

Modellwahl und Parameterschätzung in der nichtlinearen Regressionsanalyse für ornithologische Problemstellungen

KLAUS-DIETER FEIGE

Abstract: Growth models are preferable to quasi-linear regression polynomials in many questions. The effective application of this requires a good knowledge of the properties of the model. Using practical examples, the possibilities for model selection and initial value estimation of the model parameters are presented. The characterization various, frequently used models of growth analysis is intended to indicate model-specific features.

Key words: growth models, nonlinear regression, growth process, die-off process, time series, parameter estimation, Excel solver method, initial value estimation

1. Problemsituation

Wachstumsfunktionen dienen der modellhaften Analyse des zeitabhängigen Wachstums von Organismen und Populationen und spielen nicht nur in den biologischen Wissenschaften eine zunehmend größere Rolle. Sie dienen einerseits der zeitabhängigen Darstellung eines Zielmerkmals an sich, ermöglichen darüber hinaus aber auch anhand konkret berechneter Modellparameter, Tendenzen und Grenzwerte von Entwicklungen vorherzusagen oder Entscheidungen vorzubereiten. In der Biologie beziehen sich die entsprechenden Analysen immer auf eine Population oder Beziehungen zwischen mehreren Populationen.

Im Unterschied zur quasilinearen Regressionsanalyse zeichnet sich die **Eigentlich Nichtlineare Regression** durch eine in der Regel leicht nachvollziehbare Interpretation der Modellparameter sowie oft auch durch eine geringere Parameterzahl aus. Diese Vorteile bestehen aber nur dann, wenn der Nutzer der Analyseverfahren eine begründete Übereinstimmung zwischen dem betrachteten biologischen Prozess und dem abstrakten mathematischem Modell annehmen kann. Eine derartige Voraussetzung ist nicht immer einzuhalten, doch es sollte in jedem Fall ein Modell gewählt werden, das zumindest in seinen Grundzügen nicht sofort dem biologischen Sachverhalt widerspricht. Diese Form der Modellierung ist meist auch den GAM-Verfahren vorzuziehen (FAHRMEIR et al. 2013).

Um mit den Regressionsmodellen effektiv arbeiten zu können, muss der Anwender die speziellen Eigenschaften der ausgewählten Modelle kennen. Dazu ist oftmals auch eine Auseinandersetzung mit den Ableitungen der Wachstumsfunktionen (Differenzialrechnung) erforderlich. Die 1. Ableitung ist dabei im Sinne einer Zuwachsfunktion verständlich, die 2. Ableitung, praktisch der Zuwachs des Zuwachses, stellt die zeitabhängige Krümmung der Regressionsbeziehung dar.

Dabei werden zwei grundsätzliche Modell-Philosophien unterschieden. Das sind das exponentielle Wachstum ohne eine Begrenzung des Wachstums und das logistische Wachstum mit einer Begrenzung des Wachstums durch Ressourcen. Es ist aber so, dass das exponentielle Wachstum lediglich bestimmte und begrenzte Zeitphasen einer Populationsentwicklung abbilden kann. Ein unendliches Wachstum ist aufgrund der Endlichkeit der Welt (Erde) nicht vorstellbar und daher auch nicht projizierbar.

Gelegentlich liegen dem Fachwissenschaftler auch Datenpunktwolken vor, die bereits erkennen lassen, dass das betrachtete Wachstum nicht monoton voranschreitet. In solchen Fällen liegt der Verdacht auf sich gegenseitig überlagernder Wachstumsverläufe bzw. auch negatives Wachstum („Absterbeprozesse“) nahe. Die so vorgenommene Zerlegung des Wachstumsverlaufs in einzelne

Wachstumsphasen (oder Wachstumsschübe) stellt u. a. PEIL (1974) beispielhaft dar. Andererseits sind auch zyklische Prozessverläufe denkbar. Diese Bewertung bleibt einer fachlichen Betrachtung der konkreten Situation vorbehalten.

Einen umfangreichen Überblick über den erreichten Stand der Forschungsarbeiten zur nichtlinearen Regressionsanalyse gaben SAGER & SAMMLER (1982), RASCH et al. (1984) schon vor Jahren oder neueren Datums SCHUSTER (2009).

Hier sollen die Modelle vorrangig für Anwendungsfälle in der ornithologischen Forschung dargestellt werden. Sie sind sicher weitgehend auch für die Medizin (z. B. Epidemie-Verläufe), Biologie und Biochemie sowie Wirtschaftsprozesse oder technische Verläufe anwendbar. Um die Akzeptanz der Anwendung der hier dargestellten Regressionsverfahren bei den meist mathematischen Ornithologen nicht zu schmälern, soll auf detaillierte mathematische Ableitungen verzichtet werden.

2. Modelle

Modelle sind mathematische Funktionen, die die Korrelation oder auch die Beeinflussung eines Merkmals durch eine oder mehrere Einflussgrößen soweit es möglich ist, als Formel darstellen. In Abhängigkeit von der Realitätsnähe des Modells verbleiben aber restliche Abweichungen von den berechneten Erwartungswerten. Diese Streuung ist in der Regel ein Signal, dass es weitere Einflussgrößen gibt, die die Zielgröße beeinflussen. Die Qualität des Modells kann durch das sogenannte Bestimmtheitsmaß B charakterisiert werden.

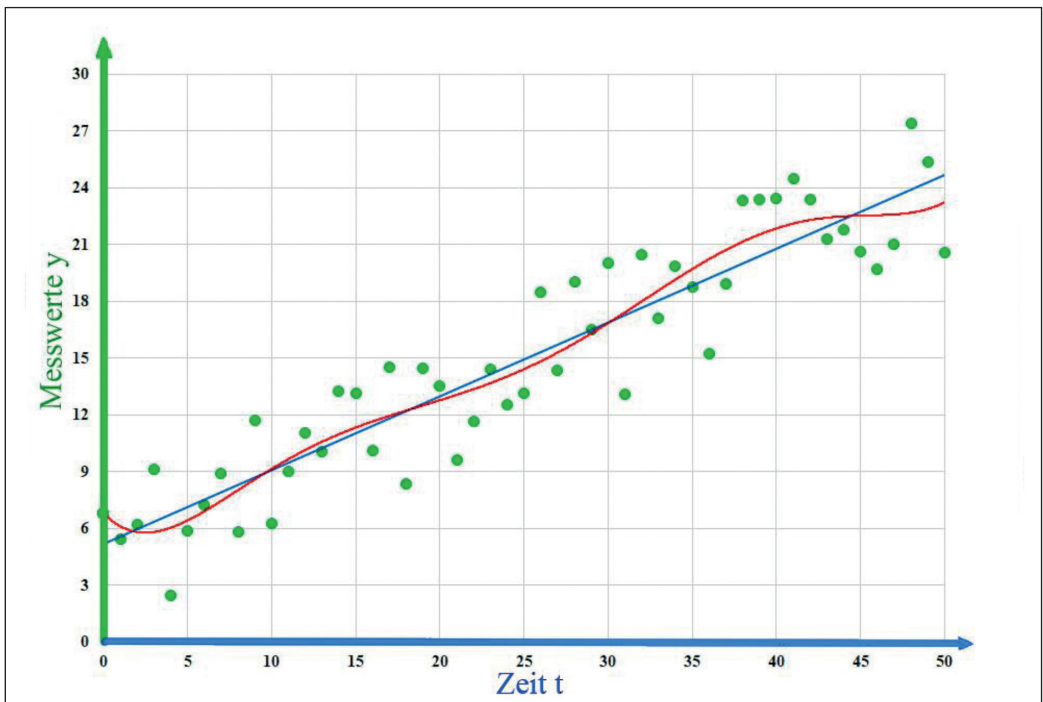


Abb. 1 Allgemeine Problemsituation einer Messwertfolge mit linearer Trendfunktion (blau) und einem Polynom-Modell 6. Grades (rot).

Bei einer *Linearen Regression* wird versucht diese Modellbeziehung durch ein Gerade des Typs

$$y = a + b_1 * t \quad (1)$$

abzubilden.

Bei einer *Quasilinearen Modellierung* sind additiv auch weitere Potenzen der Einflussgröße modelltypisch:

$$y = b_1 + b_2 * t + b_3 * t^2 + \dots + b_{m+1} * t^m \quad (2)$$

zulässig. Für diese Modelle sind die Parameter b_1 numerisch leicht zu berechnen. Die Interpretation der Ergebnisse ist dagegen oft nicht plausibel.

Die Eigentlich Nichtlineare Regressionsanalyse verwendet vielfach Modellkomponenten, die auf Exponential- oder Winkelfunktionen aufbauen. Die Parameter-Berechnung ist dann oft nur iterativ möglich.

