

## ZUSAMMENFASSUNG

---

Verallgemeinerte additive Modelle (GAM) bieten eine breite Palette an Modellen für den Umgang mit nichtlinearen Zusammenhängen. Zur Modell-schätzung von GAMs in R steht das von Simon Wood entwickelte `mgcv` Paket zur Verfügung. Die Standardwerte für die Parameter der `mgcv::gam()` Funktion zu verwenden, ist jedoch nicht immer ausreichend. Viele Modellierungsprobleme benötigen manuelle Anpassungen des Modells und erfordern deshalb vom Nutzer ein tieferes Verständnis für die Methoden. Der Schwerpunkt dieser Masterarbeit liegt auf der Untersuchung von gängigen Fragestellungen, die während der Anpassung eines GAMs auftreten, der Entwicklung von Entscheidungskriterien für eine geeignete Modell- und Parameterwahl und der Implementierung des `pagam` Pakets in R (Partially Automated Fitting of GAMs). Das Paket bietet mit den neuen Funktionen eine automatisierte Unterstützung bei der Modell- und Parameterwahl. Die wichtigsten Beiträge sind dabei die Identifizierung der zu verwendenden Residuenverteilung, die Auswahl geeigneter Basisdimensionen der Glättungsterme, der Umgang mit Concurvity, die Durchführung einer Variablenselektion, die Berücksichtigung von Variablenrestriktionen und die Erleichterung des Umgangs mit Hierarchien in den Daten. Einige Funktionen werden anhand zweier Datensätze getestet. In beiden Fällen ist erkennbar, dass die neuen Funktionen im Vergleich zur `mgcv::gam()` Funktion mit ihren Standardwerten zu kleinen Verbesserungen in der Modellkomplexität und der Anpassungsgüte führen.

**Schlagwörter:** Verallgemeinerte Lineare Modelle, Verallgemeinerte Additive Modelle, Glättungsverfahren, Splines, Automatisierung, Modellspezifikationen, Variablenselektion, Concurvity, Variablenrestriktionen

## ABSTRACT

---

Generalized additive models (GAM) offer a wide range of modeling tools for dealing with non-linear relationships. For the estimation of GAMs in R, the `mgcv` package developed by Simon Wood is available. However, the default values for some of the parameter in the `mgcv::gam()` function are usually not adequate. Many modeling problems need manual adjustments of the model and its parameter settings and thus require a deep understanding of the methods from the user. In the context of generalized additive models, the focus of this research is to explore common challenges during the fitting process, establish decision criteria for model specifications and develop the `pagam` package in R (Partially Automated Fitting of GAMs). The package provides a user-friendly supporting environment for model selection and choice of parameter values. In detail, the main contributions are: Identifying error distribution to use in fitting, choosing appropriate basis dimensions of smooth terms, dealing with concurvity, performing variable selection, considering variable constraints and facilitating hierarchical data management. Example applications are presented for two different datasets. In both cases some of the functions lead to small improvements of model complexity and goodness of fit compared to the `mgcv::gam()` function with its default values.

**Keywords:** Generalized Linear Models, Generalized Additive Models, Splines, Smoothing, Automation, Model Specification, Variable Selection, Concurvity, Variable Restrictions