

365 | 1976

**SCHRIFTENREIHE SCHIFFBAU**

G.O. Andersson

**Statistische Analyse des Widerstandes  
von Schiffsmodellen**

**TUHH**

*Technische Universität Hamburg-Harburg*

**Statistische Analyse des Widerstandes von Schiffsmodellen**

G.O. Anderson

Hamburg, Technische Universität Hamburg-Harburg, 1976

© Technische Universität Hamburg-Harburg

Schriftenreihe Schiffbau

Schwarzenbergstraße 95c

D-21073 Hamburg

<http://www.tuhh.de/vss>

Statistische Analyse des Widerstandes  
von Schiffsmoellen

von  
G.O. Andersson

Teilprojekt C2 des Sonderforschungsbereiches 98

"Schiffstechnik und Schiffbau"

Hamburg 1976

## V o r w o r t

Immer wieder werden in der Literatur neue Widerstandsanalysen, Zusammenstellungen von Versuchsergebnissen veröffentlicht, immer noch werden in Versuchsanstalten Widerstandsversuche mit Schiffsmodellen ausgeführt. Das beweist, daß die vorhandenen Unterlagen in den meisten Fällen nicht ausreichen, den Widerstand eines Schiffes ohne Modellversuch genügend genau vorauszusagen. Das gilt sogar für konventionelle Schiffsformen, in besonderem Maße natürlich für neuartige, ungewöhnliche Formen.

Wenn sich der Sonderforschungsbereich Schiffbau entschlossen hat, in einem Teilprojekt eine weitere Widerstandsanalyse auszuarbeiten, waren dafür zwei Motive entscheidend: Erstens stehen heute für die Auswertung von Meßergebnissen wirkungsvolle statistische Methoden, wie z.B. die multiple Regressionsanalyse, zur Verfügung, in die dank der elektronischen Rechentechnik viele Daten, und vor allem viele Variable, eingeführt werden können, und zweitens liegen in der HSVA viele Versuchsergebnisse, für die bisher keine so vollständige statistische Auswertung bekannt geworden ist.

Der Bericht liegt nun vor. Ausgewertet wurde der Restwiderstandsbeiwert nach der Froudeschen Methode und daneben der Formfaktor. Großer Wert wurde darauf gelegt, herauszufinden, welche Formparameter des Schiffes von dominierender Bedeutung sind und für welche Kombination von Formparametern lineare Beziehungen zu den gesuchten Werten bestehen. Hierfür wurde neben der Regressionsanalyse auch die Faktorenanalyse eingesetzt, von denen die letztere bisher wohl noch nie zu diesem Zweck verwandt wurde.

Das Ergebnis stellen Regressionsgleichungen dar. Es werden zahlreiche statistische Daten wiedergegeben, die es erlauben, für jede Anwendung die Vertrauensgrenzen zu bestimmen sowie festzustellen, ob das aktuelle Schiff innerhalb des Bereiches des zugrunde liegenden Kollektivs liegt. Das ermöglicht nicht nur, die gesuchten Werte vorauszusagen, sondern sich auch ein Urteil zu bilden, welche Genauigkeit diesen Werten zugemessen werden darf. Dies und die Auswertung für jeweils eine Froude-Zahl - unabhängig von den benachbarten

Froude-Zahlen - führte allerdings zu Punkten, die über der Geschwindigkeit aufgetragen keine strakende Kurve ergeben; die strakende Kurve muß innerhalb der Vertrauensgrenzen gezogen werden.

Ob diese Widerstandsanalyse, diese Art der Auswertung der Meßergebnisse einen Fortschritt gebracht hat und verwendet werden wird, muß der Zukunft überlassen werden. Die Analyse zeigt wieder einmal, und dieses Mal besonders deutlich, in wie unübersichtlicher Weise der Widerstand von den Formparametern abhängt und warum es daher bisher keine vollständig ausreichende Analyse gibt. Die Analyse muß als Versuch gewertet werden, mit modernen Methoden ein altes Problem zu bearbeiten. Durch die Ergebnisse wurden die Erwartungen nur zu einem geringen Maße erfüllt. Die Schwierigkeiten wurden zu Beginn der Arbeit unterschätzt. Vielleicht wird es möglich werden, den Versuch fortzusetzen und die benutzten statistischen Methoden zu verbessern.

O. Grim  
wissenschaftlicher Leiter des  
Teilprojektes

Inhaltsverzeichnis

	<u>Seite</u>
Symbolverzeichnis	iv
Literaturverzeichnis	viii
1. Einleitung	1
2. Überblick über Regressionsanalysen des Widerstandes	4
3. Statistische Methoden im Zusammenhang mit der Regressionsanalyse	6
3.1. Allgemeines	6
3.2. Beschreibende statistische Größen	7
3.2.1. Mittelwert, Streuung, Standardabweichung	7
3.2.2. Die Kovarianz	8
3.2.3. Der Korrelationskoeffizient	8
3.2.4. Der multiple Korrelationskoeffizient	9
3.3. Statistische Begriffe und Testgrößen	10
3.3.1. Wahrscheinlichkeit	10
3.3.2. Die statistische Größe $t$	10
3.3.3. Signifikanztest	12
3.3.4. Der $\chi^2$ - Test auf Normalverteilung	13
3.3.5. Weitere Prüfungen auf Normalverteilung	14
3.3.6. Die statistische Größe $F$	16
3.3.7. Untersuchung auf Zufälligkeit	17
3.3.7.1. Zurückweisung von Extremwerten (Ausreißern)	18
3.3.7.2. Test auf runs	18
3.3.7.3. Methode der Mean Square successive Differences (MSSD)	18

	<u>Seite</u>	
3.4.	Regressionsanalyse	19
3.4.1.	Beschreibung des Verfahrens	19
3.4.2.	Annahmen, Forderungen und Anwendung	23
3.4.3.	Einige ergänzende Formeln	24
3.5.	Faktorenanalyse	26
3.5.1.	Definition und Anwendungszweck	26
3.5.2.	Mathematische Grundlagen	28
3.5.3.	Rotation	32
4.	Verfahren der Widerstandsanalyse	34
4.1.	Graphische Verfahren Eine Anwendung auf modernen Schiffsformen	34
4.2.	Mathematisch-statistische Verfahren	35
4.2.1.	Der geschwindigkeitsabhängige Ansatz	37
4.2.1.1.	Ansatz von OORTMERSSEN	37
4.2.1.2.	Ansatz nach INUI	38
4.2.1.3.	Auswertung der Ansätze	38
4.2.2.	Der Ansatz bei konstanter Froudezahl	41
5.	Die verfügbaren Daten	43
5.1.	Die abhängige Variable	44
5.2.	Die unabhängigen Variablen	46
5.2.1.	Wahl der Variablen	46
5.2.2.	Faktorenanalysen der unabhängigen Variablen	48
5.2.3.	Funktionale Zusammenhänge mit dem Restwiderstandsbeiwert	56

	<u>Seite</u>	
6.	Ergebnisse der Restwiderstandsanalyse	58
6.1.	Gruppeneinteilung und Bemerkungen zur Durchführung der Berechnungen	58
6.2.	Darstellung der Ergebnisse	61
6.3.	Diskussion der Ergebnisse	64
7.	Der Formfaktor $k$	69
7.1.	Theoretische Ableitung	69
7.2.	Ermittlung von Formfaktoren	71
7.3.	Darstellung und Diskussion der Ergebnisse	73
8.	Zusammenfassung	75

Symbolverzeichnis

$a_{jl}$	Wert der Ladung des l-ten Faktors durch die j-te Variable
$b_j$	Regressionskoeffizient der j-ten unabhängigen Variablen einer Stichprobe
$c_{kp}$	Element der linear transformierten Korrelationsmatrix
$g$	Erdbeschleunigung
$k$	Formfaktor
$i_{E,R}$	Wasserlinienwinkel (Entrance bzw. Run)
$m$	Anzahl sämtlicher unabhängiger Variabler in der Regressionsanalyse
$m_{lj}$	Element der Transformationsmatrix
$n$	Anzahl der Stichprobendaten eines Merkmals
$p$	Anzahl der in die Regressionsgleichung aufgenommenen unabhängigen Variablen
$p_{lj}$	Beobachtungswert einer unabhängigen Variablen nach der Faktorenanalyse
$r_{x,y}$	Korrelationskoeffizient zwischen den Variablen x und y
$s$	Standardabweichung
$s^2$	Streuung oder Varianz
$s_{x,y}$	Kovarianz der Variablen x und y
$s_{y;x_1 \dots x_p}$	Standardabweichung einer Regressionsgleichung für y mit den unabhängigen Variablen $x_1$ bis $x_p$
$t$	statistische Testgröße; t - Verteilung

$x$	Symbol für eine unabhängige Variable
$\bar{x}$	Mittelwert einer Stichprobe der Variablen $x$
$x_B$	Verdrängungsschwerpunkt der Länge nach
$x_F$	Wasserlinienschwerpunkt
$y$	Symbol für die abhängige Variable
$\bar{y}$	Mittelwert einer Stichprobe der abhängigen Variablen
$\tilde{y}$	aus der Regressionsgleichung berechneter Näherungswert für einen gegebenen $y$ -Wert
$z_{ji}$	Element der standardisierten Variablenmatrix
$z_B$	Verdrängungsschwerpunkt der Höhe nach
$A$	Faktormatrix, sog. Faktorenmuster
$B$	größte Breite des Unterwasserschiffes
$B$	Index für Bulb = Wulst
$C$	Korrelationsmatrix zwischen schiefwinkligen Faktoren
$C_B$	Blockkoeffizient
$C_{FO}$	Reibungswiderstandsbeiwert der äquivalenten ebenen Platte
$C_M$	Hauptspantvölligkeit
$C_P$	Schärfegrad
$C_R$	Restwiderstandsbeiwert
$C_T$	Gesamtwiderstandsbeiwert
$C_V$	Beiwert des Zähigkeitswiderstandes
$C_W$	Beiwert des Wellenbildungswiderstandes

$C_{WP}$	Völligkeitsgrad der Wasserlinienfläche
E	Index für Entrance
F	statistische Testgröße; F-Verteilung
$F_n$	Froudezahl
I	Einheitsmatrix
L	Schiffslänge zwischen den Loten
M	Index für Mittelschiff
P	Matrix der Beobachtungswerte nach der Faktorenanalyse
R	Index für Run
R	Korrelationsmatrix
$R_n$	Reynoldszahl
$R_{F0}$	Reibungswiderstand der äquivalenten ebenen Platte
$R_R$	Restwiderstand
$R_T$	Gesamtwiderstand
$R_V$	Zähigkeitswiderstand
$R_W$	Wellenbildungswiderstand
S	Kovarianzmatrix
S	Schiffsoberfläche
T	Konstruktionstiefgang
T'	mittlerer Tiefgang für alle Beladungszustände außer voll beladen
V	Symbol für die Vertrauensgrenzen
V	Schiffsgeschwindigkeit

$V_M$	Modellgeschwindigkeit
$W$	Symbol für Wahrscheinlichkeit
$X$	Wert einer unabhängigen Variablen eines neuen Entwurfs
$Y$	aus der Regressionsgleichung berechneter Wert der abhängigen Variablen eines neuen Entwurfs
$Z$	Matrix der standardisierten Variablenwerte
$\alpha$	Wert der statistischen Wahrscheinlichkeit
$\beta$	Regressionskoeffizient der Grundgesamtheit
$\epsilon$	Fehlerglied, Differenz zwischen gegebenem und aus der Regressionsgleichung berechnetem $y$ -Wert
$\lambda$	Eigenwert der Koeffizientenmatrix in der Hauptachsenmethode
$\mu$	Mittelwert der Daten einer Grundgesamtheit
$\nu$	Freiheitsgrad
$\rho$	spezifische Dichte
$\sigma^2$	Streuung der Daten einer Grundgesamtheit
$\chi^2$	statistische Testgröße; $\chi^2$ - Verteilung
$\Delta$	Masse der Verdrängung
$\nabla$	Verdrängungsvolumen

Literaturverzeichnis

- [1] Doust, D.J. and O'Brien, T.P.:  
Resistance and Propulsion of Trawlers,  
Trans. N.E.C.I., Vol. 75, 1958/1959
  
- [2] Doust, D.J.:  
Ship Design and Power Estimating Using Statistical Methods,  
Norwegian Ship Model Experiment Tank  
Publication No. 70, Dec. 1962
  
- [3] Sabit, A.S.:  
Regression Analysis of the Resistance Results of the  
B.S.R.A. Series,  
ISP Vol. 18, Jan. 1971
  
- [4] Sabit, A.S.:  
An Analysis of the Series 60 Results,  
Part I: Analysis of Forms and Resistance Results,  
ISP Vol. 19, March 1972
  
- [5] Swift, P.M.; Nowacki, H. and Fischer, J.P.:  
Estimation of Great Lakes Bulk Carrier Resistance  
Based on Model Test Data Regression,  
Marine Technology, Oct. 1973
  
- [6] Oortmerssen, G. van:  
A Power Prediction Method and its Application to Small Ships,  
ISP Vol. 18, Nov. 1971
  
- [7] Tagano, H.:  
Prediction of the Wave Resistance of Ships by  
Statistical Analysis,  
Mitsubishi Technical Bulletin No. 90, June 1974

- [17] International Towing Tank Conference:  
Standard Symbols 1971, B.S.R.A. Naval Architecture  
Technical Memorandum, Aug. 1971 sowie  
Collatz, Dr. G.:  
International vereinbarte Buchstabensymbole und Bezeichnungen  
auf dem Gebiet der Schiffshydrodynamik,  
Schiff + Hafen Jahrg. 27, Okt. 1975, Heft 10
- [18] Westervelt, F.H.:  
A Study of Automatic System Simulation Programming and the  
Analysis Using an Internally Stored Program Computer,  
Thesis, University of Michigan Industry Program, IP470, Oct. 1960
- [19] Todd, Dr. F.H. and Pien, Dr. P.C.:  
Series 60 - The Effect upon Resistance and Power of Variation  
in LCB Position,  
SNAME Vol. 64, 1956
- [20] Todd, Dr. F.H.; Stuntz, G.R. and Pien, Dr. P.C.:  
Series 60 - The Effect upon Resistance and Power of Variation  
in Ship Proportions,  
SNAME Vol. 65, 1957
- [21] Hughes, G.:  
An Analysis of Ship Model Resistance into Viscous and Wave  
Components, Part I, II, III.  
Trans R.I.N.A. Vol. 108, 1966
- [22] Prohaska, C.W.:  
A Simple Method for the Evaluation of the Form Faktor and the  
Low Speed Wave Resistance,  
Proceedings of the 11<sup>th</sup> International Towing Tank Conference,  
Tokyo, Oct. 1966
- [23] Report of the Performance Committee,  
13<sup>th</sup> International Towing Tank Conference 1972,  
Berlin/Hamburg

- [8] Fairlie-Clarke, A.C.:  
Regression Analysis of Ship Data,  
ISP Vol. 22, July 1975
- [9] Bennett, C.A. and Franklin, N.L.:  
Statistical Analysis in Chemistry and the Chemical Industry,  
John Wiley & Sons, Inc. New York, London, Sidney 1954
- [10] Linder, A.:  
Statistische Methoden,  
Birkhäuser Verlag Basel und Stuttgart, 1960
- [11] Graf/Henning/Stange:  
Formeln und Tabellen der mathematischen Statistik,  
Springer Verlag Berlin/Heidelberg/New York,  
2. Auflage 1966
- [12] Brownlee, K.A.:  
Statistical Theory and Methodology in Science and Engineering,  
John Wiley & Sons, Inc. New York, London, Sidney, 2<sup>nd</sup> Ed. 1965
- [13] Überla, K.:  
Faktorenanalyse,  
Springer-Verlag Berlin/Heidelberg/New York,  
2. Auflage 1971
- [14] Ralston, A. and Wilf, H.S.:  
Mathematische Methoden für Digitalrechner,  
R. Oldenbourg Verlag München-Wien 1967
- [15] Guldhammer, H.E. and Harvald, S.A.:  
Ship Resistance. Effect of Form and Principal Dimensions,  
Akademisk Forlag Copenhagen 1965
- [16] Inui, T.:  
Simplified Method of Calculation for Wave-Making Resistance,  
SNAJ, 60<sup>th</sup> Anniversary Series, Vol. 2 (1957), p. 197

## 1. Einleitung

Die Aufgabe, ein Schiff auf rein theoretischer Grundlage zu entwerfen, ist bis heute ungelöst. An vielen Stellen des Schiffsentwurfs werden empirische Hilfsmittel benötigt, und es ist abzusehen, daß sie noch längere Zeit unentbehrlich bleiben werden. Ein wichtiges Gebiet, auf dem man sich in der Frühphase eines Entwurfs empirischer Unterlagen bedienen muß, ist das der Widerstands- und Leistungsprognose. Wegen der hohen Geldsummen, die hier auf dem Spiel stehen, vertraut man den ermittelten Werten allerdings nicht allzu sehr, sondern versucht stets, noch vor einem Vertragsabschluß größere Sicherheit durch Modellversuche zu erlangen.

Die Summe der Ergebnisse solcher Modellversuche wiederum bildet den willkommenen und notwendigen Nachschub für die Auffrischung älterer empirischer Unterlagen. Besonders sinnfällig und notwendig wird eine solche Auffrischung dann, wenn durch Wandel oder Neueinführung von Schiffstypen neue Bereiche von Schiffsformen erschlossen wurden, für die ältere Prognoseunterlagen nur "weiße Flecke" bereit halten. Eine solche Entwicklung fand gerade in der jüngsten Vergangenheit statt, als unvorhergesehen ein Bedarf an großen völligen Tankschiffen und Bulkern einerseits, schnellen, schlanken Containerschiffen andererseits einsetzte. Umfangreiche Modellversuche wurden durchgeführt, und die Ergebnisse lagen bald darauf vor. Es war eine wichtige und dringende Aufgabe, diese Daten der Allgemeinheit zugänglich zu machen, das heißt daraus Entwurfsunterlagen zu erstellen. Die vorliegende Arbeit ist das Ergebnis eines solchen Prozesses, der sich allerdings nicht nur auf die genannten neuen Schiffstypen beschränkte, sondern auch Schiffe altbekannter Form einbezog.

Für die Arbeit wurden die Ergebnisse von Widerstandsversuchen der Hamburgischen Schiffbau-Versuchsanstalt (HSVA) aus den Jahren Ende 1967 bis Anfang 1973 gesichtet und, soweit bestimmte Voraussetzungen erfüllt waren, verwendet. Das aus zahlreichen Einzeluntersuchungen der Industrie zusammengesetzte Datenmaterial war naturgemäß unsystematisch, d.h. die Verteilung der Formparameter entsprach keineswegs der Regelmäßigkeit etwa der Taylor- oder der Series 60-Modellreihe. Bei der Auswertung war Rücksicht zu nehmen auf diese Unsystematik. Ferner war zu fordern, daß dem modernen Hilfsmittel Computer Rechenschaft gezollt wurde. Beide Forderungen ließen sich erfüllen, indem

die Daten mit Hilfe statistischer Methoden untersucht und analysiert und die Ergebnisse in die computergeeignete Form von Formeln gekleidet wurden. Das hierfür benutzte Werkzeug war die Regressionsanalyse.

Schon mehrfach ist diese Methode zur Analyse des Widerstandes eingesetzt worden, zuerst in der Arbeit von DOUST und O'BRIEN (1959) über Fischkutter. In der Praxis jedoch werden die damit erzielbaren Ergebnisse offenbar nur ungern angewendet; nach wie vor spielen graphische Interpolationsmethoden die bedeutendere Rolle. Zum Teil mag das an den speziellen, heute weniger begehrten Schiffsformen liegen, die analysiert wurden; SABIT z.B. hat Regressionen der SERIES 60 und der BSRA-Serie vorgelegt. Ein weiterer Grund aber mag auch in dem verbreitet anzutreffenden, leider unzureichendem Verständnis für statistische Methoden zu suchen sein. Es muß als ein Mangel der bisher veröffentlichten Arbeiten angesehen werden, daß sie die Technik der statistischen Verfahren nicht ausführlich genug darstellung und damit eher Mißtrauen als die verdiente Wertschätzung hervorrufen.