2.1 Einfaktorielle Modelle

Hierbei wird versucht ein Modell der Beziehung zwischen einer Einflussgröße (hier in der Regel der Faktor **Zeit**, aber auch andere linear messbare Faktoren) und einer Zielgröße (z. B. Körpergröße, Gewicht, Entfernung u.a.) darzustellen. Klassische Beispiele sind so die Modellierung des Längenwachstum oder die Gewichtsentwicklung von Menschen oder Tieren über den Lauf der Lebensjahre oder das Wachsen einer Zellkultur über wenige Stunden. Aber auch die Lautstärke einer Explosion in Abhängigkeit von der Entfernung zum Explosionsort ließe sich mit derartigen Modellen berechnen.

2.1.1 Monotone Wachstumsverläufe

Monotone Wachstumsverläufe sind dadurch gekennzeichnet, dass sie keine Stagnationen oder gar Rückschritte hinsichtlich des Verlaufs der Zielvariablen besitzen. Solche Entwicklungskurven sind recht selten und oft nur unter Laborbedingungen zu erwarten. Andererseits lassen sich aber auch mit etwas „Gelassenheit“ manche Prozesse hinreichend genau darstellen.

Erste mathematische Modellfunktionen sind schon im frühen 19. Jahrhundert entwickelt worden. PEIL (1974) stellte eine Reihe derartiger Modelle und deren Eigenschaften zusammen. SAGER & SAMMLER (1982) verglichen am Beispiel des Längenwachstums der Petrale-Seezunge (*Eopsetta jordani*) die Eigenschaften verschiedener monotoner Wachstumsfunktionen. MATTHES et al. (1983) analysierten das Wachstum bei Rindern. Die Mannigfaltigkeit von Anwendungsbeispielen bis heute ist vor allem in der Tierzucht, Meeresbiologie, Medizin und Forstwirtschaft überwältigend. Ornithologen waren in der Modellierung dagegen doch noch sehr zurückhaltend.

Zunächst müssen wir uns mit den Vor- und Nachteilen sowie Interpretationsmöglichkeiten der Grundmodelle auseinandersetzen. Wann und warum die eine oder andere Funktion entwickelt wurde, ist hier wohl nicht zu bewerten, da weniger zielführend.

Zunächst unterscheiden wir Modelle danach, ab wann das Wachstum beginnt. Dies ist vielfach für $t \geq 0$ oder zumindest $t > 0$ sinnvoll. Das Wachstum hat hier also einen Startpunkt.

Eine zweite Gruppe zeichnet sich dadurch aus, dass man von einem schon seit „ewigen“ Zeiten verlaufenden Wachstum ausgeht. Hier gilt dann zumindest theoretisch: $-\infty < t < +\infty$. Das macht zwar in einigen Fällen theoretisch Sinn, wenn man einen zurückliegenden Startwert der Zielvariablen nachvollziehen will. Im Weiteren soll jedoch der praktisch nützlichere Fall von $t \geq 0$ betrachtet werden, selbst dann, wenn auch Modellwerte für Zeiträume vor dem Erfassungsbeginn denkbar wären.

Darüber hinaus unterscheiden wir Modelle, bei denen das Wachstum des Zielfaktors endlich ist gegenüber von Modellen mit einem unbegrenzt fortschreitendem Wachstum (wie man es oft einigen Fischarten unterstellt).

Zudem unterscheiden sich die Funktionen dahingehend, ob es im Anstiegsverlauf einen Wendepunkt t_w gibt. Das bedeutet dann, dass das Wachstum zunächst mit zunehmender Stärke vorankommt, dann aber einen Zeitpunkt t_w des maximalen Wachstums hat und danach doch wieder in eine immer mehr abflachende Wachstumsanstieg übergeht. Dies scheint in der Praxis der „normale“ Fall sein.

Andererseits gibt es auch ein „negatives Wachstum“. Das bedeutet, dass ausgehend von einem Startwert die Zielgröße mit der Zeit fortwährend abnimmt. Solche Verläufe werden **Absterbeprozesse** genannt.

Die Tabelle 2 soll jedoch einen Einblick in die wundersame Welt mathematischer Fantasie geben. Dabei soll nur eine kleine, aber vielfach praxiserprobte Auswahl an Modellen dargestellt werden.

Tab. 1 Symbole und Grundformeln.

Symbol	Bezeichnung	Formel / Erläuterung
t_i	i-ter Messzeitpunkt auf der Zeitachse t (i = 1, ..., n)	Bei den t_i kann es sich z. B. um ein Datum, eine Uhrzeit, ein Jahr handeln.
y_i	i-ter Zielwert zum Zeitpunkt t_i (i = 1, ..., n)	die y_i sind die Messwerte des betrachteten Merkmals zum Zeitpunkt t_i , z. B. Gewicht, Anzahl der Brutpaare, Entfernung von einem Ort
\hat{y}_i	i-ter Modellwert zum Zeitpunkt t_i (i = 1, ..., n)	die \hat{y}_i sind die berechneten Werte des betrachteten Merkmals zum Zeitpunkt t_i
b_i	Modellparameter (i=1, ..., m)	zu berechnende Variablen, die die Anpassung der Modellfunktion an die Daten sichern
a_j	Modellkonstanten (j=1, ..., k)	im Modell fest vorgegebene, weil bekannte Konstanten
F	Regressionsmodell	eine mathematische Formel $F=f(t, b_1, \dots, b_m)$ mit der die Entwicklung der Zielwerte über die Zeit dargestellt werden soll.
F(i)	Modellwert zum Zeitpunkt t_i (i = 1, ..., n)	berechneter Erwartungswert für das Regressionsmodell zum Zeitpunkt t_i
B	Bestimmtheitsmaß	B ist eine Kennzahl zur Beurteilung der Anpassungsqualität einer Regressionsfunktion. Das Bestimmtheitsmaß zeigt, wie gut sich die Messwerte durch das Modell darstellen lassen. Es gilt $0 \leq B \leq 1$, wobei $B=1$ bedeutet, dass sich die Messwerte ohne Abweichung durch das Modell darstellen lassen. $B=0$ bedeutet, dass das Modell völlig ungeeignet ist. $B = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad \text{mit } \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$

Tab. 2 Häufig verwendete nichtlineare Wachstumsmodelle.

Nr.	Modell-Autor / -Name	Funktion	n	Wendepunkt t_w =	$F(0)$ =	$F(\infty)$ =
1	PÜTTER 1920, BERTA-LANFFY 1934	$F(t) = b_1 - b_2 * e^{-b_3 * t}$ mit $b_1 > b_2$ und $b_1 > 0$; wenn jedoch $b_2 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	3	ohne	$b_1 - b_2$	b_1
2	Exponential-Funktion	$F(t) = b_1 - b_2 * (1 - b_3)^t$ mit $0 < b_3 < 1$, $b_2 > 0$ und $b_1 - b_2 \geq 0$; wenn jedoch $b_2 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	3	ohne	$b_1 + b_2$	b_1
3	JOHNSON 1935	$F(t) = b_1 * b_2^{(t/(t-b_3))}$ mit $b_1 > 0$; wenn jedoch $b_3 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	3	ohne	b_1	$b_1 * b_2$
4	Logistische Funktion	$F(t) = b_1 / (1 + b_2 * e^{-b_3 * t})$ mit $b_1 > 0$; wenn jedoch $b_3 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	3	$\ln(b_2)/b_3$ mit $y(t_w) = b_1/2$	$b_1/(1+b_2)$	b_1
5	GOMPERTZ 1825	$F(t) = b_1 * e^{(-b_2 * e^{-b_3 * t})}$ mit $b_1 > 0$; wenn jedoch $b_3 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	3	$\ln(b_2)/b_3$ mit $y(t_w) = b_1/e$	$b_1 * e^{-b_2}$	b_1
6	KORF 1939	$F(t) = b_1 * e^{-b_2 * t^{-b_3}}$ $t > 0$ und $b_1 > 0$; wenn jedoch $b_2 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	3	$[(b_2 * b_3)/(1+b_3)]^{(1/b_3)}$ mit $y(t_w) = b_1 * e^{(b_3-1)/b_2}$	≈ 0	b_1
7a	Sigmoid-Funktion 3 -parametrisch	$F(t) = b_1 * (1 - \tanh(-b_2 * (t - b_3)))$ mit $b_1 > 0$; wenn jedoch $b_3 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	3	b_3 mit $y(t_w) = b_1$	$b_1 * (1 - \tanh(b_2 * b_3))$	$2 * b_1$
7b	Sigmoid-Funktion 4 -parametrisch	$F(t) = b_1 - b_2 * (1 - \tanh(b_3 * (t - b_4)))$ mit $b_1 > 0$; wenn jedoch $b_3 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	4	b_4 mit $y(t_w) = b_1 - b_2$	$b_1 - b_2 * (1 - \tanh(-b_3 * b_4))$	b_1
8	MITSCHERLICH 1919, RICHARDS 1959	$F(t) = b_1 * (1 - b_2 * e^{-b_3 * t})^{b_4}$ mit $b_1 > 0$; wenn jedoch $b_2 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	4	$\ln(b_2 * b_4)/b_3$ mit $y(t_w) = b_1 * [(b_4 - 1)/b_4]^{b_4}$	$b_1 * (1 - b_2)^{b_4}$	b_1
9	JANOSCHEK 1957	$F(t) = b_1 * (1 - b_2 * e^{-(b_3 * t)^{b_4}})$ mit $b_1 > 0$; wenn jedoch $b_2 < 0$ wird aus dem Wachstums- ein Absterbe- prozess	4	$[(b_4 - 1)/(b_3 * b_4)]^{(1-b_4)}$ mit $y(t_w) = b_1 * (1 - b_2 * e^{[(1-b_4)/b_4]})$	$b_1 * (1 - b_2)$	b_1
10	Sinus-Modell	$F = b_1 * \sin(b_2 * (t - b_3))$ zyklische Funktion	3	mehrere	$b_1 * \sin(-b_2 * b_3)$	nicht definiert

