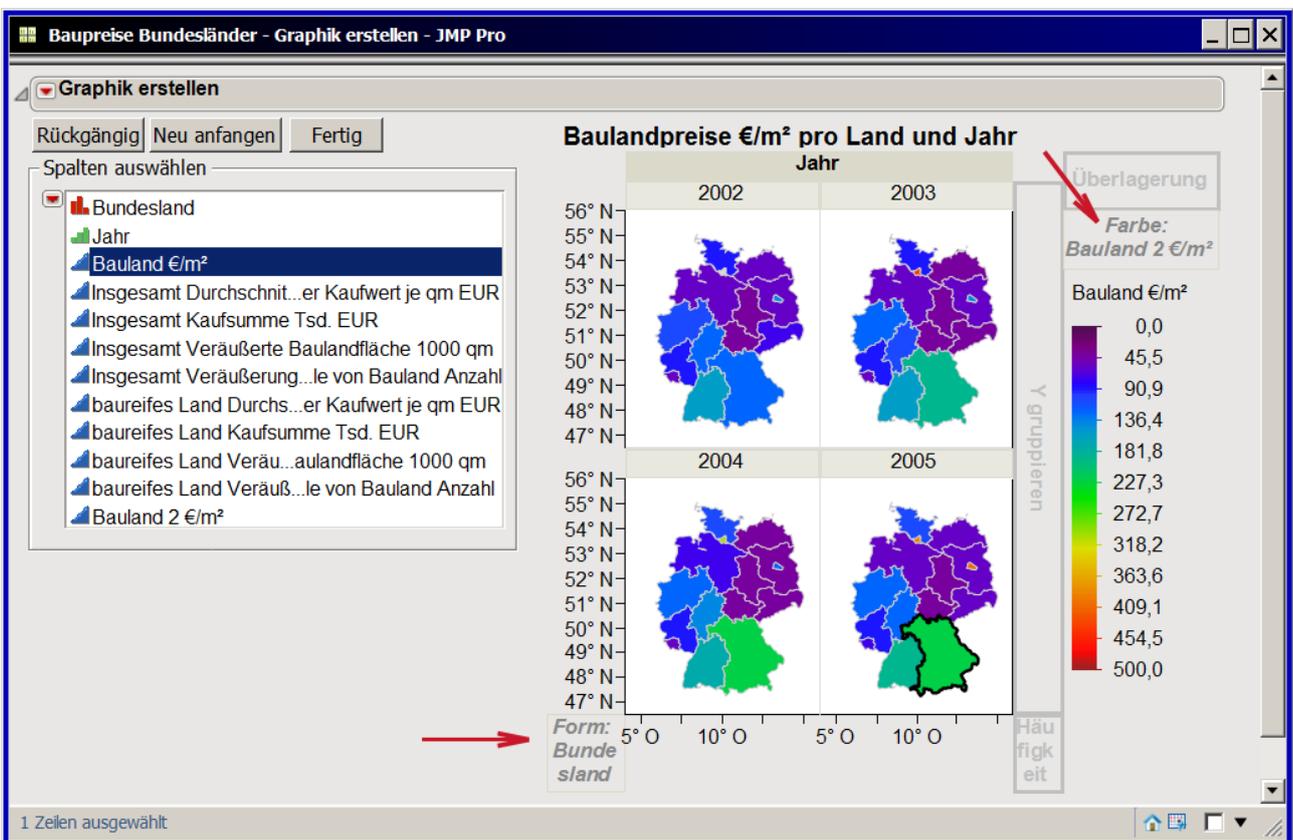


Abbildung 12: Geographische Einheiten als Formfaktor

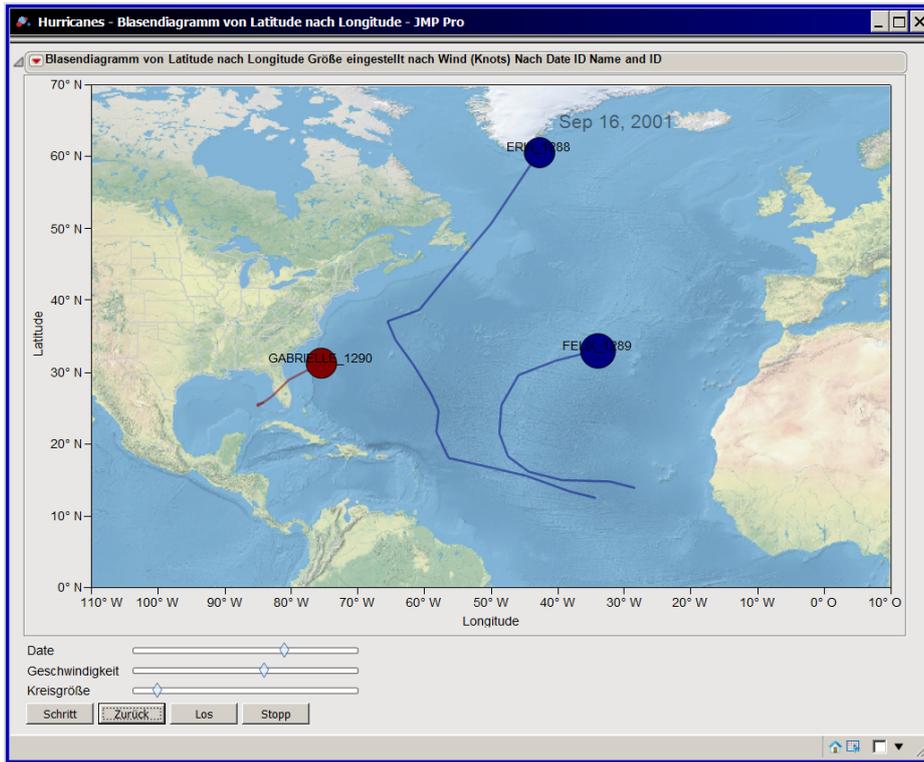


Abbildung 13: Bubble Plot mit Karte als Hintergrund

Das Blasendiagramm (Bubble Plot) kann nun auch mit geografischem Bezug dargestellt werden (Abb. 13). Diese und andere animierte Grafiken lassen sich als Objekte für die Adobe Flash Software speichern, so dass sie auch mit einem Browser betrachtet werden können. Dabei bleiben die Animation und die Verlinkung innerhalb einer Grafik erhalten.

2.4 JMP als Analyseschaltelpult

In JMP 9 sind drei Neuerungen erhalten, die ermöglichen, JMP als eine zentrale Drehscheibe für eine Vielzahl an Erweiterungen einzusetzen. Dabei handelt es sich um:

- Add-Ins zum einfachen Verteilen und Installieren von (JMP-)Scripten
- EXCEL Schnittstelle zum Erstellen von interaktiven JMP Grafiken direkt aus EXCEL und zur Nutzung des JMP Profilers für Optimierung und Simulation in EXCEL Modellen.
- R-Schnittstelle zur Übergabe von Daten aus JMP an R und Rückgabe der Analyseergebnisse an JMP zur Speicherung und grafischen Darstellung.

2.5 Weiteres

Die hier vorgestellten Änderungen zeichnen sich durch ihre besondere Bedeutung für die Datenanalyse oder als besondere Innovation aus. Es gibt viele weitere Ergänzungen, die quer über alle Plattformen den Funktionalitätsumfang von JMP erweitern. Zusätzliche Statistiken, verbesserte Grafiken, erweiterte Interaktivität und flexiblere Darstellungsmöglichkeiten erhöhen den Gebrauchswert von JMP in vielen Bereichen. Weitere Informationen finden Sie unter www.jmp.com/de.