Auf zahlreiche wichtige Fragen gehen diese Arbeiten außerdem nur kurz oder gar nicht ein. Diese Fragen können wie folgt umrissen werden: Optimale Auswahl der Variablen, mit denen der Widerstand beschrieben werden soll; Prüfung auf Unabhängigkeit der Variablen, damit möglichst viele unterschiedliche Informationen für die Beschreibung zur Verfügung stehen; Ermittlung der bestgeeigneten Funktion, in der eine Variable mit dem Widerstand zusammenhängt; Test auf Zufälligkeit der Daten, um herauszufinden, ob das Regressionsergebnis vom Einfluß einer Variablen mit allzu häufig vorkommendem gleichem oder sehr ähnlichem absolutem Wert geprägt ist; Ermittlung der Vertrauensgrenzen als direktes Maß für die Genauigkeit berechneter Widerstände. All dies ist jedoch von großer Bedeutung, insbesondere dann, wenn die Daten aus einem unsystematischen Kollektiv wie Routinemeßergebnissen einer Versuchsanstalt stammen. In diesem Bericht werden daher die Grundlagen der statistischen Theorie und die Tests beschrieben, die im Zusammenhang mit der Regressionsanalyse zu beachten sind. Dabei wird die Mathematik nur in dem Ausmaß wiedergegeben, wie es zum Verständnis der Zusammenhänge unbedingt erforderlich ist; bezüglich Einzelheiten, z.B. der programmtechnischen Durchführung des Regressionsverfahrens, kann auf die zahlreichen Literaturquellen verwiesen werden.

Auswahl der Widerstandsdaten und Untersuchungen auf die Zufälligkeit ihrer Anordnung werden diskutiert. Ausführlich und, soweit es gegenwärtig zu übersehen ist, ohne Vorbild auf schiffbaulichem Gebiet, werden Untersuchungen über die gegenseitige Unabhängigkeit der Variablen durchgeführt, die für die Korrelation mit dem Widerstand benutzt werden. Desgleichen wird eingehend über die Suche nach den bestgeeigneten Funktionen berichtet, die die beschreibenden Variablen und den Widerstand miteinander verbinden.

Das Ergebnis der Arbeit sind Regressionsgleichungen für den Restwiderstandsbeiwert  $C_R$  sowie, aus einer angeschlossenen Untersuchung hervorgegangen, für den Formfaktor  $k$ , ferner umfangreiches statistisches Material zur Beurteilung der Berechnungsgenauigkeit. Zur Anwendung der Formeln ist die Aufmessung der Konstruktionsspannen eines Schiffsentwurfs erforderlich sowie die Ermittlung einiger neu eingeführter, relativ ungeläufiger Formparameter. Die berechneten Zahlenwerte für  $C_R$ , die für bestimmte feste Froudezahlen erhalten werden, bilden im allgemeinen keinen strakenden Verlauf, sondern müssen durch einen vermittelnden Kurvenzug ausgeglichen werden. Insgesamt ist der Aufwand, den das Verfahren erfordert, wohl größer als bei herkömmlichen Verfahren zur Bestimmung des Widerstands.

## 2. Überblick über Regressionsanalysen des Widerstandes

Die Regressionsanalyse ist schon mehrfach auf den Widerstand angewendet worden. Ein Studium der bisher erschienenen Arbeiten läßt erkennen, daß, vielleicht verständlich aus der erstmaligen Beschäftigung mit einer neuen, für den Computer geeigneten Technik, diese selbst in den Vordergrund rückte, ihren statistischen Grundlagen aber unzureichende Beachtung geschenkt wurde. Das Ergebnis sind häufig Regressionsgleichungen, die nur unsichere Prognosen gestatten.

Mehrere Gleichungen für den Widerstand von Fischkuttern und Handelsschiffen wurden von DOUST und Mitarbeitern angegeben [1,2]. Sie enthalten Hauptabmessungsverhältnisse, Blockkoeffizienten und einen Wasserlinienwinkel als unabhängige Variable, die außer in linearer Form als einfache Potenzfunktionen auftreten. Gemeinsames Merkmal der Ergebnisgleichungen ist die hohe Zahl von Gliedern, die von rund 30 bis 86 reicht. Die etwa 130 gegebenen Fischkutterdaten werden zwar bis auf einen mittleren Fehler von drei bis fünfzehn Prozent angenähert, doch kann die gleiche Genauigkeit nicht für Prognosewerte erwartet werden. Signifikanzuntersuchungen und Angabe der Vertrauensgrenzen fehlen. Letzteres gilt ohne Ausnahme auch für sämtliche auf DOUST folgende Regressionsanalysen.

Nach sehr gründlichen Überlegungen über Wechselwirkungen der Hauptabmessungsverhältnisse miteinander untersucht SABIT mehrere Gruppierungen entsprechender Variablen und ihrer Kreuzprodukte in Regressionsanalysen des Widerstandes der BSRA-Standardserie [3] und der SERIES 60 [4]. Die endgültige Entscheidung zugunsten einer der Gruppierungen mit 16 Variablen scheint durch Anwendung von Signifikanztests zu fallen, doch wird nicht erklärt, welche es sind. Die Annäherung der 40 bis 67 gegebenen Widerstände durch die Regressionsgleichungen ist im Mittel besser als bei DOUST, und auch die Vorhersagegenauigkeit für Entwürfe, deren Formen denen der Serien entsprechen, dürfte hoch sein. Es sollte jedoch nicht übersehen werden, daß sich Serien leichter analysieren lassen als unsystematische Kollektive.

Unzureichende, soweit vorhanden möglicherweise auch falsche Anwendung statistischer Signifikanzmethoden kennzeichnen Regressionsanalysen des Widerstandes von Great-Lakes-Massengutschiffen von SWIFT et alii [5]. Sie teilen für eine Stichprobe von 50 Schiffen Gleichungen mit 12 bis 16 Gliedern mit,

deren Aufbau weitgehend dem Ansatz von DOUST entspricht. Die statistische Verlässlichkeit der Ergebnisse steht in Frage.

Während in den erwähnten Arbeiten der Widerstand an festen Froudezahlen bzw. an dem in angelsächsischen Ländern gebräuchlichen Geschwindigkeitsparameter  $V/\sqrt{L}$ ,  $V$  = Geschwindigkeit in Knoten,  $L$  = Länge zwischen den Loten in Fuß, analysiert wird, macht OORTMERSSEN [6] einen geschwindigkeitsabhängigen Widerstandsansatz. Er besteht aus fünf auf theoretischer und experimenteller Grundlage abgeleiteten Gliedern, deren Koeffizienten als Funktionen der Schiffsform angeschrieben werden und in Regressionsanalysen zu ermitteln sind. Die endgültige Gleichung enthält etwa 50 Terme. Die Beschreibung der gegebenen 93 Widerstandskurven von kleinen Schiffen wie Schleppern und Fischkuttern kann nicht ganz befriedigen. OORTMERSSENS Leistung liegt in der Entwicklung des theoretischen Ansatzes, doch der statistische Aspekt der Regressionsanalyse kommt zu kurz.

Ähnlich OORTMERSSEN geht auch TAGANO [7] in einer Analyse schneller Handelsschiffe von einer theoretischen, geschwindigkeitsabhängigen Widerstandsformel aus. Die Koeffizienten werden als Funktion von Formvariablen aufgefaßt, für die im Unterschied zu allen anderen Arbeiten Aufmaße der Spantarealkurve verwendet werden, um lokale Formeinflüsse besser zu berücksichtigen. Durch Umformung wird der geschwindigkeitsabhängige Ansatz jedoch soweit verstümmelt, daß er nur noch bei konstanter Froudezahl ausgewertet werden kann. Der Bericht nennt nicht die Zahlenwerte der Regressionskoeffizienten, und an keiner Stelle werden Nachweise statistischer Signifikanz geführt.

Soweit bekannt, wurden Regressionsanalysen im Schiffbau auch auf Leistungsdaten sowie Probefahrts- und Reisedaten angewendet. Eine umfassende Literaturübersicht des gesamten Komplexes findet sich bei FAIRLIE-CLARKE [8], der außerdem eine ausgezeichnete Darstellung der statistischen Grundlage der Regressionsanalyse gibt.

### 3. Statistische Methoden im Zusammenhang mit der Regressionsanalyse

#### 3.1. Allgemeines

Statistische Methoden sind überall dort am Platze, wo aus einer begrenzten Zahl von Daten, z.B. Versuchsergebnissen, Schlüsse gezogen werden sollen. Von höchstem Nutzen sind sie dann, wenn die Daten streuen. Unter bestimmten Annahmen über die Natur der Streuung ermöglichen sie Aussagen über die Verlässlichkeit der gezogenen Schlüsse bzw. über das Risiko, daß ein falscher Schluß gezogen wird.

Zu solchen Aussagen gelangt man durch Beurteilung der Verteilung, in der die Daten vorliegen. Experimentelle Daten können vielen Verteilungstypen angehören. Doch diese Typen haben gemeinsame Eigenschaften, die zum Zweck der Beschreibung und Vorhersage genutzt werden.

Die in dieser Arbeit verwendeten Daten werden als Stichproben aus den entsprechenden normal verteilten Grundgesamtheiten angesehen. Ob die Elemente einer Stichprobe der zugehörigen Grundgesamtheit zufällig entnommen sind, wird im Einzelfall geprüft.

Das Element einer Stichprobe weist im allgemeinen eine oder mehrere gleichzeitig beobachtete Eigenschaften auf. Eine spezielle Eigenschaft wird im folgenden mit Merkmal bezeichnet, andere Ausdrücke hierfür sind Parameter oder Variable. Merkmale eines Schiffes sind z.B. sein Widerstand, sein Längenbreitenverhältnis oder seine metazentrische Höhe. Jedes Merkmal besitzt eine eigene Grundgesamtheit.

Die von GAUSS für Meßfehler astronomischer Beobachtungen gefundene Normalverteilung ist eine Verteilungsform, die bei sehr vielen Grundgesamtheiten festgestellt werden kann. Aus der Annahme, daß Normalverteilung der Grundgesamtheit vorliegt, leiten sich Tests auf Verlässlichkeit statistischer Schlußfolgerungen ab, die aus Stichproben gezogen werden. Will man mit Hilfe der Stichprobenergebnisse eine Prognose für ein nicht der Stichprobe angehörendes Objekt abgeben, so muß dies Objekt natürlich der Grundgesamtheit angehören; insbesondere müssen seine Merkmale auch innerhalb der für die Stichprobenelemente geltenden Grenzen liegen. Das ist nötig, weil eine Stichprobe nichts über die Teile der Grundgesamtheit aussagt, die außerhalb der Stichprobengrenze liegen. Man darf jedoch extrapolieren, wenn das Verhalten

der Grundgesamtheit jenseits der Stichprobengrenzen theoretisch oder aus Erfahrung bekannt ist.

Im folgenden sollen die für die Regressionsanalyse wichtigsten statistischen Größen und Begriffe und die grundlegenden Gedanken der Testverfahren beschrieben werden. Hierdurch soll es möglich werden, richtigen Gebrauch von den Ergebnissen dieser Arbeit zu machen. Die Darstellung stützt sich im wesentlichen auf das Buch von BENNET/FRANKLIN [9], doch werden Einzelheiten auch [10 bis 14] entnommen.

### 3.2. Beschreibende statistische Größen

#### 3.2.1. Mittelwert, Streuung, Standardabweichung

Größere Datenmengen sind gewöhnlich unübersichtlich. Man braucht wenige einfache Größen, die eine möglichst sichere Information über die Gestalt der Stichprobe geben. Solche Größen sind der Mittelwert, die Streuung (Varianz) und die Standardabweichung.

Es liegen  $n$  Beobachtungen des Merkmals  $x$  vor. Ihr Mittelwert ist definiert durch

$$\bar{x} = \frac{\sum_{i=1}^n (x_i)}{n}, \quad (1)$$

die Streuung durch

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{1}{n-1} \left[ \sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n} \right] \quad (2)$$

und die Standardabweichung durch

$$s = \sqrt{s^2}. \quad (3)$$

Der Mittelwert  $\bar{x}$  kennzeichnet den Schwerpunkt der Stichprobenwerte. Ein großer Wert  $s^2$  spricht für weit um den Mittelwert streuende Einzelwerte.

Die Standardabweichung, deren Dimension mit der der Stichprobenwerte übereinstimmt, hat folgende anschauliche Bedeutung: vorausgesetzt, daß die  $x_i$  einigermaßen normal verteilt sind und  $n > 50$  ist, liegen im Intervall

$[\bar{x} + s ; \bar{x} - s]$	etwa 2/3 der Beobachtungen,
$[\bar{x} + 2s ; \bar{x} - 2s]$	etwa 19/20 der Beobachtungen,
$[\bar{x} + 3s ; \bar{x} - 3s]$	etwa 997/1000 der Beobachtungen.

Ein anderer, seltener gebrauchter Kennwert ist der Median. Er ist der Wert, der in der Mitte der nach aufsteigender Größe geordneten Stichprobendaten steht. Bei den Zahlen von 1 bis 11 ist 6 der Median.

### 3.2.2. Die Kovarianz

Die Streuung oder Varianz  $s^2$  sagt etwas aus über die Lage von Daten in Bezug auf ihren Mittelwert. Die Kovarianz enthält Informationen über die mögliche lineare Beziehung zwischen Paaren von Daten. Sie ist definiert durch

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad (4)$$

wo  $(x_i, y_i)$  ein Datenpaar und  $\bar{x}, \bar{y}$  die jeweiligen Mittelwerte bezeichnen. Ist die lineare Beziehung zwischen  $x$  und  $y$  eng, so sind große Differenzen  $(x_i - \bar{x})$  mit großen Differenzen  $(y_i - \bar{y})$  verbunden. Das Kreuzprodukt der Differenzen wird einen hohen Wert annehmen.

### 3.2.3. Der Korrelationskoeffizient

Ein dimensionsloses Maß für die Strenge einer linearen Beziehung zwischen den Daten aus zwei Merkmalsgebieten ist der Korrelationskoeffizient. Er ist definiert als

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{s_{xy}}{s_x \cdot s_y} \quad (5)$$

Dies ist die Kovarianz, umgewandelt in ein Relativmaß mittels Division durch geeignete Maßstabsfaktoren. Der Korrelationskoeffizient kann Werte zwischen -1 und +1 annehmen. Sein Vorzeichen ist abhängig davon, ob  $y$  proportional zu  $x$  steigt oder fällt. Hohe Werte  $r$  sprechen für enge Beziehungen zwischen den Merkmalen. Im Fall  $r = 0$  bezeichnet man die beiden Merkmale als unkorreliert. Insbesondere sind sie unkorreliert, wenn sie voneinander unabhängig sind. Allerdings darf umgekehrt aus Unkorreliertheit nicht auf Unabhängigkeit geschlossen werden.

### 3.2.4. Der multiple Korrelationskoeffizient

Gelingt es, eine vielfache Beziehung zwischen einem bestimmten Merkmal  $y$  (Zielgröße, abhängige Variable) und  $p$  anderen Merkmalen  $x_j$  (unabhängige Variable) in der Form

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_p x_p \quad (6)$$

herzustellen, wo  $(y_i; x_{1i} \text{ bis } x_{pi}; i = 1 \text{ bis } n)$  zusammengehörende Wertetupel bezeichnen, so verwendet man als Maß für die Streuung der Beziehung den multiplen Korrelationskoeffizienten, dessen Quadrat lautet

$$r^2 = 1 - \frac{(n-p)s_{y;x_1 \dots x_p}^2}{(n-1) s_y^2} \quad (7)$$

Hierin berechnet sich  $s_{y;x_1 \dots x_p}^2$  aus den Differenzen zwischen den gegebenen und den durch Gl. (6) berechneten  $y$ -Werten;  $s_y^2$  ist die Streuung der gegebenen  $y$ -Werte. Die beiden Klammerausdrücke sind notwendige Korrekturen, um erwartungstreue Schätzungen für die beiden Streuungen zu erhalten. Zur Definition der Erwartungstreue muß auf die Literatur verwiesen werden.

$r^2$  kann Werte zwischen Null und 1 annehmen. Bei gegebener Stichprobe ist lediglich  $s_{y;x_1 \dots x_p}^2$  durch die Güte einer Näherung nach Gl. (6) zu beeinflussen. Je besser die berechneten und die gegebenen  $y$ -Werte übereinstimmen, desto kleiner wird die Streuung  $s_{y;x_1 \dots x_p}^2$ . Damit nähert sich  $r^2$  dem Wert 1. Man kann den multiplen Korrelationskoeffizient als Maß für die Erklärung der Varianz der Zielgröße durch eine Näherung mittels unabhängiger Variabler auffassen.

### 3.3. Statistische Begriffe und Testgrößen

Der statistische Begriff der Wahrscheinlichkeit und einige weitere Begriffe und statistische Tests sollen anhand einer Untersuchung des Mittelwertes einer Stichprobe erläutert werden.

#### 3.3.1. Wahrscheinlichkeit

Angenommen es gibt eine Grundgesamtheit, deren Elemente zur Hälfte ein bestimmtes Merkmal tragen, zur Hälfte nicht. Dann weist die Wahrscheinlichkeit dafür, daß Teile der Elemente von zufälligen Stichproben aus dieser Grundgesamtheit dieses Merkmal besitzen, eine bestimmte Verteilung (die binomische Verteilung) auf. Im Gesetz dieser Verteilung spielt die Größe  $n$  der zufälligen Stichproben eine wichtige Rolle. Für jede einzelne Stichprobe hat die Wahrscheinlichkeit einen Zahlenwert, und es ist in das Ermessen des Untersuchenden gestellt, ob er dieses Ergebnis für mit der Merkmalsgesetzmäßigkeit der Grundgesamtheit (die auch eine Hypothese sein kann) verträglich hält oder nicht. Es muß dann eine Grenze gezogen werden, die dem Sicherheitsbedürfnis oder dem Risiko, das man eingehen will, gerecht wird. Diese Sicherheitsschwelle wird häufig bei technischen Anwendungen bei einer Wahrscheinlichkeit von  $1/20$  oder  $5\%$  gesetzt. Ergibt die Untersuchung der Stichprobe die Wahrscheinlichkeit, eine ebenso große oder größere Abweichung von der Hypothese zu erhalten, von gleich oder mehr als  $1/20$ , so wird man die Stichprobe für in Einklang mit der für die Grundgesamtheit aufgestellten Hypothese halten.

Ist  $\mu$  der unbekannte Mittelwert von Zahlen einer Grundgesamtheit,  $\bar{x}$  der Mittelwert einer zufälligen Stichprobe aus dieser Grundgesamtheit, so kann man unter Beachtung der gewählten Sicherheitsschwelle Grenzen, sogenannte Vertrauensgrenzen, angeben, außerhalb derer  $\mu$  mit der Wahrscheinlichkeit  $\alpha$  liegt. Zu diesem Zweck wird die statistische Größe  $t$  benötigt.