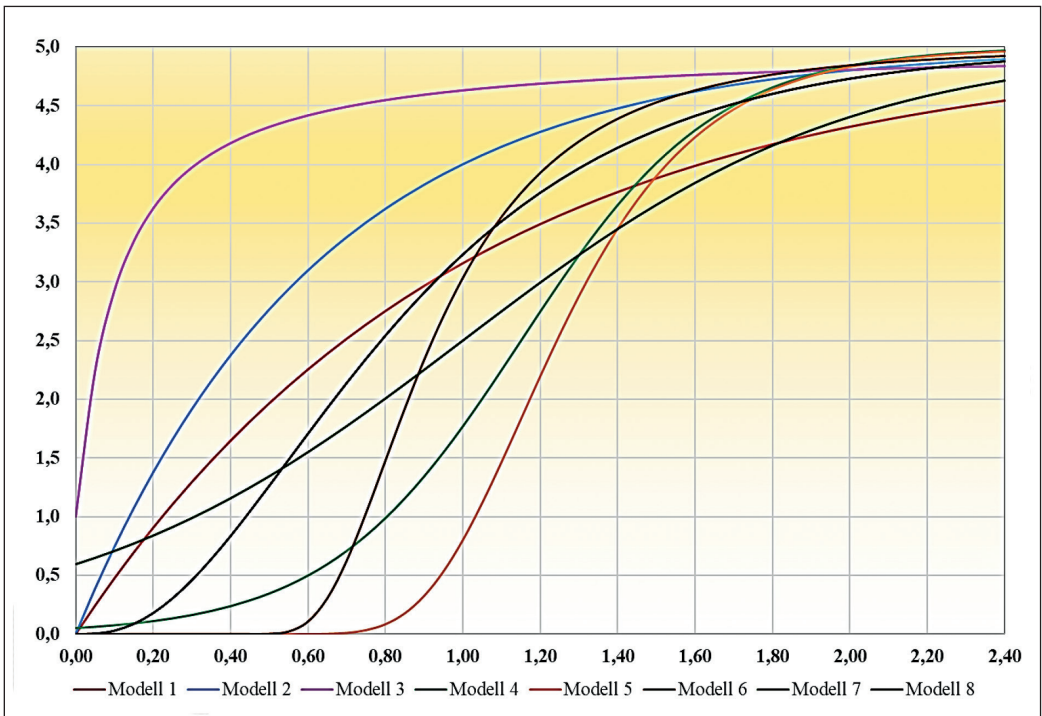


Abb. 2 Beispielsverläufe nichtlinearer Wachstumsmodelle.

Und die Menge erdachter Modellvarianten ist bei weitem noch nicht zu Ende. Allein die vorgeschlagenen Beispiele verwirren den Nicht-Mathematiker oft bestimmt schon sehr.

Was können uns diese Modelle - angepasst an die Daten - aber wirklich bieten? Zunächst bekommen wir bei endlichen Wachstumsprozessen eine Aussage zum **Maximum** [$F(\infty)$] des Prozessverlaufs.

Nicht unwichtig ist dann auch die Bestimmung des **Wendepunkts** t_w mit $F(t_w)$ eines Wachstumsprozesses, also dem Zeitpunkt, an dem der Anstieg (oder analog Absterbeverlauf) am stärksten ist. Für viele Prozesse sind diese Wendepunkte schließlich auch kritische Prozesspunkte. Und nicht unwichtig ist auch der Modellwert zum Beginn des Prozesses [$F(0)$]. Und hat man für diese Parameter selbst nur grobe Verdachtswerte, dann lässt sich die digitale Modellbestimmung sicher gestalten und beschleunigen. Und hier helfen oft auch grafische Darstellungen der x-y-Diagramme der Beobachtungspunkte erste Vermutungen anzustellen. Das betrachten wir näher an einzelnen Beispielen in einem späteren Kapitel.

Insgesamt lassen sich so mit relativ wenigen Parametern die typischen Wachstumsverläufe charakterisieren und geeignete **Prognosen** entwickeln.

Nehmen wir uns zum besseren Verständnis einmal ein Beispiel vor.

Die Wildvogelhilfe (<https://wp.wildvogelhilfe.org/vogelwissen/die-aufzucht/jungvogel-nichtgleich-jungvogel/entwicklung-eines-haussperling-nestlings/>) schreibt so: „Die privaten Vogelpfleger Stefan und Birgitt Sattler haben gemeinsam die Entwicklung eines Haussperling-Nestlings dokumentiert. Täglich wurde der Vogel ab dem Funddatum gewogen und fotografiert. Die Fütterung erfolgte gemäß unseren Empfehlungen für Gemischtfresser während der Nestlingsphase mit

Insekten und den entsprechenden Zusätzen.“ Zwischen dem 5. und 14. Lebenstag wurde dann der Jungvogel täglich gewogen.

Für die 10 Messwerte ergab das Polynomial-Modell das Ergebnis:

$$F(t) = 0,0011t^4 - 0,0422t^3 + 0,381t^2 + 1,6381t - 1,5385 \quad (3)$$

mit $B = 0,9881$.

Nur wenn man das Modell auch über den Messzeitraum hinaus benutzen möchte, ergibt sich schon für den 30. Lebenstag des Haussperlings ein Gewicht von etwa 170 g.

Verifizieren wir die Datenmenge mit den Eigentlich-nichtlinearen Modellen 1 bis 9, so ergeben sich jeweilig Modellparameterkombinationen mit Bestimmtheitsmaßen $B = 0,98183$ und $B = 0,98696$. Das sind ebenfalls gute Vorhersagen.

Das beste Ergebnis wurde mit Modell 6 mit mit $B = 0,9869$ erreicht.

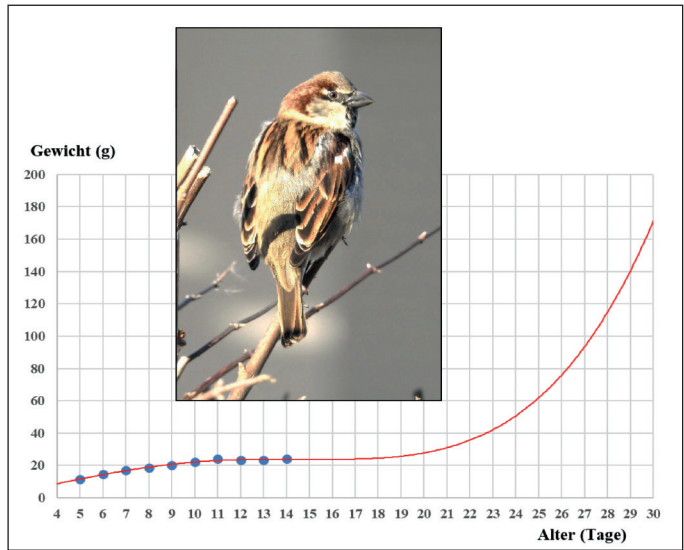


Abb. 3 Polynomial-Modell und Prognose der Gewichtsentwicklung über den Messzeitraum hinaus (Haussperling) (Aufn.: K.-D. FEIGE).

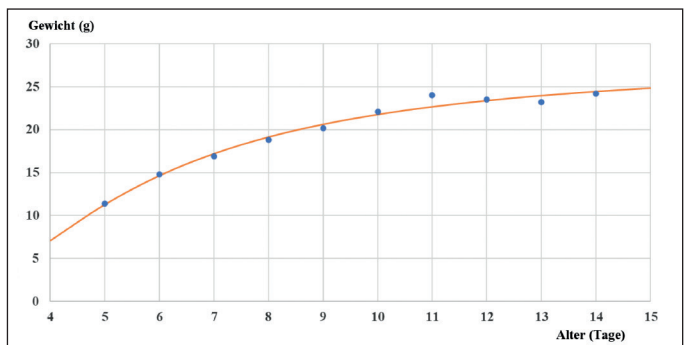


Abb. 4 Korf-Modell (Haussperling).