#### 3.3.2. Die statistische Größe $t$

Man betrachtet den Ausdruck

$$t = \frac{\bar{x} - \mu}{s} \sqrt{n}, \quad (8)$$

der die Abweichung  $(\bar{x} - \mu)$  des Mittelwertes  $\bar{x}$  einer Stichprobe von dem Mittelwert  $\mu$  der Grundgesamtheit mit der Standardabweichung  $s/\sqrt{n}$  des Stichproben-

mittelwertes in Beziehung setzt. Die Verteilung der Größe  $t$ , berechnet für viele zufällige Stichproben aus der Grundgesamtheit, ist charakteristisch; sie heißt  $t$ -Verteilung. Wenn die Grundgesamtheit der  $x$ -Werte symmetrisch ist, ist es auch die  $t$ -Verteilung. Für die Wahrscheinlichkeit, daß  $t$  einen Wert zwischen  $t$  und  $t+dt$  annimmt, gilt ein Gesetz, dessen Kurve ebenfalls symmetrisch und glockenförmig ist. Die wesentliche Variable in diesem Gesetz ist  $n$ , und für jedes  $n$  gibt es eine eigene Kurve. Wählt man auch hier eine Sicherheitsschwelle von z.B.  $\alpha = 0.05$ , so hält man zwei Punkte  $\pm t_{\alpha}$  auf der  $t$ -Achse, außerhalb derer die Fläche unter der Kurve 5 % der gesamten Fläche ausmacht. Solche Werte  $t_{\alpha}$  sind für verschiedene Freiheitsgrade  $\nu$  (im Fall der Gl. (8) ist  $\nu = n - 1$ ; zur Definition des Freiheitsgrades s. Literatur) berechnet und tabelliert worden. Sie liegen derart vor, daß die Wahrscheinlichkeit, daß  $|t| > t_{\nu, \alpha}$ , gleich  $\alpha$  ist:

$$w ( |t| > t_{\nu, \alpha} ) = \alpha. \quad (9)$$

Durch Umformung der Gl. (8) können nun Vertrauensgrenzen um  $\bar{x}$  berechnet werden, außerhalb derer der Wert  $\mu$  mit der Wahrscheinlichkeit  $\alpha$  liegt:

$$\bar{x} \pm t_{\nu, \alpha} s / \sqrt{n}. \quad (10)$$

Legt man die Sicherheitsschwelle statt auf 0,05 auf 0,01, so wird  $t_{\nu, \alpha}$  wachsen, und damit weiten sich die Vertrauensgrenzen. Will man ein kleineres Risiko laufen,  $\mu$  zu verfehlen, so muß man eben die Grenzen, innerhalb derer es liegen kann, erweitern.

In ähnlicher Weise wie für den Mittelwert  $\bar{x}$  können Vertrauensgrenzen für die meisten aus Stichproben berechneten Werte angegeben werden. Man sagt z.B.,  $\bar{x}$  sei die bestmögliche Schätzung von  $n$  aus der Stichprobe, und die Vertrauensgrenzen sind ein Maß für die Genauigkeit der Schätzung. Wo immer möglich sollten die Vertrauensgrenzen zu den mittels statistischer Methoden berechneten Werten angegeben werden.

### 3.3.3. Signifikanztest

Die t-Verteilung hat eine besondere Bedeutung für die Berechnung von Vertrauensgrenzen erlangt, des weiteren aber auch für Testzwecke. Dies soll an dem folgenden Beispiel gezeigt werden.

Aus einer Grundgesamtheit seien viele Stichproben (Anzahl sei m) gezogen worden, deren Werte durch den Ansatz

$$y = b_{j0} + b_{j1}x ; j = 1 \text{ bis } m \quad (11)$$

beschrieben werden. Die m Werte  $b_{j0}$  und  $b_{j1}$  werden z.B. durch Ausgleichsrechnung bestimmt. Es sei ferner angenommen, daß die Werte  $\beta_0$  und  $\beta_1$ , die für die Grundgesamtheit gelten, bekannt sind.

Man bildet das Verhältnis der wirklichen Abweichungen ( $b_{j1} - \beta_1$ ) zur Standardabweichung der  $b_{j1}$ , welches (unter Verzicht auf die mathematische Ableitung) für eine einzelne Stichprobe mit den Werten  $x_i$ ,  $i = 1$  bis  $n$  gegeben ist durch

$$t^* = \frac{(b_1 - \beta_1) \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{s_{yx}}, \quad (12)$$

wo  $s_{yx} = \sqrt{\sum_{i=1}^n (y_i - \tilde{y}_i)^2 / (n-2)}$ ;  $\tilde{y}_i$  = aus der Geraden berechnete Näherungswerte der gegebenen Werte  $y_i$ . Es läßt sich zeigen, daß die Verteilung der Gesamtheit der  $t^*$ -Werte gerade der t-Verteilung entspricht, wobei der Freiheitsgrad  $\nu = n-2$  verwendet werden muß.

Man benutzt diese Übereinstimmung einerseits, um Vertrauensgrenzen für die Steigung  $b_1$  irgendeiner der Ausgleichsgeraden zu berechnen. Für bekanntes  $n$  und  $\alpha$  wird das tabellierte  $t_{\nu, \alpha}$  ermittelt. Löst man Gl. (12) nach  $\beta_1$  auf und setzt  $t_{\nu, \alpha}$  für  $t^*$  ein, so erhält man die Vertrauensgrenzen

$$b_1 \pm \frac{t_{\nu, \alpha} s_{yx}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad \text{mit } \nu = n-2. \quad (13)$$

Andererseits kann man auch die sogenannte "Signifikanz" von  $b_1$  testen. Hierzu nimmt man an, daß  $\beta_1 = 0$  sei (irgendein anderer, z.B. ein theoretisch erwarteter Wert für  $\beta_1$  kann ebenso eingesetzt werden). Man berechnet  $t^*$  aus Gl. (12) und vergleicht es mit dem für  $\nu = n - 2$  und der gewählten Wahrscheinlichkeit tabellierten Wert  $t_{\nu, \alpha}$ . Ist  $t^* > t_{\nu, \alpha}$ , so ist  $b_1$  "signifikant" von Null verschieden (die Hypothese  $\beta_1 = 0$  ist mit der Wahrscheinlichkeit  $\alpha$  widerlegt); anderenfalls ist  $b_1$  "nicht signifikant"

Mit Hilfe entsprechender Überlegungen und Vergleiche können Vertrauensgrenzen auch für neue, mittels Gl. (11) interpolierte  $y$ -Werte angegeben werden. Dies ist von erheblicher Bedeutung, denn diese Vertrauensgrenzen geben dem Benutzer von Interpolationsgleichungen wie der im Beispiel gewählten Ausgleichsgeraden ein unmittelbares Bild von der Genauigkeit einer Prognose. Auf die Mitteilung der Formel an dieser Stelle wird verzichtet, sie findet sich im Kontext der Regressionsanalyse (Abschnitt 3.4.).

#### 3.3.4. Der $\chi^2$ -Test auf Normalverteilung

Der Gebrauch der statistischen Größe  $t$  zu Testzwecken wie im vorigen Kapitel unterliegt gewissen Beschränkungen. Am bedeutsamsten ist die Forderung, daß die zu testende Größe annähernd normalverteilt ist. Um dies zu überprüfen, eignet sich in manchen Fällen der  $\chi^2$ -Test.

Zur Definition der  $\chi^2$ -Verteilung sei auf die Literatur verwiesen. Werte für  $\chi^2$  liegen in Abhängigkeit vom Freiheitsgrad  $\nu$  und der Wahrscheinlichkeit  $\alpha$  tabelliert vor. Hier nun werden diese Werte dazu benutzt, um zu prüfen, ob eine Anzahl von Beobachtungen aus einer Normalverteilung mit dem Mittelwert 0 und einer Standardabweichung  $\sigma$ , die durch  $s$  abgeschätzt wird, gekommen sein könnte.

Zur praktischen Durchführung teilt man die beobachteten Werte und die Normalverteilung in Klassen ein. Zweckmäßig ist z.B. die Wahl von Flächen gleicher Größe unter der Verteilungskurve. Wenn z.B. jede Klasse 10 % der Gesamtfläche unter der Normalverteilungskurve enthalten soll, so müssen die Klassengrenzen mit der Standardabweichung  $s$  aus

$$0; \pm 0.253s; \pm 0.524s; \pm 0.842s; \pm 1.282s; \pm \infty$$

berechnet werden. Die theoretische Anzahl von Beobachtungen pro Klasse bei

einer Stichprobe vom Umfang  $n$  aus einer normalverteilten Grundgesamtheit ist dann einfach  $f = n/10$ . Von dieser theoretischen Häufigkeitsverteilung wird eine normale Stichprobe gewöhnlich abweichen. Man kann sich natürlich auf graphischem Wege eine qualitative Vorstellung von der Abweichung verschaffen. Zu einer quantitativen Abschätzung der Gültigkeit der Annahme, daß die vorgelegte Stichprobe aus einer normal verteilten Grundgesamtheit verhilft aber der  $\chi^2$ -Test. Man berechnet

$$\chi^2 = \sum_{k=1}^r \frac{(N_k - f_k)^2}{f_k}, \quad (14)$$

wo  $N_k$  die pro Klasse tatsächlich ausgezählte Anzahl von Beobachtungen und  $f_k$  ihr theoretischer Wert ist, und  $r$  die Anzahl der Klassen bezeichnet. Das so berechnete  $\chi^2$  vergleicht man mit dem tabellierten Wert, wobei die Zahl der bei 10 Klassen zu berücksichtigenden Freiheitsgrade  $\nu = 7$  ist. Eine Art der möglichen Aussagen ist die, daß die gemachte Annahme verworfen werden sollte, wenn  $\chi^2 > \chi^2_{\nu, \alpha}$  (unter Verwendung einer gewählten Wahrscheinlichkeit  $\alpha$ ). Einen anschaulichen Zahlenwert zur Abschätzung der Situation erhält man aber dann, wenn man den Wert für  $\alpha$  bestimmt, der dem berechneten  $\chi^2$  entspricht.

### 3.3.5. Weitere Prüfungen auf Normalverteilung

Die einfachste Art, sich ein Bild von der Gestalt einer Stichprobe zu verschaffen, ist die graphische Auftragung. Ein Beispiel zeigt Bild 1 (wie die Bilder 2 bis 4 entnommen aus [9]). Dieselben Daten wie Bild 1, nur in anderer Klasseneinteilung sowie mit der Angabe von  $\bar{x}$  und der Bereich  $\bar{x} \pm ns$ ,  $n = 1, 2, 3$ , zeigt Bild 2. Die Umrißkurven solcher Histogramme besitzen häufig zwar die Form von Glockenkurven, doch genügt dies keineswegs als Nachweis der Übereinstimmung mit der ebenfalls glockenförmigen Normal- oder Gaußverteilung.

Ein graphischer Nachweis für Übereinstimmung oder Abweichung von der Normalverteilung kann mit Hilfe der Summenhäufigkeitskurve geführt werden. Um sie zu erhalten, addiert man die in Prozent aller Stichprobenwerte ausgedrückte Häufigkeit der Daten pro Histogrammklasse von links nach rechts auf und bildet sie als Polygonzug ab (Bild 3; die Bedeutung der sogenannten Quantile ergibt sich aus ihren Ordinatenwerten). Verwendet man zur Auftragung ein Papier,

dessen Ordinate gemäß der Normalverteilungsfunktion

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] dx \quad (15)$$

geteilt ist (Wahrscheinlichkeitspapier), so erhält man bei Übereinstimmung der Stichprobenverteilung mit der Normalverteilung eine Gerade (Bild 4). Größere Abweichungen von der Geraden sind Anzeichen für schlechte Übereinstimmung mit der Normalverteilung.

Zu zahlenmäßigen Aussagen über Abweichung von der Normalverteilung gelangt man auf folgendem Weg. Es werden die semiinvarianten Größen  $k_3$  und  $k_4$  berechnet (s. Literatur), die u.a. die dritten Potenzen der aus den Daten berechneten aufsummierten Differenzen  $(x_i - \bar{x})$  bzw. die Differenzen der vierten und der zweiten Potenz dieses Ausdruckes enthalten.  $k_3$  ist Null, wenn die untersuchte Verteilung symmetrisch ist, sonst zeigt ein negatives Vorzeichen ein Übergewicht nach links, ein positives Vorzeichen ein Übergewicht nach rechts vom Mittelwert an. Mithin beurteilt die Semiinvariante  $k_3$  die Schiefe der Verteilung.  $k_4$  ist Null für die Normalverteilung. Ein errechneter positiver Wert hingegen spricht für das Vorkommen höherer Differenzen als in einer Normalverteilung, ein negativer für kleinere Differenzen. Die Semiinvariante  $k_4$  mißt die sogenannte Überhöhung der gegebenen Verteilung.

Ergänzend sei erwähnt, daß  $k_3$  und  $k_4$  als dritte bzw. vierte Semiinvariante bezeichnet werden. Die ersten beiden Semiinvarianten  $k_1$  und  $k_2$  sind gerade der Mittelwert und die Varianz der untersuchten Verteilung.

Für die dritte und vierte Semiinvariante liegen auf ihren Standardabweichungen basierende Grenzwerte tabelliert vor, innerhalb derer die für eine Stichprobe berechneten Werte für **Schiefe** und **Überhöhung** liegen sollten, wenn die Stichprobe aus einer normal verteilten Grundgesamtheit stammt. Durch Vergleich der für eine gegebene Stichprobe berechneten Werte für Schiefe und Überhöhung kann ein qualitatives Urteil über die Gestalt der Stichprobe gefällt werden, das aber im allgemeinen noch durch eine graphische Darstellung unterstützt werden muß.

### 3.3.6. Die statistische Größe F

Es soll die Frage beantwortet werden, ob gegebene Daten besser durch den Ansatz  $y = b_0 + b_1 x_1$  mit nur einer unabhängigen Variablen oder durch  $y = c_0 + c_1 x_1 + c_2 x_2$  mit zwei unabhängigen Variablen beschrieben werden.

Eine einfache Aussage läßt sich schon mit dem in 3.3.3. angegebenen Signifikanztest für die Koeffizienten  $b_1$  bzw.  $c_1$  und  $c_2$  machen. Ist  $b_1$  wohl signifikant, andererseits aber auch  $c_1$  und  $c_2$ , dann dürfte der zweite Ansatz besser sein. Sind im zweiten Ansatz entweder nur  $c_1$  oder nur  $c_2$  signifikant, ist der einfache Ansatz gemäß der ersten Gleichung wohl vorzuziehen.

Weiterhin läßt sich der F-Test anwenden. Er basiert auf dem Verhältnis

$$F = s_1^2 / s_2^2 \quad (16)$$

von zwei Varianzen  $\sigma_1^2$  und  $\sigma_2^2$ , die unabhängigen Schätzungen der Varianzen  $\sigma_1^2$  und  $\sigma_2^2$  zweier Normalverteilungen sind. Dieses Verhältnis ist nach der sog. F-Verteilung verteilt. In Tabellen ist F in Abhängigkeit von  $v_1$ ,  $v_2$  und  $\alpha$  derart gegeben, daß

$$W(F \geq F_{v_1, v_2, \alpha}) = \alpha \quad (17)$$

Zur Anwendung des F-Tests berechnet man  $F = s_{yx_1x_2}^2 / s_{yx_1}^2$  ( $s_{yx_1}^2$  und  $s_{yx_1x_2}^2$  sind die mittleren quadratischen Abweichungen der Stichprobenwerte  $y_i$  von den nach den eingangs erwähnten Ansätzen berechneten  $y$ -Werten) und vergleicht dies mit dem tabellierten Wert  $F_{1, v, \alpha}$ , wo  $v = n-1-p$  der Freiheitsgrad,  $p$  die Anzahl der  $b$ 's des "längeren" Ansatzes ist und  $\alpha$  die gewählte Wahrscheinlichkeit. Ist  $F \geq F_{1, v, \alpha}$ , so kann mit der Sicherheit  $\alpha$  angenommen werden, daß der längere Ansatz besser die gegebenen Daten annähert als der kürzere. Ist  $F < F_{1, v, \alpha}$ , so darf dieser Schluß nicht gezogen werden. Formal läßt sich feststellen, daß

$$t_{v, \alpha}^2 = F_{1, v, \alpha} \quad (18)$$

### 3.3.7. Untersuchung\_auf\_Zufälligkeit

Die Forderung nach Zufälligkeit von Stichprobendaten kann wie folgt begründet werden. Es sei eine Stichprobe von Schiffswiderständen auf Abhängigkeit von den Hauptabmessungsparametern L/B und B/T zu untersuchen. Die Verteilung der L/B - und B/T-Werte sei grundsätzlich sehr gleichmäßig, d.h. annähernd normal. Aber es mögen auch die Widerstände einer Schiffsserie gemessen worden sein, in der B/T konstant, L/B jedoch variabel war.

Letztere Messungen fallen nicht aus dem Rahmen des Untersuchungszieles. Dennoch stören sie, weil ihr wiederholt auftretendes konstantes B/T eine Sonderstellung in der Stichprobe einnimmt. Es liegen eigentlich zwei verschiedenartige Abhängigkeiten von B/T vor, aber die Analyse kann nur eine einzige gemeinsame ermitteln, deren Maß kleiner ist als bei schwerpunktfreier Zusammensetzung der Stichprobe.

Wenn man annehmen muß, daß die Grundgesamtheit keine Schwerpunkte enthält, sollte auch die Stichprobe frei davon sein. In diesem Sinne kann eine Stichprobe dann als zufällig gelten, wenn in ihren Merkmalen keine schwerpunktartigen Häufungen auftreten.

Daten aus Versuchen, die bereits nach Zufallsgesichtspunkten geplant wurden, sind zufällig, alle anderen sollten auf Zufälligkeit geprüft werden. Ein negatives Ergebnis besagt, daß sich die Stichprobe in gewissem Maße der statistischen Kontrolle entzieht. In der Analyse ermittelte Abhängigkeiten können nicht als Beweis für deren tatsächliche (physikalische) Existanz gewertet werden.

Die Prüfung auf Zufälligkeit ist schwierig, weil eine eindeutige mathematische Definition des Begriffes nicht existiert. Die entwickelten Tests prüfen daher nur gewisse Anzeichen von Nichtzufälligkeiten. Solche Anzeichen sind:

- 1) das Auftreten extremer Variationen,
- 2) das Auftreten von Trends,
- 3) das Auftreten periodischer Variationen,
- 4) das Auftreten von Diskontinuitäten in den gegebenen Daten.

### 3.3.7.1. Zurückweisung von Extremwerten (Ausreißern)

Sind  $x_1$  bzw.  $x_n$  die als Ausreißer verdächtigten Werte, so berechnet man das Verhältnis ihrer Differenz zur Standardabweichung der Stichprobe,

$$z_B = \frac{x_n - x_1}{s} . \quad (19)$$

Aus der Verteilungsfunktion der statistischen Größe  $z_B$  lassen sich in Abhängigkeit von  $n$  Grenzen  $z_0$  und  $z_u$  berechnen, innerhalb derer  $z_B$  mit der Wahrscheinlichkeit  $\alpha$  liegen sollte. Ist  $z_B > z_0$ , so wird für  $(x_n - \bar{x}) > (\bar{x} - x_1) x_n$ , für  $(x_n - \bar{x}) < (\bar{x} - x_1) x_1$  als Ausreißer fortgelassen. Mit dem Rest der Stichprobe wird die Prüfung wiederholt [11].

### 3.3.7.2. Test auf runs

Man zählt aus, wieviele sog. runs in den Daten vorhanden sind. Ein run umfaßt alle unmittelbar aufeinanderfolgenden Daten oberhalb bzw. unterhalb des Medians. Er kann z.B. aus nur einem Wert bestehen, wenn dieser oberhalb, die beiden benachbarten unterhalb des Medians liegen.

Die Theorie der runs liefert eine Methode, die aufgrund der Anzahl der runs in Abhängigkeit von der Anzahl der Daten zu prüfen erlaubt, ob evtl. in den Daten auftretende Trends, periodische Variationen oder Diskontinuitäten auf Nichtzufälligkeit hindeuten. Der Test verwendet die t-Verteilung. Auf eine eingehende Darlegung der Methode wird verzichtet, sie findet sich in [9].

### 3.3.7.3. Methode der Mean Square Successive Differences (MSSD)

Man berechnet

$$\delta^2 = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{n - 1} \quad (20)$$

Wenn die Daten eine zufällige Stichprobe aus einer normalverteilten Grundgesamtheit darstellen, so ist  $\delta^2/2$  eine Schätzung von  $\sigma^2$ . Man vergleicht  $\delta^2/2$  mit  $s^2$ , der anderen Schätzung von  $\sigma^2$ . Das Verhältnis  $\delta^2/2s^2$  wird klein, wenn

Trends in den Daten vorliegen, hingegen groß, wenn kurzweilige Oszillationen auftreten. Der notwendige Test, ob derartige Erscheinungen tatsächlich signifikant sind, wird wie folgt durchgeführt: Man berechnet  $\epsilon = 1 - \delta^2/2s^2$  sowie (für  $n > 25$ ) näherungsweise die Varianz der Verteilung von  $\epsilon$  durch  $\sigma_\epsilon = (n - 2)/(n - 1)(n + 1)$ , und damit  $t = \epsilon/\sigma_\epsilon$ . Dieser t-Wert ist anhand der tabellierten t-Verteilung in Abhängigkeit von  $\nu$  und  $\alpha$  zu überprüfen. Ein größerer berechneter t-Wert als  $t_{\nu, \alpha}$  ist Anzeichen für Nichtzufälligkeit, ein kleinerer weist die Nichtzufälligkeitshypothese zurück [9].

### 3.4. Regressionsanalyse

#### 3.4.1. Beschreibung des Verfahrens

Die Regressionsanalyse hat das Ziel, eine abhängige Variable  $y$  mit Hilfe unabhängiger Variabler  $x_j$ ,  $j = 1$  bis  $m$ , so genau wie möglich zu schätzen. Aus der Schätzformel  $y = f(x_j)$  können unbekannte  $y$ -Werte für beliebige Kombinationen der  $x_j$  interpoliert werden.

Der Ausdruck Variable wird hier anstelle des früher verwendeten Ausdrucks Merkmal gebraucht. Die Begriffe unabhängige und abhängige Variable haben sich eingebürgert, sind aber streng genommen falsch. Wirklich unabhängige (orthogonale) Variable könnten zu der sog. abhängigen Variablen überhaupt nicht in Beziehung stehen. In Wirklichkeit muß man bestrebt sein, Variable  $x_j$  mit möglichst guter Beziehung zu  $y$  zu finden. Die Begriffe abhängig und unabhängig treffen nur insofern zu, als man die unabhängigen Variablen z.B. in Experimenten frei variieren kann, während man die abhängige Variable lediglich mißt. Der übliche Ansatz für die Schätzformel lautet:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + \dots + b_mx_{mi} + \epsilon_i, \quad i = 1 \text{ bis } n. \quad (21)$$

$b_0$  ist eine Konstante,  $b_1$  bis  $b_m$  sind die sog. Regressionskoeffizienten,  $\epsilon_i$  ist das Fehlerglied, bis auf das genau die Schätzformel die gegebenen  $y$ -Werte annähert, und  $n$  ist die Anzahl der Beobachtungen oder Messungen. Die abhängige Variable wird somit als Linearkombination der unabhängigen Variablen aufgefaßt. Werden Zusammenhänge z.B. von der Art  $y = b_0 + b_1x^2 + b_2x \sin x$  vermutet,

so kann dies durch die Substitutionen  $x_1 = x^2$ ,  $x_2 = x \sin x$  auf die Form der Gl. (21) gebracht werden.

Die Aufgabe besteht darin, die Fehlerglieder  $\varepsilon_i$  zu minimieren. Dies geschieht in der Regel nach der Methode der kleinsten Fehlerquadrate.

In verkürzter Schreibweise lautet Gl. (21):

$$y_i = b_0 + \sum_{j=1}^m b_j x_{ji} + \varepsilon_i \quad (22)$$

Die Bedingung zur Minimierung der Fehlerquadratsumme ist

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - b_0 - \sum_{j=1}^m b_j x_{ji})^2 = \text{Min.} \quad (23)$$

Durch Bildung der partiellen Ableitung nach  $b_0$  und Nullsetzen der rechten Seite erhält man:

$$\sum_{i=1}^n y_i - n b_0 - \sum_{j=1}^m (b_j \cdot \sum_{i=1}^n x_{ji}) = 0, \quad (24)$$

woraus folgt:

$$b_0 = \bar{y} - \sum_{j=1}^m b_j \bar{x}_j. \quad (25)$$

Hierin sind  $\bar{x}_j$  bzw.  $\bar{y}$  die Mittelwerte der unabhängigen bzw. der abhängigen Variablen. Setzt man (25) in (23) ein, wird

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [(y_i - \bar{y}) - \sum_{j=1}^m b_j (x_{ji} - \bar{x}_j)]^2 = \text{Min.} \quad (26)$$

Partielle Ableitungen nach den  $b_j$  liefern schließlich:

$$\sum_{j=1}^m \left[ \sum_{i=1}^n (x_{ji} - \bar{x}_j) (x_{ki} - \bar{x}_k) \right] \cdot b_j = \sum_{i=1}^n (x_{ki} - \bar{x}_k) (y_i - \bar{y}). \quad (27)$$

Dies ist ein System simultaner, linearer, algebraischer Gleichungen zur Bestimmung der  $b_j$ . Die Berechnung kann nach einem beliebigen Verfahren ablaufen, z.B. nach der Gaußschen Eliminationsmethode. Zweckmäßig ist die Benutzung der Matrizenrechnung.

Bis hierhin unterscheidet sich das Verfahren nicht von einer gewöhnlichen Ausgleichsrechnung. In [14] jedoch wird ein EFROYMSON angegebener Algorithmus beschrieben, der speziell auf statistische Belange zugeschnitten ist. Dieses Verfahren benutzt die Gaußsche Eliminationsmethode und macht von der Tatsache Gebrauch, daß bei jedem Schritt aus der Hinzufügung von jeweils einer Variablen partielle Regressionsgleichungen in folgender Weise entstehen:

$$\begin{aligned} (1) \quad y &= b_0 + b_1 x_1 \\ (2) \quad y &= b'_0 + b'_1 x_1 + b'_2 x_2 \\ (3) \quad y &= b''_0 + b''_1 x_1 + b''_2 x_2 + b''_3 x_3 \end{aligned} \tag{28}$$

Die Schreibweise deutet an, daß bei jedem Schritt neue Koeffizienten  $b_j$  erzeugt werden. Ihr Zahlenwert hängt von der Summe aller jeweils berücksichtigten unabhängigen Variablen ab.

Die Taktik des Verfahrens von EFROYMSON besteht darin, daß schrittweise jeweils diejenige unabhängige Variable eliminiert wird, die die größtmögliche Verkleinerung des restlichen Fehlergliedes bewirkt. Dies gelingt, indem man Gl. (27) bzw. das aus ihr folgende Normalgleichungssystem in eine dimensionslose Form umschreibt, in der die Koeffizienten vor der ersten Elimination Korrelationskoeffizienten entsprechen. Es sind dies gerade die Korrelationen zwischen den unabhängigen Variablen einerseits und zwischen den unabhängigen und der abhängigen Variablen andererseits. Als erste wird die Variable mit der größten Korrelation zur abhängigen Variablen eliminiert. Anschließend lineare Transformationen der Korrelationskoeffizienten ermöglichen es, vor dem zweiten und jedem weiteren Schritt mit Hilfe der statistischen Größe  $F$  zu prüfen, welche unabhängige Variable als nächste im Sinne bestmöglicher Fehlergliedverminderung in die Regressionsgleichung aufgenommen werden soll.

Die Verwendung von Korrelationskoeffizienten erlaubt Kontrollen, ob starke

Beziehungen zwischen P a a r e n der unabhängigen Variablen bestehen. Ist eine unabhängige Variable angenähert eine Linearkombination einer anderen, so wird nur die bedeutendere in die Regressionsgleichung aufgenommen, die andere ausgeschieden, da sie nur unwesentlich zur Fehlergliedverkleinerung beitragen wird.

Die linearen Transformationen erzeugen multiple Korrelationskoeffizienten, aufgrund derer geprüft werden kann, ob eine unabhängige Variable angenähert eine Linearkombination der S u m m e anderer unabhängiger Variablen ist. Eine solche unabhängige Variable wird nicht in die Regression aufgenommen, oder sie wird, wenn sie bereits aufgenommen war, wieder ausgeschieden.

Kontrolle der Unabhängigkeit der "unabhängigen" Variablen und Auswahl der nächstanzunehmenden Variablen mit Hilfe des F-Tests steuern die Regressionsanalyse in statistischer korrekter Weise. Der zu verwendende  $F_{1, \nu, \alpha}$  - Wert ist variabel, da sich der Freiheitsgrad  $\nu = n - p - 1$ ,  $p =$  Anzahl der bereits aufgenommenen unabhängigen Variablen, bei jedem Aufnahmeschritt um 1 verkleinert. Die Regressionskoeffizienten sind mit der vorgegebenen Wahrscheinlichkeit  $\alpha$  signifikant, sie brauchen nicht auf Signifikanz geprüft zu werden. Berechnet man dennoch den t-Wert jedes Regressionskoeffizienten  $b_j$  unter der Hypothese  $\beta_j = 0$ , so kann aus der nach dem Größenbetrag geordnete Reihenfolge der t-Werte die Bedeutung der zugehörigen unabhängigen Variablen für die Beschreibung der Zielgröße abgelesen werden. Ob diese Aussage quantitativ oder nur qualitativ richtig ist, wird durch Vergleich der Fehlergliedverteilung mit der Normalverteilung überprüft.

Die Auswahl der für ein Problem wichtigen unabhängigen Variablen in der Regressionsanalyse geschieht zwar statistisch korrekt, doch hat diese "Automatik" einen bisweilen unerwünschten Nebeneffekt. Man kann z.B. nicht erzwingen, daß bestimmte unabhängige Variablen unter allen Umständen, andere "nach Belieben" im Regressionsergebnis auftauchen. Dies aber könnte bisweilen erwünscht sein bei Problemen, wo eine strenge Abhängigkeit bestimmter Variablen von der abhängigen Variablen erwartet wird oder aber der tatsächliche Grad der Abhängigkeit untersucht werden soll. Ist die Abhängigkeit dann tatsächlich klein, so taucht die entsprechende unabhängige Variable nicht im Ergebnis auf, sie ist für den Betrachter damit gleichsam unsichtbar und verloren. Um hier dennoch ein Ergebnis zu bekommen, bleibt nur der Weg, die Wahrscheinlichkeit  $\alpha$  zu verringern, bis die Variable erscheint.

Wie die Wahrscheinlichkeit spielt auch der Freiheitsgrad  $\nu = n - p - 1$  eine bedeutende Rolle. Wächst die Zahl  $p$  der aufgenommenen unabhängigen Variablen, so nimmt er ständig ab und geht gegen Null für  $p \rightarrow n$ . Eine vernünftige statistische Aussage ist dann nicht mehr möglich. Anschaulich läßt sich das für den Fall demonstrieren, daß die Zahl der Beobachtungen und der unabhängigen Variablen übereinstimmt. Die Interpolationsformel geht dann exakt durch alle beobachteten Punkte, unabhängig davon, ob die unabhängigen Variablen vernünftig gewählt werden oder überhaupt nichts mit dem Problem zu tun haben.

### 3.4.2. Annahmen, Forderungen und Anwendung

Von den Werten  $y_i$  der abhängigen Variablen wird angenommen, daß sie einer normalverteilten Grundgesamtheit angehören und daraus eine zufällige Stichprobe darstellen. Prüfverfahren für letztere Annahme wurden in 3.3.7. beschrieben. Die Verteilung der gegebenen  $y_i$  ist zwar nicht entscheidend, doch sollte man sich stets ein Bild davon machen (s. 3.3.6.) und Ausreißer eliminieren.

Die Werte  $x_{ji}$  der unabhängigen Variablen unterliegen keinerlei statistischen Beschränkungen. Sie können von beliebiger Form und voneinander abhängig \* sein. Aus praktischen Gesichtspunkten sollten jedoch auch sie einige Forderungen erfüllen:

- 1) Ihre Werte sollten aus numerischen Gründen etwa von gleicher Größenordnung sein. Es ist nicht nötig, daß die unabhängigen Variablen dimensionslos gemacht oder auf den Wertevorrat  $-1$  bis  $+1$  normiert werden. Lediglich ihre Minimal- und Maximalwerte sollten angegeben werden.
- 2) Sie sollten in der Stichprobe ungefähr so verteilt sein, wie sie es in der großen Gesamtheit auch sind. Sie dürfen daher ruhig Schwerpunkte aufweisen, wenn dies als normal angesehen werden kann. Doch sollte man sich hierüber in jedem Fall Klarheit verschaffen und zu beurteilen versuchen, wieweit die vorgefundene Verteilung der erwarteten entspricht.

---

\* Vgl. hierzu jedoch den vorangehenden Abschnitt.

Die Regressionsgleichung darf zur Vorhersage des Verhaltens eines neuen Entwurfs verwendet werden, wenn die Werte seiner unabhängigen Variablen aus dem Wertevorrat der Regressionsstichprobe stammen und sein Charakter mit dem Charakter der Stichprobenobjekte übereinstimmt. Extrapolation ist nur erlaubt, wenn die gefundenen Zusammenhänge zwischen unabhängigen Variablen und abhängiger Variabler Trends aufweisen, die theoretisch oder durch Erfahrung zu bestätigen sind.

Die Optimierung eines neuen Entwurfs entspricht nicht den Fähigkeiten der Regressionsanalyse. Sie liefert lediglich ein Ausgleichspolynom für gegebene Punkteverteilungen. Natürlich kann jedoch die wiederholte Berechnung der Regressionsgleichung mit geänderten "Parametern" des neuen Entwurfs die Suche nach dem Optimum fördern.

Schwerpunktbildungen in den Verteilungen der unabhängigen Variablen müssen beachtet werden. Das Verhalten eines neuen Entwurfs kann falsch vorhergesagt werden, wenn die Werte seiner unabhängigen Variablen extrem weit von solchen Schwerpunkten entfernt liegen.

### 3.4.3. Einige ergänzende Formeln

Die Vertrauensgrenzen der Koeffizienten in der multiplen linearen Regressionsgleichung berechnen sich aus

$$b_k \pm t_{v,\alpha} s_{y;x_1 \dots x_p} \sqrt{c_{kk}} \quad (29)$$

Hierin sind  $s_{y;x_1 \dots x_p}$  die Standardabweichung des Fehlergliedes mit dem Freiheitsgrad  $v = n-p-1$  und  $c_{kk}$  ein modifiziertes Diagonalelement der multiplen, mehrfach linear transformierten Korrelationsmatrix, die bei Beendigung der Regressionsgleichung den Platz der ursprünglichen Matrix der Korrelationskoeffizienten einnimmt;  $c_{kk}$  ist also bekannt.

Das eigentliche Ergebnis einer Regressionsanalyse ist die Regressionsgleichung. Wenn man zu ihrer Bestimmung, was im allgemeinen zutrifft, nur eine einzige Stichprobe benutzen kann, so stellt die Regressionsgleichung allenfalls

die bestmögliche Schätzung der für die Grundgesamtheit gültigen Interpolationsgleichung dar. Die Lage ihrer graphischen Darstellung, der Regressions"fläche", im vieldimensionalen Variablenraum weicht im allgemeinen von der "richtigen" Lage ab. Um dem Rechnung zu tragen, gibt man Vertrauensgrenzen für die Regressionsfläche  $y$  an, die sich berechnen aus

$$y \pm t_{v,\alpha} s_{y;x_1 \dots x_p} \sqrt{\frac{1}{n} + \sum_{k=1}^p (x_k - \bar{x}_k)^2 c_{kk} + \sum_{k=1}^p \sum_{l=1, l \neq k}^p (x_k - \bar{x}_k)(x_l - \bar{x}_l) c_{kl}} \quad (30)$$

Mit der Wahrscheinlichkeit  $\alpha$  liegt die "richtige" Regressionsgerade innerhalb des hierdurch gekennzeichneten Vertrauensbereiches. Die  $c_{kl}$  sind wie  $c_{kk}$  modifizierte Elemente der Korrelationsmatrix, die am Ende der Regressionsrechnung erhalten wird. Aus der Symmetrie der Korrelationsmatrix folgt  $c_{kl} = c_{lk}$ , womit sich Gl. (30) vereinfacht berechnen läßt aus

$$y \pm t_{v,\alpha} s_{y;x_1 \dots x_p} \sqrt{\frac{1}{n} + \sum_{k=1}^p (x_k - \bar{x}_k)^2 c_{kk} + 2 \sum_{k=1}^p \sum_{l=k+1}^p (x_k - \bar{x}_k)(x_l - \bar{x}_l) c_{kl}} \quad (30a)$$

Nimmt man einmal an, die Interpolationsgerade für die Grundgesamtheit wäre bekannt, so würden die Einzelwerte  $y_i$  dennoch um diese Gerade streuen, es sei denn, die Interpolationsgleichung enthält soviele Koeffizienten wie die Grundgesamtheit Einzelwerte, so daß diese exakt miteinander verbunden werden. (Hundertprozentige Prognosegenauigkeit ist also selbst bei Verwendung einer Stichprobe von der Größe einer Grundgesamtheit kaum sinnvoll zu erzielen).

Diese Streuung tritt auch bei einer normalen Stichprobe auf, und es müssen Vorhersagebereiche für jedes aus der Regressionsgleichung berechnete  $\hat{y}_i$  angegeben werden, in die mit der Wahrscheinlichkeit  $\alpha$  alle Werte fallen würden, die der Grundgesamtheit angehören. Die Grenzen dieser Vertrauensbereiche sind

$$y \pm t_{v,\alpha} s_{y;x_1 \dots x_p} \sqrt{1 + \frac{1}{n} + \sum_{k=1}^p (x_k - \bar{x}_k)^2 c_{kk} + 2 \sum_{k=1}^p \sum_{l=1, l \neq k}^p (x_k - \bar{x}_k)(x_l - \bar{x}_l) c_{kl}}. \quad (30b)$$

wobei wiederum von der Identität  $c_{kl} = c_{lk}$  Gebrauch gemacht wurde.

### 3.5. Faktorenanalyse

#### 3.5.1. Definition und Anwendungszweck

Es sei angenommen, mehrere unabhängige Variable korrelierten stark miteinander. Sie besagen dann weitgehend dasselbe, oder anders formuliert, in ihnen prägt sich eine gemeinsame Größe aus, die aber nicht direkt meßbar ist. Die Frage stellt sich, ob diese gemeinsame Größe nicht herausgefiltert werden kann. Mit ihr müßte es möglich sein, die Grundinformation einfacher zu beschreiben als durch die Vielzahl der Variablen.

Die Faktorenanalyse ist eine statistische Methode, die dieses Problem löst. Sie extrahiert aus den Variablen sog. Faktoren, die die Struktur der gegebenen Daten durch eine einfachere Struktur ersetzen. Das Wort Faktor ist dabei nicht im üblichen Sinne zu verstehen, sondern als Bezeichnung für die im Hintergrund stehende gemeinsame Größe.

Im Prinzip können aus gegebenen Daten so viele Faktoren extrahiert werden, wie Variable vorhanden sind, doch wird eine kleinere Zahl von Faktoren angestrebt. Von der Gesamtstreuung der Daten erklärt jeder Faktor einen Teil, und seine Bedeutung für die Reproduktion der Daten kann durch die Höhe seines Anteils an erklärter Varianz gemessen werden.