Die Modellparameter besagen nun, dass das Wachstum mit 27,9 g beendet sein wird. Der Wert liegt jetzt im Normal-Spektrum für die Art, eher sogar am unteren Rand. Der Wendepunkt des Wachstums wurde bereits nach 3,8 Tagen mit $F(t_w) = 6$ g erreicht.

Nehmen wir uns noch ein anderes Beispiel vor.

Petra Salm und Birgit Beckers haben von 2007-2014 die Zahl der Weißstorch-Paare in Soest veröffentlicht (<https://www.abu-naturschutz.de/natur-im-kreis-soest/voegel/weissstorch>). Der Trend zu mehr Storchepaaren war unübersehbar. Das Polynomial-Modell 5. Grades brachte ein Bestimmtheitsmaß von $B=0,9917$.

$$F(t) = 0,0004t^5 - 0,0235t^4 + 0,581t^3 - 7,1903t^2 + 44,534t - 107,53 \tag{4}$$

Die Vorhersage für das Jahr 2023 erbrachte hier die Erwartung von 93 Paaren.

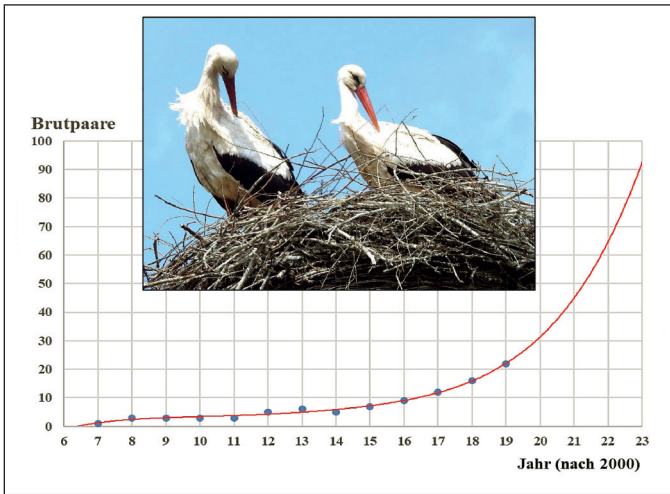


Abb. 5 Polynomial-Modell und Prognose der Gewichtsentwicklung über den Messzeitraum hinaus (Weißstorch) (Aufn.: K.-D. FEIGE).

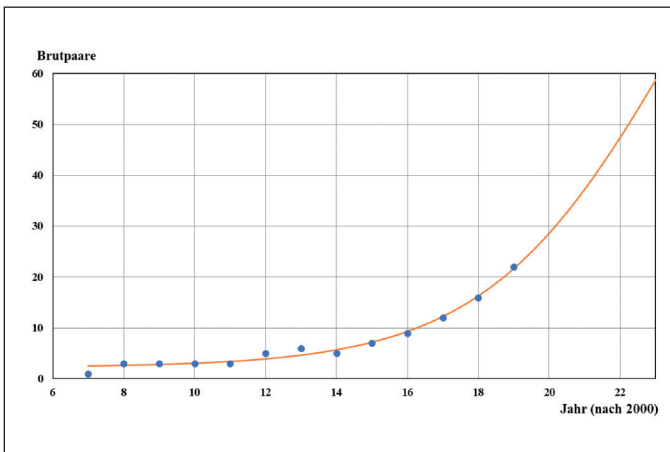


Abb. 6 4-parametrische Sigmoid-Funktion (Weißstorch).

Gehen wir davon aus, dass das Populationswachstum nicht unbegrenzt ist und die bisherigen Daten so einen Wendepunkt der Bestandsentwicklung erforderlich machen, kommen die ersten drei Modelle nicht mehr infrage. Die Verifizierung der anderen Modelle brachte für das Modell **7b** die höchste Bestimmtheit mit $B=0,9859$:

$$F(t) = 276,61 - 137,26 * (1 - \tanh(0,17348 * (t - 26,403))) \quad (5)$$

Für das Beobachtungsjahr 2023 werden hier 66 Brutpaare erwartet. Der Wendepunkt der Populationsentwicklung ($= b_4$) liegt zwischen den Jahren 2026 und 2027.

Und nun die Realität: Frau Beckers teilte mir auf Anfrage mit, dass 2023 genau 70 Brutpaare im Kreis nachgewiesen wurden. Das sind sogar noch mehr Paare als die Vorhersage erwarten ließ. Und 2022 waren es auch schon 57 Storchenpaare (Vorhersage 51 BP). Wissenschaftlich gesehen bedeutet dies, dass das verwendete Modell noch zu ungenau war - oder aber, dass nach 2019 weitere Faktoren der Populationsentwicklung dazu gekommen sind. Das schauen wir uns dann aber im nächsten Kapitel näher an.

Übrigens fällt es sicher auf, dass die Modelle 7a und 7b sich insbesondere durch eine Konstante zu Beginn der Formel unterscheiden. Diese Möglichkeit einer summerenten Konstanten b_1 als Parameter sichert dann, dass der Kurvenverlauf diesen Wert nicht unterschreitet.

2.1.2 Mehrphasige Wachstumsverläufe

2.1.2.1 Wachstumsschübe

So einfach wie die Modelle nun auch scheinen (das ist jetzt nicht ironisch gemeint), so viel komplexer kann die Realität sein. So hat es sich gezeigt, dass z. B. das frühkindliche Wachstum in mehreren „Schüben“ verläuft.

Ähnlich sieht dies auch bei Entwicklungskurven von Populationen aus. Hier können externe Faktoren nachhaltig verschiedene Entwicklungsphasen begründen. Diese Phasen müssen dabei noch nicht einmal durch dasselbe Grundmodell darstellbar sein. Hier hat sich insbesondere die additive Kombination von tanh-Modellkomponenten bewährt:

$$F(t) = b_1 - b_2 * (1 - \tanh(b_3 * (t - b_4))) + b_5 * (1 - \tanh(b_6 * (t - b_7))) + \dots + b_q * (1 - \tanh(b_{q+1} * (t - b_{q+2}))) \quad (6)$$

Ausführlich haben dies HELWIN & PEIL (1977) am Beispiel des Körperlängenwachstums des Menschen geprüft. Sie kamen dabei sogar auf bis 9 Wachstumsschübe.

KIRMSE & FREUND (2002) beschrieben die Brutbestandsentwicklung des Seeadlers in Sachsen von 1965 bis 2001. Die grafische Darstellung der Entwicklung der Population zeigte mehrere Schübe des Wachstums. Das quasilineare Modell erbrachte:

$$F(t) = -1E-06t^6 + 0,0006t^5 - 0,1191t^4 + 12,68t^3 - 755,73t^2 + 23918t - 314067 \quad (7)$$

mit einer Bestimmtheit von $B = 0,9938$. Die Prognose ist hier jedoch ein völliger Zusammenbruch der Population in nur wenigen Jahren.

Die Analyse der Daten mittels des Modells der Formel (6) mit 3 tanh-Komponenten ergab folgendes Ergebnis:

$$F(t) = 48,65 - 2,75 * (1 - \tanh(0,692 * (t - 77,7))) - 19,58 * (1 - \tanh(0,22 * (t - 92,8))) - 1,57 * (1 - \tanh(5,561 * (t - 97,7))) \quad (8)$$

mit $B = 0,9958$.

Die Interpretation lautet dann: Es gab drei Wachstumsschübe. Der erste mit einer Ausdehnung von 5,5 Paaren ($=2*2,75$) erreichte 1978 seinen maximalen Schub. Dem folgte ein stärkerer, aber langsamerer Bestandsaufschwung mit 39 Paaren ($=2*19,58$) der 1993 seinen stärksten Zuwachs

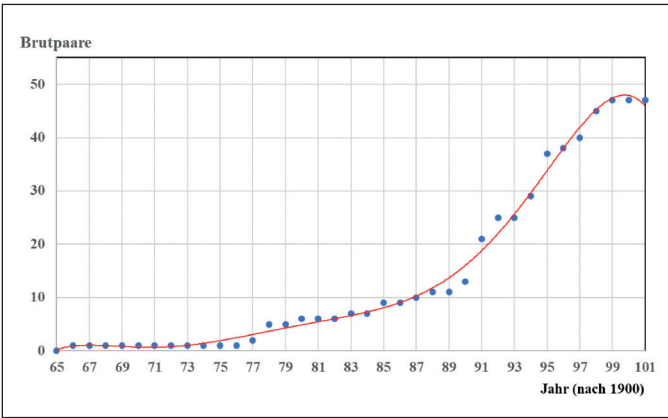


Abb. 7 Polynomial-Modell und Prognose der Brutbestandsentwicklung im Messzeitraum (Seeadler Sachsen).