Die Faktorenanalyse geht von der Korrelationsmatrix aus, in der sämtliche Informationen über die Beziehung der unabhängigen Variablen zueinander enthalten sind. Die Aufgabe lautet, die  $p \cdot p$ -Korrelationsmatrix durch eine kleinere  $p \cdot r$ -Matrix, das sog. Faktorenmuster, zu ersetzen, wo  $r$  die Zahl der Faktoren und  $p$  die Zahl der Variablen ist. Jeder Faktor bezieht demnach einen Beitrag, die sog. Ladung, von jeder Variablen, und je mehr eine Variable mit einem Faktor zu tun hat, desto höher lädt sie ihn. Zahlenwerte der Ladungen liegen zwischen  $-1$  und  $+1$ . Je mehr Variable einen Faktor laden, desto größer ist sein Anteil an der Varianzerklärung.

Die klassische Anwendung der Faktorenanalyse hat das Ziel, wenige Faktoren mit möglichst hohen Anteilen an erklärter Varianz zu extrahieren. In dieser Arbeit jedoch wird die Faktorenanalyse zu einem anderen Zweck eingesetzt. Mit ihrer Hilfe sollen die zur Beschreibung des Widerstandes verwendeten unabhängigen Variablen auf Gemeinsamkeit, speziell aber auf differierende Informationsinhalte untersucht werden, um zu verhindern, daß der Regressionsanalyse zu wenig unterschiedliche Grundinformationen angeboten werden. Es kommt darauf an, Variablenkonstellationen zu finden, die möglichst viele differenzierte, voneinander unabhängige Grundinformationen enthalten, aus denen sich also viele Faktoren mit kleineren Anteilen an erklärter Varianz extrahieren lassen.

Um Mißverständnissen vorzubeugen, sei ausdrücklich hervorgehoben, daß die Faktorenanalyse nur zur Klärung der Korrelation der unabhängigen Variablen untereinander eingesetzt wird, nicht aber zur Auffindung guter Korrelationen zwischen der abhängigen Variablen und den unabhängigen Variablen. Diese Aufgabe löst die Regressionsanalyse. Die abhängige Variable spielt in der Faktorenanalyse überhaupt keine Rolle.

Untersuchungen auf wirkliche Unabhängigkeit der "unabhängigen" Variablen im Zusammenhang mit Widerstandsregressionen sind nicht bekannt. Sie sollen hier dazu dienen, frühzeitig die Notwendigkeit zur Bereitstellung weiterer unabhängiger Variablen erkennen zu können, wenn die bis dahin vorgesehenen Variablen einen gravierenden Mangel an differenzierten Grundinformationen befürchten lassen. In einem weiteren Sinne kann man daher in der Faktorenanalyse der unabhängigen Variablen einen rationalen Ansatz zur Systematisierung der Variablenauswahl erblicken.

Es erscheint möglich, den Rechenaufwand in der Regressionsanalyse erheblich zu senken, wenn anstelle der unabhängigen Variablen direkt mit den Faktoren gearbeitet wird. Dagegen spricht jedoch, daß nur in wenigen Fällen eine anschauliche (physikalische) Deutung der extrahierten Faktoren gelingen dürfte. Die unabhängigen Variablen wird man anschaulich wählen, doch fehlen im allgemeinen die Begriffe für die ihnen gemeinsamen Größen, sonst würde man diese ja von Anfang an verwenden.

Eine ausführliche Beschreibung der Faktorenanalyse findet sich in dem Buch von ÜBERLA [13]. An dieses angelehnt, wird im folgenden kurz die zugrundeliegende Mathematik behandelt.

### 3.5.2. Mathematische Grundlagen

Die Matrix  $X = (x_{ji})$  bezeichne die Ausgangsdaten, wo  $j = 1$  bis  $p$  die Variablen und  $i = 1$  bis  $n$  die Beobachtungen sind. Der Korrelationskoeffizient zwischen zwei Variablen  $j$  und  $k$  berechnet sich aus

$$r_{jk} = \frac{\sum_{i=1}^n (x_{ji} - \bar{x}_j) (x_{ki} - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 \sum_{i=1}^n (x_{ki} - \bar{x}_k)^2}} = \frac{s_{jk}}{s_j s_k} \quad (31b)$$

Hierin sind  $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ji}$  der Mittelwert,

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2} \text{ die Standardabweichung,}$$

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ji} - \bar{x}_j) (x_{ki} - \bar{x}_k) \text{ die Kovarianz.}$$

Durch die Transformation

$$z_{ji} = \frac{x_{ji} - \bar{x}_j}{s_j} \quad (32)$$

wird die Matrix X in die Matrix  $Z = (z_{ji})$  überführt, deren Variable auf den Mittelwert Null und die Varianz 1 standardisiert sind. Die Korrelationskoeffizienten der standardisierten Variablen vereinfachen sich zu

$$r_{jk} = \frac{\sum_{i=1}^n (z_{ji} z_{ki})}{\sqrt{\sum_{i=1}^n z_{ji}^2 \sum_{i=1}^n z_{ki}^2}} = s_{jk} = \frac{1}{n-1} \sum_{i=1}^n z_{ji} z_{ki} \quad (33)$$

d.h. Korrelationskoeffizient und Kovarianz sind nach derselben Formel zu berechnen.

Die Korrelationsmatrix bzw. die Kovarianzmatrix schreibt sich somit

$$R = S = \frac{1}{n-1} ZZ', \quad (34)$$

wobei  $Z'$  die Transformierte von  $Z$  ist.

Es ist Ziel der Faktorenanalyse, die Elemente  $z_{ji}$  durch eine lineare Kombination mehrerer hypothetischer Faktoren zu beschreiben. Wenn  $r$  die Zahl der Faktoren ist, lautet der Ansatz:

$$z_{ji} = a_{j1}p_{1i} + a_{j2}p_{2i} + \dots + a_{jr}p_{ri} \quad (35)$$

Die  $a_{j1}$  sind feste Koeffizienten, die zu ermitteln sind,  $p_i$  ist eine bestimmte Beobachtung mit den Werten  $p_{1i}, p_{2i}, \dots, p_{ri}$ .

In Matrixform läßt sich Gl. (35) schreiben:

$$Z = AP \quad (36)$$

$A$  und  $P$  sind unbekannt,  $Z$  ist die Matrix der standardisierten Ausgangsdaten. Ohne Einführung von Restriktionen ist Gl. (36) nicht zu lösen. Setzt man Gl. (36) in Gl. (34) ein, ergibt sich

$$\begin{aligned} R = \frac{1}{n-1} ZZ' &= \frac{1}{n-1} AP(AP)' = \frac{1}{n-1} APP'A' \\ &= A \frac{1}{n-1} PP'A', \end{aligned} \quad (37)$$

worin  $\frac{1}{n-1} PP'$  analog zu Gl. (34) eine Korrelationsmatrix ist, und zwar

die zwischen den Faktoren:

$$R = ACA' \quad (38)$$

Fordert man, daß die Faktoren unkorreliert, d.h. orthogonal seien, ist  $C = I$ ,  $I =$  Einheitsmatrix, woraus folgt

$$R = AA' \quad (39)$$

Man gewinnt die Faktoren, d.h. die Matrix  $A$ , aus Gl. (38) oder (39).  $A$  ist eine Matrix, die, mit ihrer Transponierten multipliziert, die Korrelationsmatrix  $R$  reproduziert, wobei erwünscht ist, daß ihre Struktur wesentlich einfacher als die von  $R$  ist.

Eine geometrische Deutung des Faktorenproblems geht von einer Punktwolke in einem mehrdimensionalen Raum aus, den die zu untersuchenden Variablen aufspannen. Durch ihren Schwerpunkt sollen Achsen so gelegt werden, daß sich in möglichst dichtem Abstand von ihnen möglichst viele Punkte konzentrieren. Die erste Achse soll ein Höchstmaß von Punkten auf sich vereinigen, die zweite, senkrecht zur ersten, ein Höchstmaß der Restpunkte auf sich, usw.

Diese geometrische Vorstellung (Bild 5 aus [5]) liegt der sog. Hauptkomponenten- oder Hauptachsenmethode zugrunde. Sie ist eine von mehreren Methoden zur Auflösung von Gl. (39). Die Faktoren entsprechen dabei den Achsen durch die Punktwolke.

Gl. (39) wird dann eindeutig lösbar, wenn man verlangt, daß die Summe der Quadrate der Faktorenladungen des ersten Faktors ein Maximum der Gesamtvarianz erreichen soll, die des zweiten Faktors ein Maximum der Restvarianz usw., für den ersten Faktor also

$$S_1 = \sum_{j=1}^p a_{j1}^2 = \text{Max} \quad (40)$$

unter den  $p(p-1)/2$  voneinander unabhängigen Bedingungen

$$r_{jk} = a_{j1} a_{k1} \quad \text{mit } j, k = 1 \text{ bis } p, j < k .$$

Die hierdurch gestellte Aufgabe besteht in der Maximierung einer Funktion unter Nebenbedingungen. Ihre Lösung führt auf ein System von  $p$  homogenen Gleichungen für die  $p$  Unbekannten  $a_{j1}$ , dessen Koeffizientendeterminante für eine nichttriviale Lösung gleich Null werden muß:

$$\begin{vmatrix} (1-\lambda) & r_{12} & \dots & r_{1p} \\ r_{21} & (1-\lambda) & \dots & r_{2p} \\ \cdot & & & \\ \cdot & & & \\ r_{p1} & r_{p2} & \dots & (1-\lambda) \end{vmatrix} = 0 \quad (41)$$

Hierin sind  $r_{jk}$  die Elemente der Korrelationsmatrix,  $\lambda$  wird als Eigenwert der Determinante bezeichnet. Die Determinante hat genau  $p$  reelle Lösungen für  $\lambda$ , wobei  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ .

Die Hauptachsenmethode führt die Lösung von Gl. (39) somit auf ein Eigenwertproblem zurück.

Mit Hilfe von  $\lambda_1$  können die  $a_{j1}$  bestimmt werden, mit  $\lambda_2$  die  $a_{j2}$  usw. Die Vektoren  $A_1$  ( $A_1 = a_{j1}, a_{j2} \dots a_{jp}$ ) erfüllen die geforderten Bedingungen größtmöglicher Varianzausschöpfung.

Wegen der Standardisierung der Variablen auf den Mittelwert Null und die Streuung 1 ist die Summe der Eigenwerte gleich der Anzahl  $p$  der Variablen. Mittels Division der Eigenwerte durch  $p$  erhält man daher ein direktes Maß für die Varianzerklärung durch die einzelnen Faktoren. Eine übliche Grenzsetzung für Beendigung der Faktorenextraktion ist die Angabe eines kleinsten Wertes für  $\lambda$ , etwa  $\lambda \geq 0,5$ .

### 3.5.3. Rotation

Durch die Hauptachsenmethode werden die Richtungen der Hauptachsen (Faktoren) in Bezug auf das Koordinatensystem der Variablen festgelegt. Die Spalten der Matrix A sind wegen  $C = I$  in Gl. (39) orthogonal, sämtliche Hauptachsen stehen senkrecht aufeinander.

Ein anderes Extraktionsverfahren würde, abhängig von seinem Extraktionskriterium, eine andere Matrix A erzeugen. Es sind unendlich viele Matrizen A möglich, die alle die Matrix R gleich gut reproduzieren. In jeder Matrix A bedinden sich die Faktoren in einer durch das Verfahren bestimmten willkürlichen Position in Bezug auf die Variablen. Dies macht die anschauliche Deutung der Faktoren zuweilen unmöglich. Aus den Matrizen A soll eine ausgewählt werden, die die Bedingung erfüllt, daß die Lage der Faktoren zu den Variablen möglichst eindeutig ist. Man erreicht dies durch Rotation der Faktorenachsen.

Die Rotation versucht, die Faktorenachsen möglichst in Richtung von Variablenachsen zu drehen. Die Korrelationen einzelner Variabler miteinander werden dabei beachtet. Man bezeichnet den Vorgang mit Überführung in die Einfachstruktur und beabsichtigt damit, die Interpretierbarkeit der Faktoren zu verbessern.

Es gibt mehrere Rotationsmethoden. Schiefwinkelige Rotation ist notwendig, wenn Matrix C in Gl. (38) ungleich I ist, d.h. wenn die Faktoren untereinander korrelieren. Hier wird ein Verfahren rechtwinkliger Rotation beschrieben, das für den vorgesehenen Verwendungszweck ausreicht (Varimaxmethode).

Die Varimaxmethode definiert die Einfachheit eines einzelnen Faktors durch die Summe der Streuungen der quadrierten Faktorenladungen

$$s_1^2 = \frac{1}{p} \sum_{j=1}^p (a_{j1})^2 - \frac{1}{p^2} \left[ \sum_{j=1}^p a_{j1} \right]^2 \quad (42)$$

(dies ist eine vereinfachte, numerisch vorteilhafte Schreibweise für die Streuung, vgl. Gl. (2)). Wenn diese Varianz ein Maximum ist, dann liegen die einzelnen Ladungen nahe an 1 oder nahe an Null, so daß der Faktor möglichst einfach ist. Die Summe über die Streuungen aller Faktoren muß ebenfalls ein

Maximum annehmen, wenn für alle Faktoren die Streuung der Quadrate der Faktorenladungen möglichst groß sein soll. Dies sind Bedingungen, die iterativ die Rotation der Faktorenachsen zur Einfachstruktur erlauben.

Die rotierten Faktoren lassen sich mit Hilfe der sog. Transformationsmatrix in die ursprünglichen Faktoren zurückrechnen. Die Eigenwerte der rotierten Faktoren und damit die von jedem rotierten Faktor erklärte Varianz der Korrelationsmatrix erhält man aus

$$\tilde{\lambda}_j = \sum_{l=1}^r \lambda_l m_{lj}^2, \quad j = 1 \text{ bis } p, \quad (43)$$

wo  $r$  die Anzahl der extrahierten Faktoren,  $\lambda_l$  die Eigenwerte der unrotierten Faktoren und  $m_{lj}$  die Elemente der Transformationsmatrix sind.

Zur Durchführung von Faktorenanalysen wurde das am Rechenzentrum der Universität Hamburg verfügbare Statistikprogramm SPSS (Statistical Package of the Social Sciences) verwendet.

#### 4. Verfahren der Widerstandsanalyse

##### 4.1. Graphische Verfahren. Eine Anwendung auf moderne Schiffsformen

Die Verfahren zur Nutzung von Modellmessungen für die Prognose des Widerstandes eines neuen Entwurfs können in zwei Gruppen eingeteilt werden: Graphische Verfahren und mathematische Verfahren auf der Grundlage statistischer Methoden. Zu der ersten Gruppe zählen bekannte Verfahren wie die von TAYLOR/GERTLER (1943/1954), GULDHAMMER/HARVALD (1965) oder LAPP/AUF'M KELLER (1954/1973). Die gegenwärtige Schiffsentwurfspraxis bedient sich überwiegend der Verfahren aus dieser Gruppe. Ihres großen Bekanntheitsgrades wegen braucht nicht näher auf sie eingegangen zu werden. Es soll jedoch die Eignung eines graphischen Verfahrens zur Prognose des Widerstandes moderner Schiffsformen an einer Stichprobe überprüft werden.

Das Verfahren von GULDHAMMER/HARVALD [15] beruht auf Modellwiderständen zahlreicher Serienversuche und einiger unsystematischer Versuche. Interpolationsparameter für Schiffe mit sog. Standardform sind  $L/\nabla^{1/3}$ ,  $C_p$  und  $F_n$ . Für von der Standardform abweichendes  $B/T$  und  $x_B$  sowie andere Spant- und Vorstevenkonturen, Anhänge und Bugwulst sind Korrekturen vorzunehmen. Ermittelt wird der Beiwert des Restwiderstandes  $C_R$ .

Für 34 zufällig gezogene Schiffe wurden die in 40 Modellversuchen auf verschiedenen Tiefgängen gemessenen Widerstände den Ergebnissen nach diesem Verfahren gegenübergestellt. Die Stichprobe bestand aus 30 Fracht- und Tank Schiffen mit 34 Messungen, davon 7 auf vertrimmten Ballasttiefgängen, sowie 4 schnellen Schiffen (1 Container-, 3 Passagierschiffe) mit 6 Tiefgängen, davon 2 Trimmtiefgänge. 13 Schiffe hatten einen Wulst, 7 Schiffe waren Zweischrauber, 1 Schiff hatte drei Schrauben.

In nur 50 % aller Fälle reichte der bei GULDHAMMER/HARVALD vorhandene Wertebereich der Interpolationsparameter für die Stichprobenschiffe aus, in den übrigen Fällen mußte für ein, zwei oder alle drei Parameter extrapoliert werden.

Um einen Gesamtüberblick über die Abweichungen zwischen gemessenen und berechneten Widerständen zu erhalten, wurde für jedes Schiff eine mittlere Abweichung in Prozent des gemessenen Widerstandes berechnet, genommen über alle

Meßpunkte im jeweiligen Geschwindigkeitsintervall. Das Histogramm dieser mittleren Abweichungen für alle Schiffe zeigt Bild 6.

Abweichungen von 10 % und mehr traten insbesondere bei schnellen Schiffen mit Wulst auf, deren Widerstand überschätzt wurde. Der Mittelwert aller Abweichungen liegt bei 9 % des Gesamtwiderstandes.

Dieses Stichprobenergebnis bestätigt die bekannte Erfahrung, daß die Vorhersage des Widerstandes für einen bedeutenden Teil gegenwärtig gebauter Schiffe mit gebräuchlichen Widerstandsverfahren Schwierigkeiten bereitet. Dies ist vornehmlich darauf zurückzuführen, daß graphische Verfahren gewöhnlich auf ältere Modellserienuntersuchungen mit teilweise überholten Schiffsformen zurückgreifen.

#### 4.2. Mathematisch-statistische Verfahren

Einer statistischen Analyse des Modellwiderstandes stehen zwei grundsätzlich verschiedene Möglichkeiten offen:

- 1 ) der geschwindigkeitsabhängige Ansatz,
- 2 ) der Ansatz bei konstanter Froudezahl.

Der erste Ansatz liefert eine einzige Widerstandsgleichung mit strakendem Kurvenverlauf für das gesamte untersuchte Geschwindigkeitsintervall. Dies ist ein großer Vorzug dieser Methode, doch hat sie auch einige gewichtige, weiter unten diskutierte Nachteile. Aus dem zweiten Ansatz erhält man pro Froudezahl eine Gleichung. Um ein gewünschtes Geschwindigkeitsintervall zu überstreichen, müssen mehrere Gleichungen ausgewertet werden. Wegen der von Froudezahl zu Froudezahl aller Wahrscheinlichkeit nach unterschiedlichen Zusammensetzung der zu analysierenden Kollektive und der sich daraus ergebenden wechselnden Variablenkombinationen in der Regressionsgleichung wird die berechnete Widerstandskurve eine Streuung aufweisen, und es wird erst nach einiger Übung gelingen, einen plausiblen, glatten Kurvenzug durch die springenden Punkte hindurchzulegen. Wertvolle Hinweise für die Größe der dabei erlaubten Abweichungen wird man aus den Vertrauensgrenzen zu erhalten versuchen.

Die Entscheidung für den einen oder den anderen Ansatz ist nicht allein von diesen Überlegungen abhängig, sondern auch von der Beachtung der Gesetze zur Übertragung des Modellwiderstandes auf die Großausführung. Nach der Froudeschen Hypothese wird der Modellwiderstand in Reibungswiderstand der ebenen Platte und Restwiderstand aufgespalten, und beide Komponenten werden nach verschiedenen Gesetzen übertragen (vgl. 7.1.). Der Plattenreibungswiderstand wird über die für Modell und Großausführung geltenden Reynoldszahlen nach einer der üblichen Reibungslinien umgerechnet. Der geeignet dimensionslos gemachte Restwiderstand ist maßstabsunabhängig, das heißt sein Beiwert hat gleiche Größe für Modell und Großausführung. Es existiert jedoch keine analytische, von der Geschwindigkeit abhängige Formel für den Restwiderstand, so daß er lediglich an festen Froudezahlen analysiert werden kann (Ansatz 2).