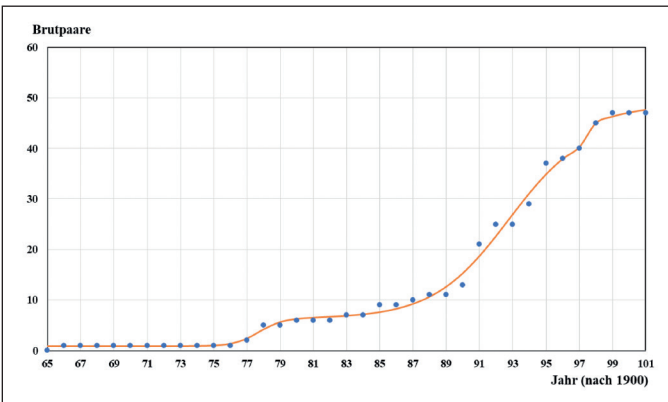


Abb. 8 10-parametrische Sigmoid-Funktion (Seeadler Sachsen).



Abb. 9 Junger Seeadler (Aufn.: K.-D. FEIGE).

hatte. Und schließlich dann noch einen schnellen Schub von gut 3 Paaren mit dem maximalen Anstieg 1998. Der Trend lässt bei unveränderten Rahmenbedingungen zukünftig keine Bestandszahlen über rund 49 Brutpaare erwarten. Soweit die Ergebnisanalyse.

Genau, aber darauf ist die zukünftige Aufmerksamkeit zu richten – wie also ändern sich die Rahmenbedingungen. Und das kann schon durch eine Gesetzesänderung des Bundesnaturschutzgesetzes erhebliche Auswirkungen haben. Aber auch um solche Folgen bewerten zu können sind diese Modelle hervorragend geeignet.

2.1.2.2 Wachstums-Absterbe-Modelle

Nicht immer gehen Wachstumsprozesse ungetrübt nach oben. Manchmal treten auch Absterbe- prozesse (Negativwachstum) in Erscheinung. Aber auch hier lassen sich die Wachstums- oder Absterbe-Komponenten verifizieren.

Anhand der durch P. HAUFF und R. FEIGE (in lit.) ermittelten und rekapitulierten Bestandszahlen der Brutpaare des Seeadlers (*Haliaeetus albicilla*) in Mecklenburg-Vorpommern von 1900 bis 2022 wurde so z. B. eine Trendanalyse für diese Art vorgenommen. Die Daten gelten als hinreichend genau gesichert, selbst wenn wenige erfassungsmethodische Ungenauigkeiten verbleiben (die Daten für 2019 und 2020 stehen wegen Unklarheiten bei der Definition eines Revierpaares nicht zur Verfügung).

Das 10-parametrische tanh-Modell brachte es dabei auf 3 Wachstumsschübe der Populationsentwicklung (mit $t = \text{Jahr}-1900$ numerisch berechnet):

$$F(t) = -164,7 + 113,34 \cdot (1 - \tanh(-0,04328 \cdot (t - 1944,25))) + 87,91 \cdot (1 - \tanh(0,0482 \cdot (t - 1956,7))) + 239,25 \cdot (1 - \tanh(-0,06014 \cdot (t - 2010,8))) \quad (9)$$

und $B = 0,9994$.

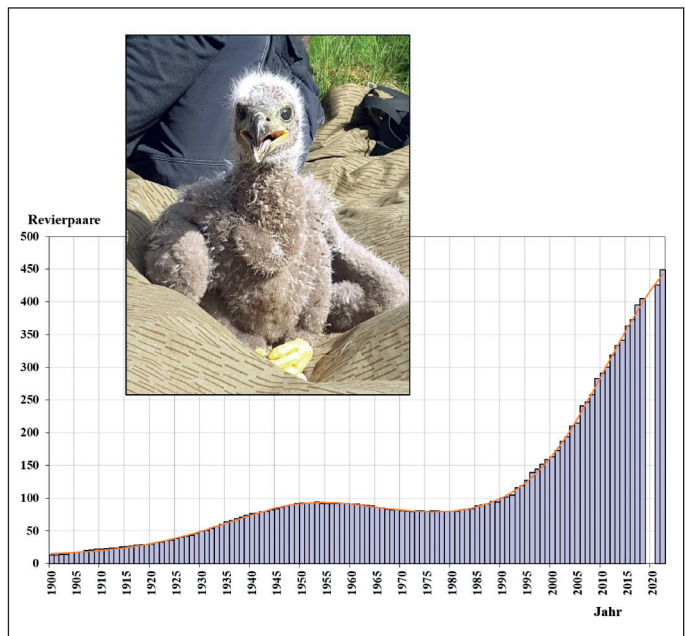


Abb. 10 10-parametrische Sigmoid-Funktion (Seeadler Mecklenburg-Vorpommern) (Aufn.: K.-D. FEIGE).

Die Interpretation: Nachdem der Bestand der Seeadler-Paare um 1900 fast völlig vernichtet wurde, nahm er aufgrund der veränderten Schutzbedingungen merklich zu, wobei diese Wachstumsphase einen Aufschwung auf bis 227 Paare (= 2*113,34) verzeichnete. Der steilste Wachstumsanstieg wurde um 1944 verzeichnet. Hierauf folgte eine Etappe des Rückgangs mit einer Spanne von fast 176 Paaren (= 2*87,91), die 1957 ihre steilste Abschwung-Phase besaß. Sie ist mit hoher Sicherheit mit dem exzessiven DDT-Einsatz in der Land- und Forstwirtschaft verbunden. Dass der Bestand nicht wieder auf ein Vorkriegsniveau zurückfiel, verdanken wir der aktuellen Regenerationsphase mit einem Aufschwung um knapp 480 Paare. Hier trat ein Wendepunkt um das Jahr 2011 ein. Daraus folgt eine perspektivische Sättigung der Bestände bei etwa 540 Paaren.

Kommen wir nun auch noch einmal auf die Entwicklung der Weißstorchbestände des Kreises Soest zurück. Hier hatte es sich gezeigt, dass neben der Besiedlung von Nisthilfen ab 2019 zunehmend Bäume (bzw. Masten) als „neue“ Nestunterlage verwendet werden. Wenn das keine Veränderung ist!

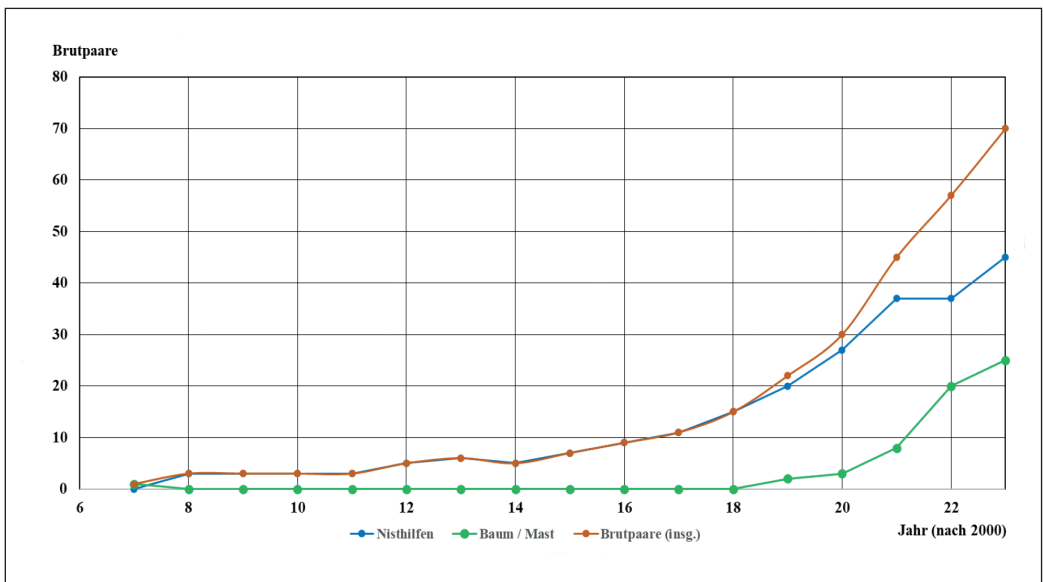


Abb. 11 Durch die Nestgrundlage differenzierte Populations-Entwicklungen (Weißstorch Soest).

Und auch hier lässt sich ein 7-parametrisches tanh-Modell anwenden, das sich summarisch aus zwei unterschiedliche Wachstumsprozesse zusammensetzt:

$$F(t) = 0,51 + 45,26 \cdot (1 - \tanh(0,151 \cdot (t - 22,95))) + 13,72 \cdot (1 - \tanh(0,800 \cdot (t - 21,46))) \quad (10)$$

und $B = 0,9922$.

Dabei haben wir mal vorausgesetzt, dass die Zahl der verfügbaren Nisthilfen nicht bereits völlig ausgereizt war. Im Ergebnis könnten so 90-91 Nisthilfen besetzt werden und gut 27 Baumbruten sind zu erwarten. Im Gesamttrend kommen wir perspektivisch auf 118-119 Brutpaare für den Kreis Soest. Lassen wir uns überraschen.

Abb. 12 Gesamt Populations-Entwicklung aufbauend auf zwei Niststätten-Grundierungen (Weißstorch Soest).

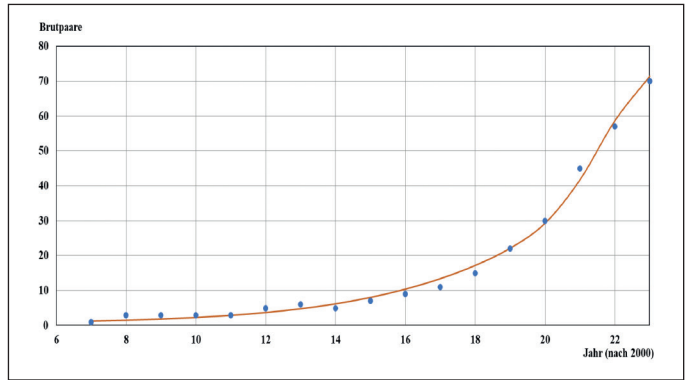


Abb. 13 Das wird noch was... (Aufn.: K.-D. FEIGE).

2.2 Zyklische Zeitreihen

2.2.1 Periodisches Modell

Oft beobachten die Biologen neben den Wachstums- und Absterbe-Prozessen aber auch Populationsentwicklungen, die offenbar zyklisch sind. Solche Schwingungen lassen sich zumindest mathematisch gut mit Winkelfunktion wie $\sin(t)$ oder $\cos(t)$ darstellen. Das Ziel ist es hier zeitlich begründete Datenreihen auf solche Perioden zu überprüfen. Das Grundmodell lautet:

$$F(t) = b_1 + b_2 * \sin(b_3 * (t - b_4)) \quad (11)$$

Damit alle Erwartungswerte des Modells nichtnegativ bleiben muss $b_1 \geq b_2$ sein. b_2 steht für die Weite der Schwingungen und b_3 für die Amplitude (Geschwindigkeit der Öszillation, je kleiner b_3 umso langsamer verläuft die Schwingung). b_4 stellt schließlich den Betrag der zeitlichen Verschiebung dar. Das jeweilige Maximum wird mit $b_1 + b_2$ erreicht.

Aber so einfach macht es uns die Natur in der Regel ja nicht. Meist sind es ja mehrere zyklische Prozesse, die unabhängig voneinander existieren und damit die Zeitreihen unterlegen:

$$F(t) = b_1 + b_2 * \sin(b_3 * (t - b_4)) + b_5 * \sin(b_6 * (t - b_7)) + \dots + b_q * \sin(b_{q+1} * (t - b_{q+2})) \quad (12)$$

2.2.2 Trendmodelle

Und dann kommt es noch aufregender, wenn sich diese periodischen Prozesse noch um einen wie auch immer verlaufenden Trend $F_{\text{trend}}(t)$ ranken. Dies kann einerseits additiv sein:

$$F(t) = b_1 + F_{\text{trend}}(t) + b_2 * \sin(b_3 * (t - b_4)) + \dots + b_q * \sin(b_{q+1} * (t - b_{q+2})) \quad (13)$$

Aber auch multiplikativ verlaufen – das bedeutet, dass die Amplitude durch die Trendfunktion multiplikativ beeinflusst ist:

$$F(t) = b_1 + F_{\text{trend}}(t) * [b_2 * \sin(b_3 * (t - b_4)) + \dots + b_q * \sin(b_{q+1} * (t - b_{q+2}))] \quad (14)$$

Und nun könnte man noch viel weiter experimentieren und Modelle entwickeln, die sich aus additiven und multiplikativen Überlagerungen zusammensetzen. Nur das geht über das Ziel dieser Arbeit im Sinne von Annäherung an ein kompliziertes Thema hinaus. Vielleicht später einmal.

Das folgende Diagramm (Abb. 14) zeigt beispielhaft für verschiedene Parameterkombinationen Funktionen entsprechend Formel (12).

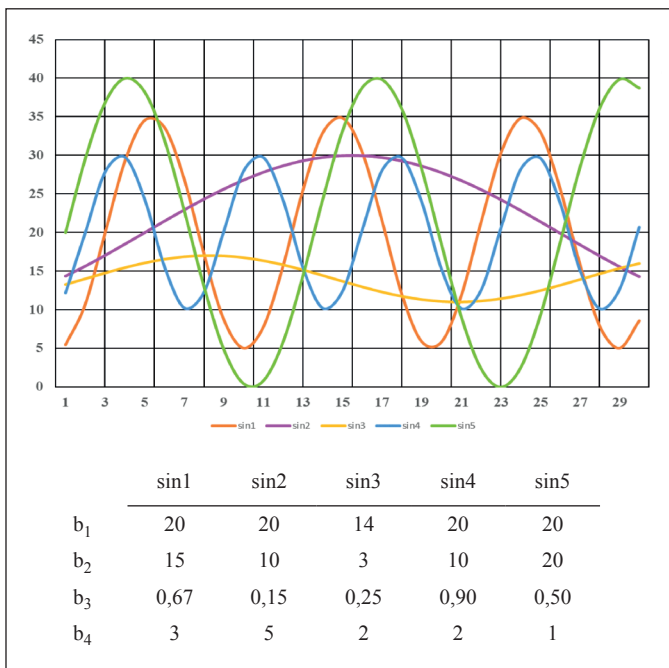
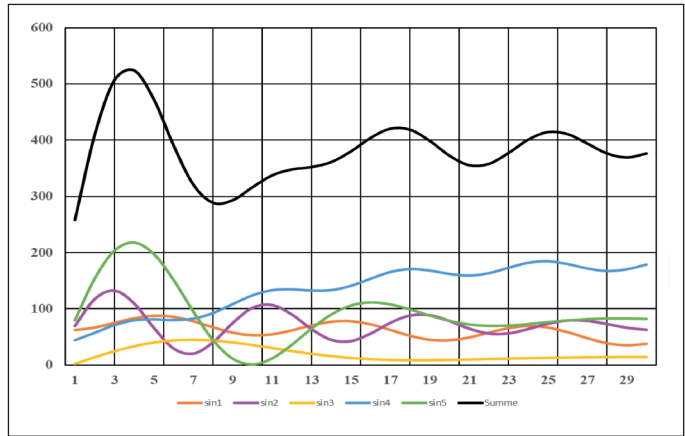


Abb. 14 Sin-Modelle mit verschiedenen Modell-Parameterkombinationen.

Abb. 15 Verständnisdarstellung von sin-Modellen mit verschiedenen Modell-Parameterkombinationen und Varianten der Trendbindung sowie der Summe verschiedener Teilmodelle.



Und die Aufspaltung eines solchen summarischen Verlaufs verschiedener Zyklen in die interpretierbaren Einzelprozesse ist dann das Ziel der Datenanalyse.

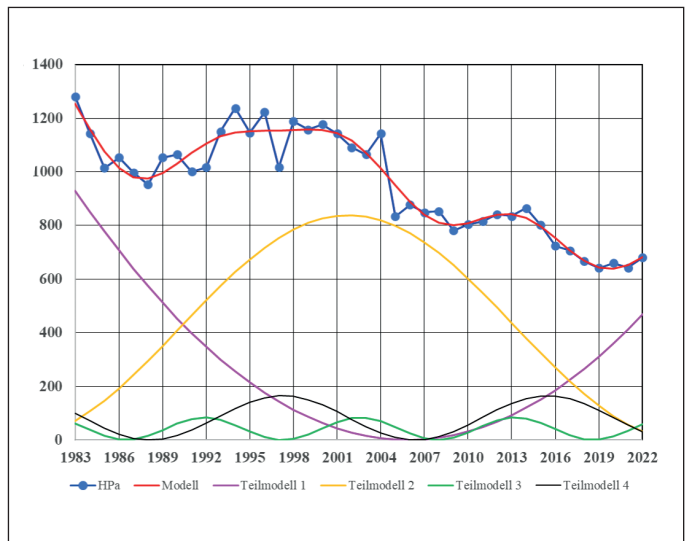
Zwei Beispiel sollen dies unterlegen. Zunächst nehmen wir uns die Populationsentwicklung der Weißstörche in Mecklenburg-Vorpommern von 1983-2022 vor. Basierend auf Modell (12) mit vier Schwingungen erhalten wir optimiert das Modell

$$F(t) = 91,8 + 6072 * \sin(0,02451*(t - 1941)) + 419,2 * \sin(0,1359*(t - 1967)) + 42 * \sin(0,5869*(t - 1994)) + 82,5 * \sin(0,3448*(t - 1984)) \quad (15)$$

mit $B=0,932$.