Modernen Forschungsergebnissen zufolge sollte der Modellwiderstand zur Übertragung in viskosen und Wellenwiderstand aufgespalten werden. Der viskose Widerstand  $R_V$  enthält außer dem Plattenreibungswiderstand  $R_{F0}$  einen weiteren, von der Schiffsform abhängigen Anteil, der  $R_{F0}$  um einen Faktor  $k$  vergrößert, so daß  $R_V = (1+k)R_{F0}$ .  $k$  kann im Modellversuch ermittelt werden.  $R_{F0}$  wird unter Berücksichtigung der unterschiedlichen Reynoldszahlen auf die Großausführung übertragen, während  $k$  unverändert gilt. Der dimensionslose Beiwert des Wellenwiderstandes ist ebenfalls maßstabsunabhängig. Für ihn lassen sich theoretische, geschwindigkeitsabhängige Formeln aufstellen, so daß er sich recht gut für eine Analyse nach Ansatz 1 eignet.

Eine unangenehme Schwierigkeit hierbei liegt jedoch in der Voraussetzung, daß  $k$  bekannt sein muß. Diese Voraussetzung ist bislang nur in Ausnahmefällen erfüllt. Von ganz wenigen Schiffen kennt man  $k$ , und die Bestimmung des  $k$ -Faktors im Experiment ist schwierig, so daß die wenigen vorliegenden Werte als unsicher anzusehen sind. Ihr Vorrat beschränkt sich außerdem überwiegend auf völlige Schiffstypen, und es besteht keine Möglichkeit, mit ihnen die Werte von schlanken Schiffen verlässlich abzuschätzen.

Angesichts dieser Gegebenheiten kam für die vorliegende Untersuchung lediglich die statistische Analyse des Restwiderstandes an festen Froudezahlen in Betracht. Dennoch erschien im Hinblick auf zukünftige Widerstandsanalysen der geschwindigkeitsabhängige Ansatz für den Wellenwiderstand verführerisch genug, um mit der anschließenden Diskussion einiger Formeln das Interesse auf ihn zu lenken.

#### 4.2.1. Der geschwindigkeitsabhängige Ansatz

Der Gesamtwiderstand wird in zähigkeitsbedingten Widerstand und Wellenwiderstand aufgespalten:

$$R_T = R_V + R_W , \quad (44)$$

oder in dimensionsloser Form nach Division durch  $0,5 \rho V^2 S$

$$C_T = C_V + C_W . \quad (45)$$

Nach Ermittlung des Zähigkeitswiderstandsbeiwertes aus

$$C_V = (1+k) C_{F0} , \quad (46)$$

wo  $C_{F0}$  der Plattenreibungswiderstandsbeiwert und  $k$  der Faktor ist, der den Formeinfluß auf den Zähigkeitswiderstand berücksichtigt (Form- oder  $k$ -Faktor), erhält man

$$C_W = C_T - C_V \text{ oder } R_W = R_T - R_V . \quad (47)$$

Es existieren mehrere geschwindigkeitsabhängige Formeln für den Wellenwiderstand. Die Ansätze von OORTMERSSEN und INUI werden näher betrachtet.

##### 4.2.1.1. Ansatz von OORTMERSSEN

Unter Beschränkung auf kleine Schiffe ohne paralleles Mittelschiff wie Schlepper und Trawler leitet OORTMERSSEN [6] eine zweidimensionale Druckverteilung für schiffsähnliche Körper her. Aufgrund statistischer Analysen der verfügbaren Daten werden einige Teile der ursprünglichen Druckformeln entfernt, und es ergibt sich als wirksame Gleichung für den Wellenwiderstand:

$$\begin{aligned} \frac{R_W}{\Delta} = & C_1 \exp\left(-\frac{m}{9} F_n^{-2}\right) + C_2 \exp(-m F_n^{-2}) + \\ & C_3 \exp(-m F_n^{-2}) \cdot \sin(F_n^{-2}) + C_4 \exp(-m F_n^{-2}) \cdot \\ & \cdot \cos(F_n^{-2}). \end{aligned} \quad (48)$$

Hierin bedeuten:

$m = b_1 C_p^{-b_2}$ ,  $b_1$  und  $b_2$  Koeffizienten, die aus den Daten zu ermitteln sind,

$$C_i = f(x_B, C_p, L/B, i_E \cdot \left[ \frac{1}{2B} (L_{pp} + L_{WL}) \right], B/T, C_M)$$

Die Koeffizienten  $C_i$  stellen den Zusammenhang zwischen Wellenwiderstand und individueller Schiffsform her.

#### 4.2.1.2. Ansatz nach INUI

Für Froudezahlen kleiner als 0,35 gibt INUI [16] Reihenentwicklungen für den aus einer Quellensenkenverteilung abgeleiteten Wellenwiderstand des dünnen Schiffes an. Der erste Anteil (Glied mit  $C_1^*$  in Gl. (49)) wächst monoton mit der Froudezahl, der zweite wechselt zwischen plus und minus. Unter Beschränkung auf die ersten Glieder der Reihenentwicklung und nach einigen Vereinfachungen erhält man nach einem Vorschlag von SHARMA:

$$C_W = C_1^* F_n^4 + C_2^* F_n^5 \cos(k_0 L + \frac{\pi}{4}) + C_3^* F_n^5 \sin(k_0 L + \frac{\pi}{4}) \quad (49)$$

Hierin ist  $k_0 = g/V^2$  die Wellenzahl und  $k_0 L = Lg/V^2 = 1/F_n^2$ .

Die Koeffizienten  $C_i^*$  haben dieselbe Aufgabe wie  $C_i^*$  bei OORTMERSSEN.

#### 4.2.1.3. Auswertung der Ansätze

Die Schwierigkeit bei der Auswertung der Formeln (48) oder (49) liegt darin, daß  $C_W$  aus dem gemessenen Gesamtwiderstandsbeiwert  $C_T$  nur dann isoliert werden kann, wenn der Formfaktor  $k$  bekannt ist. OORTMERSSEN schätzt den Formfaktor für seine Schiffstypen nach der Formel

$$k = 139 C_B^2 / \{(L/B)^2 \cdot [(B/T)^2 + 1]\}$$

ab und setzt für  $C_{FO}$  die ITTC-Linie von 1957 ein. Es existiert noch eine Reihe weiterer Schätzformeln für  $k$  (s. Abschnitt 7.2.), doch keine ist anerkannt zuverlässig. Die Anwendung der Formel INUIS für statistische Zwecke wird hier erstmals vorgeschlagen, die Wahl eines Abschätzverfahrens für  $k$  ist also noch völlig offen.

Wenn die Formfaktoren für das zu analysierende Modellkollektiv durch Messung bekannt sind, oder wenn man sich wie OORTMERSSEN für eine Schätzformel entschieden hat, steht der Analyse des Wellenwiderstandes nichts mehr im Wege. Das Verfahren erfordert zwei Schritte. Im ersten werden Zahlenwerte für  $C_i$  für jede gegebene Widerstandskurve mit Hilfe einer Ausgleichsrechnung über die Meßpunkte ermittelt. Den zweiten Schritt bilden Regressionsanalysen über die Stichprobengesamtheit eines jeden  $C_i$  in Abhängigkeit von Formvariablen.

Der Charakter der hier nur nebenbei geführten Studie erlaubte nicht, eine komplette Analyse von Modellen mit bekanntem  $k$ -Faktor durchzuführen. So wurde lediglich untersucht, in wieweit sich die durch Gln. (48) und (49) gegebenen Widerstandsformeln zur Darstellung gemessener Widerstandskurven beliebiger Modelle eignen. Zu diesem Zweck wurden aus einer größeren Datenmenge eine Stichprobe von 6 Messungen zufällig gezogen.

Dieses kleine Kollektiv umfaßte drei Tankschiffe, alle mit Bugwulst, davon eines im Ballastzustand, ein Frachtschiff, ein Zweischrauben-Passagierschiff mit Bugwulst und einen Bagger. Von keinem Schiff war der  $k$ -Faktor bekannt. Deshalb wurden die Gln. (48) und (49) modifiziert in einer Weise, die für den vorgesehenen Zweck zulässig ist, sich jedoch nicht für eine statistische Analyse eignet. Unter Verwednung des Gesamtwiderstandes läßt sich Gl. (48) schreiben:

$$\frac{R_T}{\Delta} = C_o^* \frac{R_{FO}}{\Delta} + (\text{rechte Seite von Gl. (48)}),$$

oder nach Multiplikation mit  $\Delta/R_{FO}$ :

$$\begin{aligned} \frac{R_T}{R_{F0}} = & C_0^0 + C_1^0 \exp\left(-\frac{m}{9} F_n^{-2}\right) + C_2^0 \exp\left(-m F_n^{-2}\right) + \\ & + C_3^0 \exp\left(-m F_n^{-2}\right) \sin\left(F_n^{-2}\right) + \\ & + C_4^0 \exp\left(-m F_n^{-2}\right) \cos\left(F_n^{-2}\right). \end{aligned} \quad (50)$$

Für Gl. (49) folgt:

$$C_T = C_0^* \cdot C_{F0} + (\text{rechte Seite von Gl. (49)}).$$

Dies kann geschrieben werden:

$$\begin{aligned} \frac{C_T}{C_{F0}} = & C_0^I + C_1^I \cdot F_n^4 / C_{F0} + C_2^I \cdot F_n^5 / C_{F0} \cdot \cos\left(F_n^{-2} + \frac{\pi}{4}\right) + \\ & + C_3^I \cdot F_n^5 / C_{F0} \cdot \sin\left(F_n^{-2} + \frac{\pi}{4}\right). \end{aligned} \quad (51)$$

Gln. (50) und (51) enthalten zwar  $k$  nicht in explizierter Form - es verbirgt sich in  $C_0^0$  bzw.  $C_0^I$  -, doch hängen bei der Analyse eines Versuches wegen der Benutzung des Verhältnisses  $R_T/R_{F0}$  alle Koeffizienten, auch die des Wellenwiderstandes, vom Reibungswiderstand und der gewählten Reibungslinie ab. Dieses Verhältnis ist nicht nur für Modell und Schiff verschieden, sondern auch noch innerhalb des Modellbereiches uneinheitlich, weil es vom Maßstab abhängt. Dies ist Grund genug, um die Gleichungen in dieser Form nicht für eine statistische Widerstandsanalyse einzusetzen. Doch werden die ursprünglichen Wellenwiderstandsformeln im wesentlichen nur um ein konstantes Glied erweitert, und dies ist unschädlich, wenn nur geprüft werden soll, ob solche theoretisch entwickelten Formeln eine gemessene Widerstandskurve eng genug darzustellen vermögen.

Zur Anwendung auf die sechs Stichprobenschiffe wurden die Koeffizienten  $C_i$  durch Ausgleichsrechnung mittels der Methode der kleinsten Fehlerquadrate

berechnet. Die Größe  $m$  in Gl. (50) wurde nach einer Angabe von OORTMERSSEN gleich 0,15 gesetzt.

Um ein Maß für die durchschnittliche Abweichung zwischen Näherungs- und gemessener Widerstandskurve pro Schiff zu erhalten, wurden die Abweichungen an allen Meßpunkten in Prozent des gemessenen Widerstandes berechnet und ihr Mittelwert festgestellt. Die gemittelten prozentualen Abweichungen zeigt die folgende Tabelle:

	Tanker	Tanker	Tanker (Ballast)	Frachter	Passagierschiff	Bagger
OORTMERSSEN	0,2	0,5	1,0	1,8	0,9	0,7
INUI	0,1	0,1	0,1	4,5	0,8	0,6
$F_n$ von	0,128	0,107	0,107	0,025	0,186	0,122
bis	0,170	0,175	0,184	0,222	0,319	0,209

Die Beschreibung des Widerstandsverlaufes mit Hilfe der beiden Gleichungen ist mit einer Ausnahme (Frachter) außerordentlich gut. Die Gleichung von INUI scheint etwas wirksamer zu sein als die von OORTMERSSEN. Als graphisches Beispiel wurde auf Bild 7 der Frachter gewählt, bei dem die sichtbarsten Abweichungen auftraten. Untypisch bringt hier INUIS Formel größere Unterschiede hervor als OORTMERSSENS.

#### 4.2.2. Der Ansatz bei konstanter Froudezahl

Der Gesamtwiderstand wird in Reibungs- und Restwiderstand aufgespalten:

$$R_T = R_{FO} + R_R, \quad (53)$$

oder in dimensionsloser Schreibweise nach Division durch  $0,5 \rho V^2 S$ :

$$C_T = C_{F0} + C_R \quad (54)$$

$C_{F0}$  ist wiederum der Plattenreibungswiderstandsbeiwert.

$C_R$  enthält mit dem Formanteil des Zähigkeitswiderstandes einen zähigkeitsbedingten und mit dem Wellenwiderstand einen schwerkraftbedingten Anteil. Eine theoretische Gleichung für den Restwiderstand oder seinen Beiwert in geschlossener Form existiert nicht. Daher muß die statistische Analyse an festen Froudezahlen durchgeführt werden. Dies aber macht den Ansatz besonders geeignet zur Behandlung von unsystematischen Versuchsergebnissen, die sehr unterschiedliche Froudezahlbereiche überstreichen können.

## 5. Die verfügbaren Daten

Für die Regressionsanalyse des Widerstandes moderner Schiffsformen stand eine Ausgangsmenge von 554 Modellversuchen mit 308 Schiffsmodellen zur Verfügung. Die Messungen fanden von Ende 1967 bis Ende 1973 in der HSVA an 6 bis 9 m langen Modellen statt. Turbulenz wurde auf die in diesem Hause übliche Weise durch Sandpapierstreifen am Bug der Modelle erregt.

Die Schiffe dieser Stichprobe werden im folgenden mit Datenshippen bezeichnet. Sie gehören dem Typ der seegehenden Fracht- und Personenschiffe an. Ihre an festen Froudezahlen genommenen Restwiderstandsbeiwerte sind die abhängigen Variablen, ihre die Schiffsform beschreibenden Merkmale die unabhängigen Variablen in der Regressionsanalyse.

Tabelle 1 zeigt eine Aufgliederung der Datenshippen nach Tiefgang und Schraubenzahl. Neben der Beschaffenheit der Anhänge kann die Zahl der mit einem Bugwulst versehenen Schiffe entnommen werden. Der Konstruktionstiefgang ist als jeweils größter geschleppter Tiefgang auf ebenem Kiel definiert. Mit Zweittiefgang werden alle weiteren ebenen Ladetiefgänge bezeichnet, soweit sie nur geringfügig vom Konstruktionstiefgang abweichen und eindeutig nicht als Ballasttiefgänge anzusehen sind. Mit Ballasttiefgang werden alle ebenen Schwimmlagen mit  $B/T > 3,75$  bezeichnet.

Trimmlagen traten ausschließlich in Ballastzuständen ( $B/T > 3,75$ ) auf. Es wird zwischen einer Haupt- oder Ersttrimmlage und der Zweittrimmlage unterschieden. Erstere ist durch den größten mittleren Tiefgang bzw. bei mehreren gleichen mittleren Tiefgängen durch die kleinste Tauchungsdifferenz an den Loten gekennzeichnet, letztere enthält alle weiteren Trimmtiefgänge.

Weiter unten werden dimensionslose Formvariable eingeführt, in denen teilweise der jeweilige Tiefgang als Bezugsgröße verwendet wird. Um optisch hervorzuheben, daß es sich um vollbeladene- bzw. Ballastzustände handelt, trägt der Tiefgang bei den beiden Gruppen Konstruktionstiefgang und Zweittiefgang die Bezeichnung T, während für die restlichen drei Gruppen das Symbol T' verwendet wird. Bei Trimmlagen ist T' als Mittelwert der Tiefgänge am vorderen und hinteren Lot zu verstehen.

Die Tabelle enthält weiterhin eine Klassifizierung nach zehn Schiffstypen, in die sich sämtliche Datenschiffe ohne Schwierigkeiten einordnen ließen. Unter Sonderfahrzeuge wurden z.B. Bagger und Dockschiffe eingereiht.

### 5.1. Die abhängige Variable

Grundlage für die Bestimmung der Werte der abhängigen Variablen, des Restwiderstandsbeiwertes  $C_R$ , bildeten die Originalmeßwerte der Widerstandsversuche. Die gemessene Modellgeschwindigkeit  $V_M'$  wurde nach der Formel

$$V_M = V_M' \sqrt{\frac{A_T - A_M}{A_T}}$$

für Tankeinfluß korrigiert, wo  $A_T$  der Tankquerschnitt ( $18 \times 6 \text{ m}^2$ ) und  $A_M$  der Modellquerschnitt auf Hauptspant sind. Diese einfache Formel ist deshalb erlaubt, weil die Korrektur wegen des im Verhältnis zum Modellquerschnitt sehr großen Tankquerschnitts außerordentlich klein ist. Die gemessenen Gesamtwiderstände wurden mittels Division durch  $0,5\rho V_M'^2 S$  in dimensionslose Beiwerte  $C_T$  umgewandelt und unter Berücksichtigung der jeweiligen Tankwassertemperatur auf  $15^\circ \text{ C}$  standardisiert. Subtraktion des Reibungsbeiwertes der äquivalenten Platte  $C_{F0}$  nach der ITTC-Linie von 1957,

$$C_{F0} = \frac{0,075}{(\log R_n - 2)^2} \quad , \quad (55)$$

ergab die Restwiderstandsbeiwerte  $C_R$ .

Die Prüfung auf Verlässlichkeit der Meßwerte erfolgte graphisch durch folgende Auftragungen:

- 1)  $C_T$  über der Reynoldszahl  $R_n$ ,
- 2)  $C_T/C_{F0}$  über  $F_n^4/C_{F0}$ ,
- 3)  $C_R$  über  $F_n$  (nur bei Messungen mit  $F_n < 0,13$ ).

Die Auftragung von Beiwerten ist empfindlich gegenüber Meßfehlern. Fragwürdige Meßpunkte konnten auf diese Weise in einem gewissen Umfang erkannt und ausgeschlossen werden.

Insbesondere dienten die Auftragungen der Ermittlung der niedrigsten verlässlich erscheinenden Froudezahl, bei der der gemessene Widerstand noch nicht durch Laminareffekte oder Ablösungen im Hinterschiff beeinflusst war. Zeigten sich Anzeichen hierfür, so wurden alle betroffenen Meßpunkte ausgeschieden. Beispiele der Auftragungen zeigen die Bilder 8 bis 11.

Wahrscheinlich schließen Kontrollen dieser Art die versehentliche Verwendung unzuverlässiger Meßpunkte nicht vollkommen aus. Größere Sicherheit wäre durch Beurteilung der Streuung oder der Standardabweichung der einzelnen Widerstandsmeßwerte zu erzielen, doch standen derartige Angaben nicht zur Verfügung. So ist zu befürchten, daß die Daten insbesondere bei niedrigen Froudezahlen (etwa  $F_n < 0,20$ ) Einflüsse der Versuchstechnik und der Modellgröße enthalten, die die rein formabhängige Analyse des Widerstandes erschweren.

Die Widerstandspunkte liegen im allgemeinen auf beliebigen "krummen" Froudezahlwerten. Um eine Ordnung herzustellen, war es nötig, die Restwiderstandsbeiwerte an vorher festgelegten, geradzahligem Froudezahlen zu interpolieren. Dies geschah durch Ausgleichsrechnung über die Meßpunkte mit Hilfe einer kubischen Splinefunktion innerhalb des pro Versuchs gegebenen Froudezahlbereiches. Als Glättungsmaß wurde hierbei eine Abweichung von 1 % vom jeweiligen Meßwert vorgegeben, was aufgrund der verfahrensbedingten Meßgenauigkeit als vertretbar erschien. Interpolierte Restwiderstandsbeiwerte wurden an allen auf zwei Stellen hinter dem Komma glatten Froudezahlen innerhalb des Meßintervalls berechnet.