Grafisch sieht das dann wie folgt aus (Abb. 16).

Abb. 16 Populationsentwicklung der Weißstörche in Mecklenburg-Vorpommern in Schwingungsphasen (Teilmodelle) aufgliedert.



Ist dieses Modell realitätsnah genug, läge die Prognose z. B. für 2023 bei 721 Paaren und für 2024 bei 770 HPa. Ein hoffnungsvoller Trend.

Die Frequenz ($2 \cdot \pi / b_3$, usw.) liegt bei den einzelnen Schwingungsphasen bei Teilmodell 1: 256 Jahre, Teilmodell 2: 46 Jahre, Teilmodell 3: 10,7 Jahre und Teilmodell 4: 18 Jahre. Eine ähnliche Teilschwingung um 11 Jahre wie im Teilmodell 3 hatte ich bereits einmal beim Pirol hinsichtlich der Häufigkeitsentwicklung über die Jahre in Ostdeutschland zu verzeichnen. Sie liegt wie auch dort ziemlich nahe der Aktivitätszyklen der Sonnen mit 11,1 Jahren. Nur mal so als Anregung fürs Nachdenken.

2.2.3 Retrospektierende Modelle

Sicher ist es selbst den gutmeinenden Lesern an dieser Stelle bereits genug, um sich vielleicht einmal allein an den eigenen Daten auszuprobieren zu wollen. Der Vollständigkeit halber soll jedoch darauf hingewiesen werden, dass die Wachstumsprozesse auch von sich selbst abhängen können.

So zeigt es sich beim Pirol, dass eine hohe Brutpaarzahl schon zwei Jahre vor dem aktuellen Beobachtungsjahr im selben Gebiet mit der dann vermuteten hohen Jungenzahl und mit deren Geschlechtsreife im 2. Lebensjahr auf die aktuellen Brutpaarzahlen „durchschlug“. Und sogar gute oder schlechte Ergebnisse von vor 4 Jahren waren im Modell noch nachweisbar. Allgemein gesagt ist dann

$$F(t) = F(t, F(t-x)) \quad (16)$$



Abb. 17 Pirol-Weibchen (Aufn.: A. HABICHT).

2.3 Mehrfaktorielle Modelle

Bisher haben wir nur Wachstumsmodelle betrachtet, die allein vom Faktor Zeit t abhängig waren. Anstelle des Faktors Zeit können aber auch andere Einflussgrößen stehen. So wäre auch

die Temperatur als prozessbeeinflussendes Merkmal denkbar. Oder die Nährstoffkonzentration in verschiedenen Gewässern. Wichtig ist nur, dass das zu analysierende Zielmerkmal durch die Variabilität der Einflussgröße selbst Veränderung zeigt. Und natürlich kann es so gleichzeitig auch mehrere Einfluss-Variablen geben. Aber die dann abzuleitenden Modelle sind dann schon „sehr hohe Schule“ und mit dieser Vorstellung soll nur einen Ausblick auf weitere Möglichkeiten der Regressionsanalyse gegeben werden. Hier lassen sich dann auch Räuber-Beute-Modelle einordnen.

3. Parameterschätzung

Jetzt kommen wir jedoch zum schwierigsten Teil der Modellberechnungen. Es gilt die Parameter des gewählten Modells mit den Daten so zu verbinden, dass die Parametersymbole durch konkrete Zahlenwerte ersetzt werden können. Und da es komplexe Modelle sind, die wir verifizieren müssen, bieten sich in der Regel keine direkten Parameterberechnungen an.

Der Lösungsweg besteht darin, dass man ausgehend von geeigneten Startwerten für die Parameter eine schrittweise Annäherung an ein Optimum vornehmen kann.

Das Ziel besteht hier darin, dass die Summe der Quadrate der Abweichungen zwischen beobachteten Wert und dem zugehörigen Modellwert minimiert wird.

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \text{Min.} \quad (17)$$

Hier bietet sich dann für die Iterationen ein schon 1975 von PAUL beschriebenes Verfahren, dass z. B. bei Excel mit den Solver-Algorithmus annähernd umgesetzt wurde.

Die Qualität der Startwerte für die iterativen Berechnungen ist jedoch wesentlich für die Qualität des Ergebnisses der Modellierung. Es gilt also an dieser Stelle hinreichend viel Zeit in die Berechnungs-Vorbereitung zu stecken.

Durch Vermutungen über

- die Existenz von Wendepunkten und deren Zeitpunkt Im Wachstumsverlauf,
- die Endlichkeit des Wachstums,
- den Zielwert zum Zeitpunkt $t=0$

kann man gute Startwerte der Iteration bestimmen.

Nehmen wir uns dazu noch einmal das erste Beispiel mit dem Haussperlingsnestling vor. Mit dem Modell 6 (Korf) endet das Wachstum mit dem Parameter b_1 . Da wir davon ausgehen können, dass Haussperlingen kaum größer als 35 g erreichen, wäre also $b_1=35$ ein guter Startwert (Symbole der Startwerte werden kursiv dargestellt).

Und wenn wir annehmen, dass der Jungvogel nach gut 16 Tagen sein Zielgewicht erreicht hat, kann man vermuten, dass der Wendepunkt irgendwie bei der Hälfte der Zeit lag. Also ist dann $t_w=8$ anzunehmen. Selbst die Annahme, dass dann $y(t_w)$ etwa bei der Hälfte des Endwerts liegt, also $y(t_w)=15$ ist, wirkt plausibel. Und bei den beiden anderen Parametern kann man ruhig etwas experimentieren. Stellt man die Formel für $y(t_w)$ hier für die angenommenen Werte von b_1 und $y(t_w)$ um, so ergibt sich

$$b_3 = b_2 * \ln(15 / 30) + 1 \quad (18)$$

Die so folgende Robustheit des iterativen Verfahrens wird manchen überraschen.

Betrachten wir noch ein anderes Beispiel. Nehmen wir noch einmal die Populationsentwicklung der Weißstörche im Kreis Soest und wählen das 4-parametrische tanh-Modell aus (Modell 7b).

Gehen wir einmal davon aus, dass der Lebensraum auch mit vielen Nisthilfen doch nicht mehr als 200 Paare ermöglicht, dann ist der Schätzwert für $b_1=200$. Noch befindet sich die Population im Aufwind, so dass der Wendepunkt des Aufstiegs sicher erst so um 2025 liegt. So ist dann $b_4=25$ anzusetzen. Zu dem Zeitpunkt vermuten wir mal, dass die Hälfte der Endpopulation erreicht ist. Also $y(t_w)=100$ ist. Und daraus folgt mit $y(t_w)=b_1-b_2$ sofort $b_2=200-100=100$. Bleibt nur noch b_3 zu schätzen. Wenn wir also annehmen, dass die Verdopplung der Populationsveränderung alle 3 Jahre stattfindet, dann nehmen setzen wir $b_3=1/3=0,3333$.

Die Lösung zeigt uns dann die berechneten Parameter $b_1=121,37$, $b_2=59,21$, $b_3=0,2357$ und $b_4=22,40$.

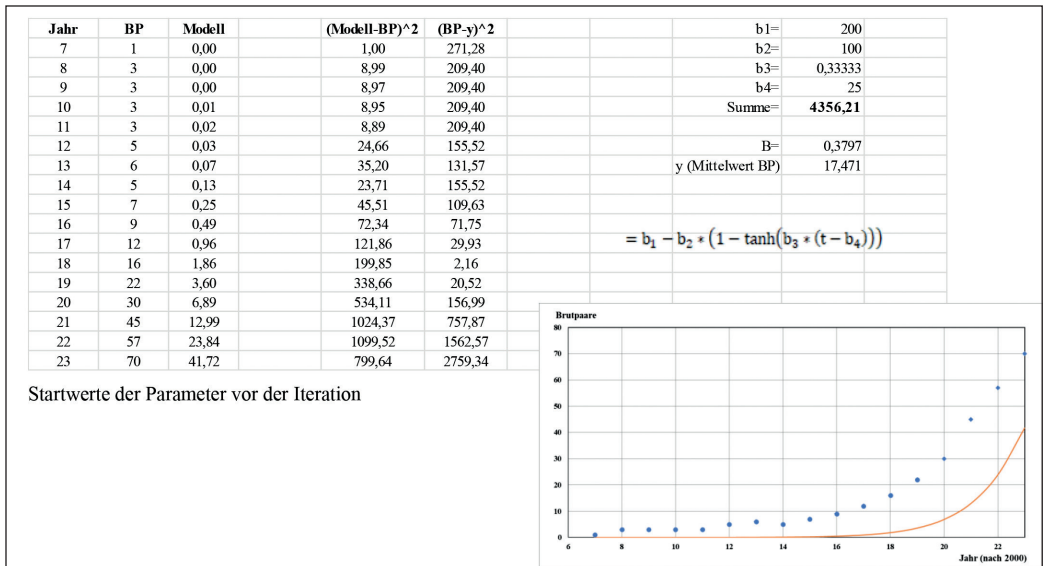


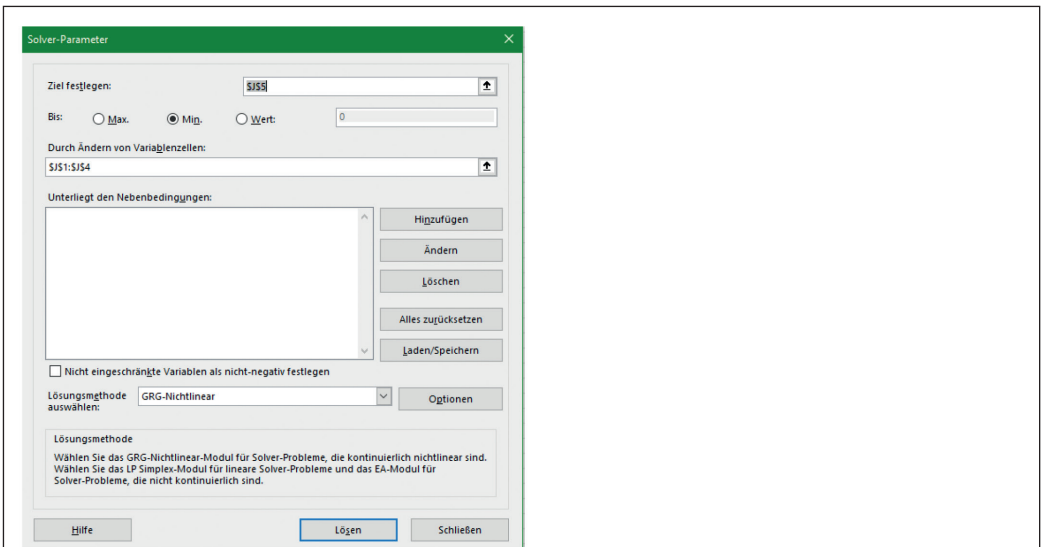
Abb. 18 Grobe Modellanpassung mit angenommenen Startwerten vor der Excel-Solver-Iteration.

Die Abbildung 18 zeigt das Datenmuster und das x-y-Diagramm der des Lösungsschemas in Excel. Und die Abbildung 19 dann das Ergebnis nach der Solver-Berechnung.

Aber machen wir uns auch nichts vor. Bereits andere Startwerte können durchaus zu Ergebnissen mit geringen Abweichungen der berechneten Parameter führen. Und manchmal ist die Abweichung auch erheblich. Das sind dann Anzeichen dafür, dass einerseits die Datenreihe noch zu kurz ist oder das gewählte Modell nicht den Anforderungen genügt. Hier sind dann ein Experimentieren mit den Startwerten und ein Durchprobieren verschiedener Modelle sinnvoll.

Nur es ist wie beim Lotto – wer nicht mitspielt hat auch keine Chance auf einen Gewinn. Und die Gewinnchancen auf den Jackpot liegen hier weit höher als bei „6 aus 49“. Am Anfang kann man sich ja zudem auch mal helfen lassen.

Natürlich gibt es neben den Möglichkeiten von Excel auch spezielle Software, mit denen man solche Modelle berechnen kann (Open-Source-Software R u.a.). Leider setzen diese aber einen erheblichen Lernaufwand voraus, der sich oft nur für Spezialisten der Statistik lohnt.



Solver-Ansatz

Jahr	BP	Modell	(Modell-BP)^2	(BP-y)^2			
7	1	3,03	4,13	271,28			b1= 121,37
8	3	3,08	0,01	209,40			b2= 59,21
9	3	3,16	0,03	209,40			b3= 0,2357
10	3	3,29	0,09	209,40			b4= 22,4
11	3	3,50	0,25	209,40			Summe= 16,72
12	5	3,82	1,39	155,52			B= 0,9976
13	6	4,34	2,75	131,57			y (Mittelwert BP) 17,471
14	5	5,17	0,03	155,52			
15	7	6,46	0,29	109,63			
16	9	8,48	0,27	71,75			
17	12	11,56	0,19	29,93			
18	16	16,17	0,03	2,16			
19	22	22,80	0,63	20,52			
20	30	31,83	3,36	156,99			
21	45	43,30	2,89	757,87			
22	57	56,59	0,16	1562,57			
23	70	70,48	0,23	2759,34			

$$= b_1 - b_2 * (1 - \tanh(b_3 * (t - b_4)))$$

Ergebnis der Iteration

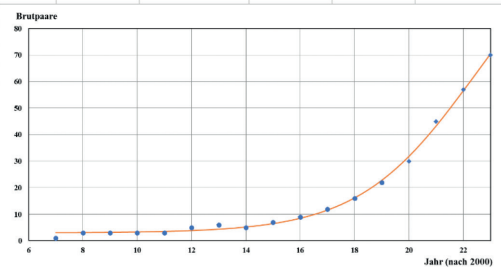


Abb. 19 Ergebnisbild der Parameterschätzung nach der Excel-Solver-Iteration.

Zusammenfassung

Wachstumsmodelle sind in vielen Fragestellungen den quasilinearen Regressions-Polynomen vorzuziehen. Die effektive Anwendung dieser setzt jedoch eine gute Kenntnis der Eigenschaften des Modells voraus. An Praxisbeispielen werden die Möglichkeiten zur Modellwahl und Anfangswertschätzung der Modell-Parameter vorgestellt. Die Charakterisierung verschiedener, häufig verwendeter Modelle der Wachstumsanalyse soll auf modellspezifische Besonderheiten hinweisen.

Literatur

- FAHRMEIR, L., KNEIB, T., LANG, S. & B. MARX (2013): Regression: models, methods and applications. – Springer Science & Business Media.
- FEIGE, K.-D. (1988): Zur Methodik der Modellwahl und Parameterschätzung in der nichtlinearen Regressionsanalyse. – Tag.-Ber. Akad. Landwirtsch.-Wiss. Berlin, 5-11.
- FEIGE, K.-D. (1988): Mathematische Modellierung und Problemanalyse im Wissenschaftlichen Gerätebau. – Tag.-Ber. Akad. Landwirtsch.-Wiss. Berlin, 29-36.
- HELWIN, H. & J. PEIL (1977): Phänomenologisch-mathematische Modellierung des Körperlängenwachstums des Menschen durch Zerlegung in Wachstumsschübe und deren quantitativ-analytische Erfassung. – Gegenbaurs morph. Jahrb. Leipzig 123, 641-655.
- KIRMSE, W. & W. FREUND (2002): Bestandsentwicklung und Lebensräume des Seeadlers (*Haliaeetus albicilla*) in Sachsen. – Corax 19, Sonderheft 1, 63-67.
- MATTHES, H.-D., SAGER, G., PANICKE, L. & H. PETERS (1983): Charakterisierung des Wachstumsverlaufs über Wachstumfunktionen bei männlichen und weiblichen Junggrindern. – Archiv für Tierzucht.
- PAUL, W. (1975): Ein direktes numerisches Verfahren zur Bestimmung der Parameter nichtlinearer Ausgleichsfunktionen. – Biom. Z., Berlin 17, 487-500.
- PEIL, O. J. (1974): Mathematische Beschreibung von Wachstumsvorgängen I. – Gegenbauers morphol. Jahrb., Leipzig 120, 832-853.
- RASCH, D. (1984): Wachstum: Arbeitsmaterial zum Forschungsthema Wachstum. – Dummerstorf-Rostock: Forschungszentrum für Tierproduktion (Probleme der angewandten Statistik; 1/1984).
- SAGER, G. & R. SAMMLER (1982): Approximation des Längenwachstums von *Eopsetta jordani* über eine Reihe von Wachstumfunktionen. – Zool. Jb. Anat. 108, 37-46.
- SCHUSTER, R. (2009): Wachstumsmodelle in Medizin, Biologie und Biochemie, Dynamik einer von der Zeit abhängigen Population. – In: Biomathematik. Vieweg + Teubner.

Dr. KLAUS-DIETER FEIGE
Zum Kirschenhof 38
D - 19057 Schwerin
E-Mail: oriolus@compuwelt.de