Individuelle Geschwindigkeitsmeßbereiche der einzelnen Modelle und das Aussondern verdächtiger Meßpunkte bei niedrigen Froudezahlen bedingten, daß von Froudezahl zu Froudezahl wechselnde Modellgruppierungen Beiträge zum Vorrat der Restwiderstandsbeiwerte lieferten. Statistisch stellen die bei einer bestimmten Froudezahl beobachteten  $C_R$ -Werte eine Stichprobe aus der Grundgesamtheit aller bei dieser Froudezahl feststellbaren Restwiderstandsbeiwerte dar. Die statistischen Richtlinien über Stichproben sind deshalb bei Auswertung derartiger Datenkollektive mittels statistischer Methoden zu beachten.

## 5.2. Die unabhängigen Variablen

### 5.2.1. Wahl der Variablen

Länge, Breite, Tiefgang und Verdrängung sind Größen, die unverzichtbare Grundinformationen über ein Schiff enthalten. Aus ihnen können drei unabhängige dimensionslose Verhältniswerte gebildet werden, gemeinhin  $L/B$ ,  $B/T$  und  $C_B = \nabla/LBT$ , aber auch z.B.  $L/B$ ,  $L/T$  und  $L/\nabla^{1/3}$  wären möglich. Solche Verhältniswerte werden im folgenden mit Hauptabmessungsvariable bezeichnet,  $C_B$  zuweilen auch als integrale Variable.  $L$  bezeichnet die Länge zwischen den Loten,  $B$  die Breite auf Spanten,  $T$  den Konstruktionstiefgang und  $\nabla$  das verdrängte Wasservolumen.  $C_B$  ist der Blockkoeffizient.

Die Bedeutung der Hauptabmessungsvariablen für die Beschreibung des Widerstandes ist nicht so groß wie ihr Wert für die Charakterisierung eines Schiffes. Versuche, den Widerstand mit diesen Variablen allein statistisch zu analysieren, scheitern immer wieder an zu wenig formspezifischer Information. Daß der Widerstand so breit wie beobachtet innerhalb relativ enger Grenzen der Hauptabmessungsvariablen streut, ist ohne Zweifel eine Folge unterschiedlichster Formdetails, auf die die statistische Analyse durch Berücksichtigung verfeinerter Formvariablen eingehen muß.

Natürlich ist es ein Unterschied, ob man die Ergebnisse einer formähnlichen Modellserie oder eines unsystematischen Kollektivs analysiert. Bei der Serie wird sich leicht auch eine starke Abhängigkeit von den Hauptabmessungsvariablen auffinden lassen, wenn diese zu den variierten Parametern zählen, in dem unsystematischen Kollektiv jedoch herrschen weniger strenge Zusammenhänge. Von Anfang an erschien es daher notwendig, außer den Hauptabmessungsvariablen weitere, der Gesamtschiffsform im allgemeinen und der lokalen Schiffsform im besonderen Rechnung tragende Variable bereitzustellen.

Die Aufgabe ist leichter formuliert als gelöst. Ohne Mühe können beliebig viele formspezifische Variable genannt werden, doch Schwierigkeiten liegen erstens in der Frage, wie sie unstrittig zu definieren sind, zweitens in der Frage, wie unabhängig sie voneinander sind, und drittens in dem Problem, wie man sie ohne übertriebenen Aufwand beschafft.

Unter der Annahme, daß eine verlässliche Vorhersage des Widerstands nur möglich ist, wenn man wenigstens über eine (vorläufige) Zeichnung der Schiffslinien verfügt, wurden die folgenden zusätzlichen Formvariablen eingeführt:

- 1) Längen- und Höhenschwerpunkt der Verdrängung  $x_B$  und  $z_B$ , Wasserlinienswerpunkt und -völligkeit  $x_F$  und  $C_{WP}$  als "integrale" Variable der Gesamtschiffsform (der Ausdruck "integral" wurde gewählt, weil zur Bestimmung dieser Variablen Integration über eine Fläche oder ein Volumen notwendig ist, diese Variablen daher im Unterschied z.B. zu L/B einen integralen Charakter der Formbeschreibung tragen),
- 2)  $(C_B, x_B, z_B, x_F, C_{WP})^E$  und  $(C_B, x_B, z_B, x_F, C_{WP})^R$  für Bug- und Heckteile der Datenschiffe sowie  $(C_B, x_B, z_B)^B$  für evtl. vorhandenen Wulst als integrale Variable der lokalen Schiffsform (Indices E = Entrance, R = Run, B = Bulb),
- 3) Dimensionslose Aufmaße und Winkel an gewählten Spanten in Vor- und Hinterschiff als punktuelle Variable der lokalen Schiffsform.

Bug und Heckteil der Schiffe wurden durch eine Länge von 20% L definiert (zur bequemen Anwendung der Simpsonschen Regel mittels der Konstruktionspannten 16 bis 20 bzw. 0 bis 4).

Mit Wulst wurde der Unterwasserteil eines Schiffes vor dem vorderen Lot bezeichnet.

Nach Sichtung zahlreicher Spantrisse erwiesen sich die Konstruktionspannten 18 und 19 im Vorschiff sowie 1 und 2 im Hinterschiff als besonders charakteristisch; Aufmaße und Winkel in ihren Spantebenen wurden an den Stellen 0,1; 0,25; 0,5; 0,75 T genommen. Die Aufmaße wurden mit der Schiffsbreite auf den Wertevorrat 0 bis 1 normiert. T ist der Konstruktionstiefgang.

Als weitere Größen, mit denen Variable gebildet werden konnten, standen die Wasserlinienlängen  $L_{WL}$  sowie die Wulstlängen  $L^B$  = größte Länge vor dem vorderen Lot,  $H^B$  = Höhe des vordersten Punktes über Basis und  $B^B$  = größte Breite am vorderen Lot zur Verfügung. Auf Variablenbildung mit der Schiffsoberfläche S wurde verzichtet, da diese Größe bereits zur Berechnung dimensionsloser Widerstandsbeiwerte verbraucht war.

Auf folgende Weise wurden die Variablen dimensionslos gemacht: Längenschwerpunkt mit  $L$ , Höhengschwerpunkte mit  $T$  bzw.  $T'$ , Blockkoeffizienten mit  $LBT$  bzw.  $LBT'$ , Wasserlinienvölligkeiten mit  $L \cdot B$ , Wulstmaße mit  $L$ ,  $B$ ,  $T$  bzw.  $L$ ,  $B$ ,  $T'$ .

Die genannten unabhängigen Variablen können vielleicht als willkürlich gewählt angesehen werden, doch sind sie größtenteils anschaulich und eindeutig definiert.

Wenn ein Linienriß vorliegt, sind sie außerdem mit erträglichem Aufwand zu beschaffen. Im Rahmen der vorliegenden Arbeit wurden die integralen Variablen  $(C_B, x_B, z_B, x_F, C_{WP})$  Schiff,  $E, R, B$  mit Hilfe des Programms ARCHIMEDES von SÖDING ermittelt. Hierfür war es erforderlich, die Datenschiffe aufzumessen. Dies erfolgte am digitalen Koordinatenmeßtisch der HSVA. Aufmaße und Winkel wurden aus den hiernach verfügbaren Unterlagen numerisch ermittelt. Das Programm ARCHIMEDES lieferte zusätzlich die integralen Variablen  $(C_B, x_B, z_B, x_F, C_{WP})^M$  des 60 %  $L$  langen Mittelschiffs.

### 5.2.2. Faktorenanalysen der unabhängigen Variablen

Von den unabhängigen Variablen eines Regressionsproblems will man in erster Linie wissen: Beschreiben sie auf wirklich vielfältige und verschiedene Weise den Gegenstand, dessen Eigenschaften man analysieren will? Oder drücken sie als Abkömmlinge eben dieses Gegenstandes immer nur dasselbe aus und belasten durch ihre Vielzahl eher die Analyse anstatt sie zu fördern? Statistisch gesehen ist dies die Frage nach der Unabhängigkeit der "unabhängigen" Variablen. Sie kann daher auch anders formuliert werden: sind hinreichend viele unabhängige Grundinformationen vorhanden, oder müssen weitere Variable bereitgestellt werden, um eine erfolgreiche Regressionsanalyse zu gewährleisten?

Über die wirkliche Unabhängigkeit der vorgesehenen Formvariablen ist wenig bekannt. Allenfalls weiß man, daß  $L/B$  und  $L/v^{1/3}$  miteinander verwandt sind und gleiches auch für  $x_B$  und  $x_F$  gelten dürfte. Starker Verdacht auf gegenseitige Abhängigkeit richtet sich zweifellos gegen Aufmaße und Winkel benachbarter Spanten, aber ist ein solcher Verdacht wirklich berechtigt? Man weiß es nicht.

Mit gutem Recht kann man argumentieren, daß die Regressionsanalyse selbst die Unabhängigkeit der unabhängigen Variablen kontrolliert und nur die wahrhaft unabhängigen in die Regressionsgleichung aufnimmt, die anderen aber ausschließt. Verläßt man sich aber ausschließlich auf diesen eingebauten Mechanismus, so müßten sämtliche vorgesehenen Variablen zugleich in die Analyse gegeben werden. Mit Rücksicht auf numerische Genauigkeit und begrenzten Raum im Computer wird dies kaum je möglich sein, umso weniger deshalb, weil nicht nur die lineare Grundform der Variablen infrage kommt, sondern auch Kreuzprodukte und eine unbegrenzte Zahl von Funktionen, die zu einer Verbesserung des Zusammenhanges mit der abhängigen Variablen beitragen könnten. Unter diesen Umständen wird man nicht um eine Aufspaltung in Variablengruppen herumkommen, und darin liegt die Gefahr, eine möglicherweise entscheidende Kombination zu verpassen.

Um diese Gefahr abzuschwächen, wurden Faktorenanalysen mit den vorgesehenen unabhängigen Variablen durchgeführt. Es war das Ziel dieser Analysen, Gemeinsamkeiten, insbesondere aber unterschiedliche Informationsinhalte der Variablen aufzuzeigen. Mit Kenntnis dieser Eigenschaften sind eher Variablenkombinationen mit in sich abgeschlossenen Informationswerten zu bilden, die keine Überschneidungen mit anderen, gerade nicht berücksichtigten Variablen aufweisen und daher unabhängig von diesen in einer Regressionsanalyse mit dem Widerstand korreliert werden können. Durch Zusammenfassen der in Zwischenregressionen als signifikant erkannten Variablen wird sich so eine reduzierte endgültige Variablenauswahl treffen lassen, die ein Optimum an Effizienz erreicht.

Von den zahlreichen Untersuchungen auf Gemeinsamkeit in den unabhängigen Variablen werden vier Gruppierungen vorgestellt, die besonderes Interesse verdienen.

Die erste Gruppe umfaßt all die Formvariablen, die in den Arbeiten von DOUST [1], SABIT [3,4], SWIFT [5] und OORTMERSSSEN [6] Verwendung fanden. Es handelt sich hierbei im wesentlichen um Hauptabmessungsvariable sowie  $C_M$ ,  $x_B$ ,  $C_{WP}$  und  $i_E$ , einschließlich Kreuzprodukten und Potenzen. Eine Auflistung der insgesamt 56 Variablen findet sich in Tabelle 2, deren spezielle Bedeutung erst weiter unten erläutert werden kann.

Die zweite Gruppe umfaßt die in 5.2.1. genannten Formvariablen mit Ausnahme der Aufmaße und Winkel, im einzelnen die drei Hauptabmessungsvariablen  $L/B$ ,  $B/T$  und  $\nabla/L^2T$ , die Formvariablen  $C_p$ ,  $x_B$ ,  $z_B$ ,  $x_F$  und  $C_{WP}$  für das gesamte Schiff sowie die Schiffsteile, ferner  $i_R$  und  $i_E$ . Letztere wurden (wie auch schon  $i_E$  in Gruppe eins) anders als in [17] als Spitzenwinkel der gleichschenkeligen Dreiecke definiert, die die Fläche  $C_{WP}^{E,R} \cdot LB$  und den Schwerpunkt  $x_F^{E,R}$  besitzen. Insgesamt handelt es sich um 28 Variable in linearer Form (Tabelle 5).

Die vierte Gruppe schließlich faßt alle in 5.2. genannten Formvariablen zusammen. Es wurde nur die lineare Form der Variablen verwendet, deren Zahl 61 beträgt (Tabelle 11).

Zahlenwerte für die Variablen der Gruppen eins und zwei wurden 217 Datenschiffen auf Entwurfstiefgang entnommen, die zum Typ Frachter, Tanker, Erz/Oboschiff und Container gehörten und mit nur einer Schraube ausgerüstet waren. Für die Variablen der dritten Gruppe wurden Aufmaße und Winkel von 277 Datenschiffen aller Typen einschließlich Schleppern und Fischereifahrzeugen verwendet. Die Daten der vierten Gruppe stammen von 250 Schiffen aller Typen auf Entwurfstiefgang außer Schleppern und Fischereifahrzeugen. Die Ergebnisse der Faktorenanalyse sind in den Tabellen 2 bis 13 zusammengestellt und werden im folgenden diskutiert.

Aus den 56, von verschiedenen Autoren bisher verwendeten Variablen der ersten Gruppe wurden nur fünf Faktoren mit Eigenwerten  $\lambda < 0,5$  extrahiert (Tabelle 3). Der erste Faktor erklärt fast 55 % der Gesamtvarianz, der zweite 21,5 %, der dritte 18 %. Auf den vierten und fünften entfallen mit 2,2 bzw. 1,5 % schon so geringe Varianzanteile, daß diese Faktoren gegenüber den ersten drei mit gutem Recht vernachlässigt werden könnten. Alle Varianzangaben beziehen sich auf die unrotierten Faktoren. Die Eigenwerte der rotierten Faktorenmatrix, die mit den Werten der Transformationsmatrix (Tabelle 4 unten) nach Gl. (43) berechnet werden können, weichen kaum von denen der unrotierten ab. Die varimax rotierte Faktorenmatrix (Tabelle 4 oben) läßt anhand der Ladungen die Zuordnung der Variablen zu den Faktoren erkennen, eine entsprechende Auflistung nach Ladungen  $\approx \pm 0,8$  zeigt Tabelle 2.

Was bedeutet dieses Ergebnis der Faktorenanalyse? Die gewählte Varimaxmethode der Rotation schafft rechtwinklig zueinander stehende Faktoren, ihre physikalischen Bedeutungen, auch Grundinformationen genannt, sind daher unabhängig voneinander, das heißt gänzlich verschieden. Ohne zunächst die physikalischen Bedeutungen mit Worten benennen zu können, kann gesagt werden, daß sich in den 56 Variablen im wesentlichen drei, mit Einschränkungen fünf, sehr verschiedene Grundinformationen verbergen. Jede dieser Grundinformationen wird von einer ganzen Reihe von Variablen zugleich getragen (Tabelle 2), d.h. aber, jede Variable einer solchen Reihe ist nur eine spezielle, uns zugängliche Erscheinungsform ein und derselben Grundinformation. Im Grunde genügt schon eine dieser Variablen (besonders die mit der höchsten Ladung), um in einer Auswahl wirklich unabhängiger Variablen die jeweilige Grundinformation ausreichend zu vertreten.

Es ist nun wohl doch überraschend, daß sich die 56 Variablen, die von den verschiedenen Autoren und vom Verfasser selbst für differenziert und mit unterschiedlichem Informationsgehalt ausgestattet gehalten wurden, auf nur drei starke Grundinformationen reduzieren lassen. Der erste Faktor vereinigt neben den Völligkeitsparametern  $C_B$  und  $C_p$  unter anderem auch den Längenschwerpunkt, die Wasserlinienvölligkeit und den Wasserlinieneintrittswinkel auf sich, was sicher kaum vermutet wurde. Da dies aber so ist, und auch Kreuzprodukte dieser Größen offenbar keine zusätzlichen Informationswerte erzeugen, läßt sich leichter erklären, warum Widerstandsanalysen mittels derartiger "unabhängiger" Variabler der Vielfalt gemessener Widerstandswerte nicht gerecht wurden und nur mäßige Ergebnisse lieferten.

Es soll im folgenden versucht werden, die Bedeutung der Faktoren zu klären, d.h. die im Hintergrund stehende Größe zu benennen. Der physikalische Hintergrund von Faktor 1 wird am ehesten durch die höchstladende Variable erklärt. Dies ist  $C_p$  (Tabelle 4), und es ist wohl nicht falsch, den Faktor als Ausdruck der Völligkeit von Schiffen zu deuten. Die Wirkung dieser Grundinformation ist so stark, daß sie sich selbst in Kreuzprodukten durchsetzt, die zwischen den eigentlichen Völligkeitsvariablen  $C_B$  bzw.  $C_p$  und anderen Variablen gebildet wurden.

Der zweite Faktor erhält seine höchsten Ladungen von L/B. Auch seine anschauliche Deutung ist damit gesichert. In Kreuzprodukten, selbst mit  $C_p$ , dominiert L/B, allerdings vorwiegend, wenn es in quadratischer Form erscheint.

In gleicher Weise ist es möglich, den dritten Faktor als von B/T abhängig zu identifizieren.

Faktor vier hat keine ausgeprägt hohen Ladungen. Die stärksten kommen von  $i_E$ , aber  $i_E$  lädt Faktor 1 noch höher. Faktor 5 wird von  $C_M$  geprägt.

Eine Reihe von Variablen lädt mehrere Faktoren gleichzeitig, dann aber mit kleineren Ladungsbeträgen. Diese Variablen enthalten gemischte Informationen.

Aus den Ergebnissen dieser ersten Faktorenanalyse wurden Schlüsse gezogen, die sich auch auf die folgenden Faktorenanalysen und die Regressionsanalyse ausgewirkt haben. Die Richtigkeit und Allgemeingültigkeit der Schlußfolgerungen können durchaus angezweifelt werden, und dieser Bericht muß die entsprechenden Beweise schuldig bleiben, aber es spricht ein gut Teil Wahrscheinlichkeit für sie. Sie bringen unter anderem den großen Vorteil mit sich, die Zahl der für die Regressionsanalyse in Betracht zu ziehenden unabhängigen Variablen erheblich einzuschränken. Die Beobachtungen und Schlüsse sind folgende:

Funktionen der Variablen scheinen die Grundinformationen der linearen Variablenform nicht zu verändern. Z.B. laden Potenzen von  $C_B$  bzw. L/B dieselben Faktoren wie  $C_B$  bzw. L/B selbst. Aus diesem Grund wurde in den weiteren Faktorenanalysen lediglich die lineare Form der Variablen berücksichtigt:

In vielen Kreuzprodukten dominiert diejenige einfache Variable, die den Kern eines Faktors bildet ( $C_p$ , L/B). Neue Grundinformationen werden durch Kreuzprodukte offenbar nicht geschaffen. Deshalb wurde im weiteren Teil dieser Arbeit auf Verwendung von Kreuzprodukten verzichtet, lediglich die Variablen der letzten Spalte in Tabelle 2 wurden in den Widerstandsregressionen eingesetzt, weil mit diesen Variablen vielleicht ein Zusammenhang mit den Widerstandsanalysen anderer Autoren herzustellen ist.

Mehrfachinformationen enthaltende Variable sind vermutlich für Regressionsanalysen von besonderer Bedeutung. Als Beispiel hierfür sei  $L/\nabla^{1/3}$  genannt, das möglicherweise wegen seiner Mittlerfunktion zwischen den Grundinformationen Völligkeit und L/B gut für die Beschreibung des Widerstandes geeignet ist. Auf Berücksichtigung solcher Variablen wurde in den Regressionsanalysen besonders geachtet.

Aus den 28 überwiegend integralen Variablen der zweiten Gruppe wurden sieben Faktoren mit Eigenwerten  $\lambda > 0,5$  extrahiert (Tabellen 6 und 7). Der erste Faktor erklärt fast zwei Drittel der Gesamtvarianz, die nächsten vier Faktoren mit etwa 10, 8, 6 und 4 % erklärter Varianz sind deutlich schwächer, Faktor 6 und 7 mit je fast 2 % nahezu unbedeutend.

Die Zuordnung der Variablen zu den Faktoren aufgrund eines Ladungsbeitrages  $\approx 0,7$  zeigt Tabelle 5. Fast sämtliche integralen Formvariablen des Gesamtschiffes und des Mittelschiffteils laden Faktor 1, die folgenden Faktoren werden in dieser Reihenfolge von Heck-, Bug-, Wulstvariablen, L/B, den  $z_B$ -Variablen und B/T geladen. Nur wenige Variable enthalten Doppel- oder Mehrfachinformationen.

Die Höhe der erklärten Varianz des ersten Faktors täuscht, sie wäre geringer, wenn die Mittelschiffsvariablen fortgelassen würden (die Zahl der Variablen, die einen Faktor laden, erhöht dessen Varianzbeitrag). Das Ergebnis dieser Analyse kann daher wie folgt zusammengefaßt werden: Die integralen Formvariablen enthalten Informationen, die durch ihre Herkunft gekennzeichnet sind, d.h. Gesamtschiff, Heck, Bug und Wulst. L/B und B/T bleiben selbständig, und die Höhenschwerpunkte spielen eine halbwegs eigenständige, wenn auch beinahe nebensächliche Rolle.

Aus den 16 dimensionslosen Aufmaßen und den 16 Winkeln der dritten Gruppe wurden 8 Faktoren mit Eigenwerten  $\lambda > 0,5$  extrahiert. Sie erklären im einzelnen 41, 19, 13, 6, 5, 3, 2 und 2 % der Gesamtvarianz (Tabelle 9 und 10).

Sämtliche Aufmaße der Vorschiffsspanten 18 und 19 laden den ersten Faktor (Tabelle 8), aber auch die Ladungen der Hinterschiffsspantaufmaße auf diesem Faktor sind nicht unbedeutend. Letztere prägen außerdem gemeinsam den Faktor 5, und die obersten Aufmaße an Spant 1 und 2 bilden den schwachen Faktor 8. Dies ist kein überraschendes Ergebnis, weil man von vornherein erwartet, daß Spantaufmaße voneinander abhängig sind.

Aufmaße und Winkel korrelieren jedoch nicht nennenswert miteinander. Man kann Aufmaßfaktoren (1, 5 und 8) sauber von Winkelfaktoren (2, 3, 4, 6 und 7) unterscheiden. Faktor 2 wird im wesentlichen von oberen Winkeln der beiden Vorschiffsspanten geladen, Faktor 3 von unteren Winkeln des Spantes 2, Faktor 4 von oberen Winkeln der Spanten 1 und 2. Faktor 6 hängt von unteren Winkeln an Spant 18 und 19 ab, Faktor 7 von unteren Winkeln an Spant 1.

Die Unabhängigkeit der Winkel untereinander ist insgesamt größer als die der Aufmaße. Die untersuchte Auswahl bringt relativ viele und, abgesehen vom ersten, auch relativ gleichstarke Faktoren hervor. Sie enthält demnach zahlreiche unterschiedliche Informationen über die Schiffsform.

Durch die Faktorenanalyse der vierten Variablengruppe sollte ermittelt werden, wieviele voneinander unabhängige Grundinformationen die Gesamtheit der in 5.2.1. erwähnten Variablen der Schiffsform in Wirklichkeit besitzt. Der grundlegende Zweck dieser Untersuchung bestand darin, ein Kriterium für eine ausreichende Anzahl der Variablen im Hinblick auf die Regressionsanalyse in die Hand zu bekommen. Leitender Aspekt hierbei war, Variable mit möglichst vielfältigem und sich nicht gegenseitig bedingendem Informationswert zur Verfügung zu haben.

Aus den 61 linearen Variablen wurden 13 Faktoren mit Eigenwerten  $\lambda \geq 0,5$  extrahiert (Tabellen 12 und 13). Der erste Faktor mit 49,5 % erklärter Varianz ist bei weitem der stärkste, der zweite und der dritte tragen mit 14,0 bzw. 9,1 % zur Varianzerklärung bei, auf die restlichen entfallen Beiträge zwischen 4,2 und 0,9 %.

Die Zahl der extrahierten Faktoren ist recht groß, und die Stärke des ersten Faktors erklärt sich überwiegend aus der Menge der ihn ladenden Variablen. Tabelle 11 zeigt, daß er mit 31 Variablen mehr als die Hälfte aller verfügbaren Variablen mit Ladungen  $\geq 0,7$  auf sich vereinigt. Diese Variablen besitzen somit etwa gleichartigen Informationswert. Zu ihnen zählen fast sämtliche integralen Formvariablen des Gesamtschiffes und der Schiffsteile sowie alle Aufmaße an den ausgewählten Spanten, wenn man auch etwas schwächer ladende Variable berücksichtigt. Analysiert man die Bedeutung des ersten Faktors mit Hilfe der Höhe seiner Ladungsbeträge, so stößt man auf die Schiffsvölligkeit als sich in ihm manifestierende Größe.

Integrale Formvariable des Heckteils laden den Faktor 2 (neben Aufmaßen an den Heckspanten 1 und 2), die Wulstvariablen den Faktor 5.  $B/T$ ,  $L/B$  und  $C_M$  erweisen sich als Merkmale mit weitgehend eigenständigem Informationsgehalt (Faktor 7, 10 und 11). Aufmaße am Fuß der Heckspanten 1 und 2 sind, wenngleich schwach, an der Erzeugung des Faktors 9 beteiligt. Die restlichen 6 Faktoren werden von den Spantwinkeln geladen.

Aus der hohen Zahl von winkelabhängigen Faktoren darf der Schluß gezogen werden, daß Gradienten sehr charakteristische und voneinander und von anderen Variablen unabhängige Beschreibungsmittel der Schiffsform sein können. Daher wird die Empfehlung ausgesprochen, bei zukünftigen Analysen von Meßwerten, die als von der Schiffsform abhängig zu verstehen sind, verstärkt auf derartige Variable zurückzugreifen. Hydrodynamische Gesichtspunkte gestatten die Vorhersage eines wahrscheinlichen Erfolges für Formgradienten, die näherungsweise in Richtung der Schiffsumströmung liegen. Aus den hier verfügbaren Variablen konnten zusätzlich auf Wasserlinien liegende Winkel mit Hilfe der Aufmaße an den Nachbarspanten 1 und 2 bzw. 18 und 19 gebildet werden, die sich in den Regressionsanalysen auch als erfolgreich erwiesen.

Das Ergebnis der vierten Faktorenanalyse erlaubt die Bildung von kleineren Variablengruppen, die voneinander unabhängige Informationswerte besitzen und daher am ehesten in den der endgültigen Analyse vorangehenden Zwischenregressionen erprobt werden können, ohne möglicherweise entscheidende Kombinationen zu verpassen. Geeignete Gruppen sind z.B. die Variablen der Faktoren 1 und 9 und die Variablen der restlichen Faktoren. Auch andere Gruppierungen sind möglich, doch wurden im wesentlichen die genannten einzeln untersucht. Der zweiten Gruppe wurden noch Variable hinzugefügt, die die aus Aufmaßen berechneten Wasserlinienwinkel und die verschiedenartige Ausbildung der Anhänge der Datenschiffe berücksichtigten.

Insgesamt entstand der Eindruck, daß die verfügbaren Variablen auf recht vielfältige Weise die Schiffsform zu beschreiben vermochten. Natürlich lieferten die Faktorenanalysen keinen Aufschluß über die Qualität, mit der die Variablen mit dem Restwiderstandsbeiwert oder dem Formfaktor korrelieren; dies zu ermitteln war wie gesagt die Aufgabe der Regressionsanalyse. Aber sie zeigten eindrucksvoll, daß viele der bisher für unabhängig gehaltenen Formvariablen dies gar nicht sind, sondern vielfach eine mit den einfachen Völligkeitsparametern  $C_B$  oder  $C_p$  identische Bedeutung haben.

### 5.2.3. Funktionale Zusammenhänge mit dem Restwiderstandsbeiwert

Während Funktionen der unabhängigen Variablen, z.B. Potenzfunktionen, die Grundinformation nicht zu verändern scheinen, so daß in Faktorenanalysen die lineare Form der Variablen untersucht werden konnte, ist in Regressionsanalysen der funktionale Zusammenhang zwischen einer unabhängigen Variablen und der Zielgröße von großer Bedeutung. Wenn in Wirklichkeit hierfür ein nichtlineares Gesetz gilt, muß die Variable in der entsprechenden Form dem Regressionsverfahren zugeführt werden. Die lineare multiple Regression nimmt jede Variable in der ihr vorgelegten Form an, besser geeignete Funktionen der Variablen erzeugt sie nicht.

Es existieren keinerlei bekannt gewordene numerische Untersuchungen über funktionale Zusammenhänge zwischen Formvariablen und dem Restwiderstandsbeiwert, und auch das Gespräch mit erfahrenen Praktikern ergab keine Anhaltspunkte. Ebenso wenig liefert die Theorie Aufschluß; als einziger direkter Formparameter tritt das Verhältnis L/B in der Theorie des Wellenwiderstandes des dünnen Schiffes auf. In den bisher veröffentlichten Regressionsanalysen finden sich einfache Variable in quadratischer oder kubischer Form, oder auch Kreuzprodukte einfacher Variabler, aber es fehlen Hinweise darauf, daß eine Prüfung auf Eignung stattgefunden hätte.

Einen offenbar brauchbaren, aber erweiterungsbedürftigen Ansatz für ein Verfahren, das das Aufsuchen geeigneter Funktionen auf dem Computer automatisiert, hat WESTERVELD [18] geliefert.

Hier jedoch wurde optisch-manuell vorgegangen. In die Punktauftragungen Restwiderstandsbeiwert über Formvariablen wurden von Hand und nach Augenschein interpolierende Kurvenzüge eingezeichnet, deren Verlauf dann durch den mathematischen Ansatz

$$C_R = \text{const. (Variable)}^a$$

beschrieben wurde. Hierin ist a der Exponent, der den gesuchten Zusammenhang mit Hilfe der einfachst möglichen Potenzfunktion darstellt.

Des bedeutenden Zeitaufwandes wegen wurden die Zusammenhänge nur an einer bestimmten Schiffsgruppe, nämlich Einschraubenschiffe der Typen Frachter, Tanker und Erz/Obo-Schiffe auf Entwurfstiefgang, bei den als charakteristisch angesehenen Froudezahlen 0,14, 0,20 und 0,26 untersucht.

Diagramme wurden für die Variablen  $L/B$ ,  $B/T$ ,  $\nabla/L^3$ ,  $L/\nabla^{1/3}$ ,  $LT/B^2$ ,  $C_M$ ,  $C_p$  sowie für die Variablen  $C_B$ ,  $x_B$ ,  $z_B$ ,  $x_F$  und  $C_{WP}$  des Gesamtschiffes und der vier Schiffsteile maschinell gezeichnet. Eine Übersicht über die ermittelten Exponenten der Interpolationskurven zeigt Tabelle 14. Beispiele der graphischen Untersuchung werden in den Bildern 12 bis 23 für die wichtigsten Hauptabmessungsvariablen und  $x_B$  gezeigt. Die dort gestrichelt eingezeichneten Linien verbinden  $C_R$ -Werte von Modellen der SERIES 60 [19,20], deren Hauptabmessungsverhältnisse und Froudezahlen jeweils angegeben sind. In einigen Fällen decken sie oder ihre Extrapolationen sich mit Interpolationskurven, die man optisch durch die Punkthaufen legen würde. Ihr Nutzen bestand häufig allein schon darin, Tendenzen in den stark streuenden Punktverteilungen überhaupt erkennbar werden zu lassen.

Vielfach erlaubt die große Streuung der  $C_R$ -Werte lediglich einen linearen Funktionsansatz, in einigen Fällen jedoch war ein tendenziöser Verlauf zu erkennen, so daß Exponenten  $a$  oder Exponentbereiche relativ sicher bestimmt werden konnten. Wegen der Größe der  $C_R$ -Stichproben besteht kaum Anlaß, an der Allgemeingültigkeit der Kurventendenzen zu zweifeln, ob sie allerdings extrapolationssicher sind, ist ungewiß.

Die Höhe mancher Exponenten mag überraschen. Sie erklärt sich aus dem Wertebereich der Variablen, der bisweilen eine hohe Variabilität in  $C_R$  mit einer geringen Variabilität in einem Formmerkmal verband. Natürlich ist zu berücksichtigen, daß die bei den Froudezahlen 0,14; 0,20 und 0,26 gefundenen Funktionen nicht unbedingt auch an den benachbarten Froudezahlen gelten müssen und den Exponenten daher noch ein gewisser Spielraum zuzugestehen ist. Dieser Spielraum aber konnte auf der Grundlage der aus der Untersuchung gewonnenen Richtwerte durch Probieren in den Regressionsanalysen relativ rasch eingeengt werden.

## 6. Ergebnisse der Restwiderstandsanalyse

### 6.1. Gruppeneinteilung und Bemerkungen zur Durchführung der Berechnungen

Nach der ausführlichen Schilderung des Regressionsverfahrens und der begleitenden Nebenuntersuchungen kann vorausgesetzt werden, daß eine knappe Darstellung der Ergebnisfindung und der Ergebnisse selbst genügt. Die Notwendigkeit statistischer Betrachtungen und Tests wurde an anderer Stelle begründet und muß gegebenenfalls dort nachgelesen werden.

Nach Durchsicht der Spantrisse der 308 Datenschiffe wurden 37 Schiffe wegen ungewöhnlicher Form ausgeschieden. Die restlichen gliederten sich in drei Gruppen, eine große, die die relativ formähnlichen Fracht- und Personenschiffe umfaßt, und zwei kleinere, die von Schleppern und Fischereifahrzeugen sowie Spezialschiffen gebildet wurden. Die beiden kleineren Gruppen wurden nicht analysiert. Lediglich die erste eignet sich aufgrund der großen Anzahl ihrer Schiffe für Widerstandsanalysen.

Die Anzahl der mit diesen Schiffen durchgeführten Widerstandsversuche mußte wegen zweifelhafter Meßergebnisse, zu geringer Meßpunktzahl oder zu weiter Punktfolge um einige Messungen reduziert werden. Des weiteren erforderten die hydrodynamischen Unterschiede der Unterwasserform die Bildung von zwei Tiefgangsgruppen. Die erste umfaßte schließlich unter der Bezeichnung Entwurfs-tiefgang 250 Versuche aus den ersten beiden Tiefgangsabteilungen der Tabelle 1, die zweite unter der Bezeichnung Ballasttiefgang 175 Versuche aus den drei letzten Tiefgangsabteilungen derselben Tabelle.

Von Froudezahl zu Froudezahl traten  $C_R$ -Stichproben unterschiedlicher Größe auf. Von jeder wurden Histogramme der Verteilung und der Summenhäufigkeit erstellt, und die Verteilung wurde mit der Normalverteilung verglichen.

Es zeigte sich, daß die Gruppe der Fracht- und Personenschiffe auch statistisch ein harmonisches Kollektiv formt, da in den  $C_R$ -Verteilungen keine unerträglichen, typenabhängigen Häufungen auftraten. Natürlich überwogen bei kleineren Froudezahlen die Anteile von völligen Schiffen wie Tankern, bei höheren die von schlanken wie Containerschiffen, doch lagen die  $C_R$ -Werte im Überlappungsbereich durchaus vermischt. Tests auf  $C_R$ -Ausreißer eliminierten vornehmlich Beiträge solcher Schiffe, die bei der jeweiligen Froudezahl mit unökonomischer Geschwindigkeit betrieben worden wären. (In diesem Zusammen-

hang sei aber darauf hingewiesen, daß es für das Ziel dieser Arbeit nicht erforderlich ist, daß Beiträge zu den einzelnen Froudezahlen aus der unmittelbaren Umgebung der Entwurfsgeschwindigkeit eines Datenschiffes stammen. Die Regressionsanalyse ist ein Verfahren der Interpolation, nicht aber der Optimierung. Daher sind auch nicht optimale  $C_R$ -Werte als zur Stichprobe gehörig anzusehen, solange sie keine Ausreißer sind.) Weiterhin wurden Untersuchungen auf Zufälligkeit durchgeführt. Die Tests zeigten, daß mit Ausnahme weniger Fälle Anzeichen der Nichtzufälligkeit vorlagen. Dies warnt vor Schwerpunktbildung von der Art, daß Häufungen gleicher Formvariabler zu einer Häufung ähnlicher  $C_R$ -Werte führten, und verweist darauf, daß die Verteilung der für die Regressionsanalysen verwendeten Formvariablen bei Anwendung der Ergebnisse beachtet werden sollten. Die statistischen Untersuchungen wurden mit der Wahrscheinlichkeit  $\alpha = 95 \%$  durchgeführt, zuweilen zur besseren Bewertung der Signifikanz auch mit  $\alpha = 99 \%$ . Sämtliche erwähnten Testergebnisse sind den Regressionsergebnissen beigelegt und sind als Teil derselben anzusehen und zu beachten.

Als untere Grenze für ein zu analysierendes  $C_R$ -Kollektiv wurde nach der Elimination von Ausreißern die Anzahl  $n = 40$  festgelegt. Für die Entwurfs-tiefgangsgruppe ergaben sich hiermit ausreichend Stichprobengrößen an den Froudezahlen 0,12 bis 0,31 für die Ballasttiefgangsgruppe und den Froudezahlen 0,12 bis 0,33 mit der Schrittweite 0,01. Für die hierdurch umrissenen Froudezahlbereiche liegen Regressionsergebnisse vor.

Vorbereitende Regressionsanalysen des Restwiderstandsbeiwertes  $C_R$  mit verschiedenen Teilkombinationen der unabhängigen Variablen und ihren Funktionen wurden unter dem Aspekt durchgeführt, daß jede Teilkombination einen so weit wie möglich abgeschlossenen Bereich der zur Verfügung stehenden Grundinformationen enthalten sollte. Die Variablengruppierungen erfolgten aufgrund der Ergebnisse der vierten in 5.2.2. wiedergegebenen Faktorenanalyse. Die jeweils signifikanten unabhängigen Variablen wurden herausgezogen und zu den in den endgültigen Regressionen verwendeten Kombinationen zusammengefaßt.

Aus diesen endgültigen Kombinationen wählte die Analyse von Froudezahl zu Froudezahl fast immer unterschiedliche unabhängige Variable aus. Wo sie übereinstimmten, wich die aufgrund der Signifikanztests ermittelte Rangfolge in der Beschreibung des Restwiderstandsbeiwertes im allgemeinen voneinander ab. Daher ist zu erwarten, daß die im Anwendungsfall über ein größeres Froude-

zahlintervall berechneten  $C_R$ -Werte streuen oder springen, so daß eine vermittelnde Kurve unter Zuhilfenahme der Vertrauensgrenzen hindurchgelegt werden muß.

Die Tabellen 15 bis 18 enthalten Numerierung und Bedeutung der unabhängigen Variablen, die in den Ergebnisdarstellungen lediglich unter der Bezeichnung  $x_j$  auftreten. Dabei bedeutet der Index  $j$  die Nummer der in den Tabellen 15 bis 18 verzeichneten Variablen. Für jeden der beiden Tiefgänge stellten sich zwei Variablenkombinationen in bestimmten Froudezahlbereichen als die besten der untersuchten Kombinationen heraus. Jede Kombination bestand, bedingt durch die Computergegebenheiten, aus 55 unabhängigen Variablen.

Bei dem Entwurfstiefgang war die Regressionsnäherung der gegebenen  $C_R$ -Verteilungen im Bereich der Froudezahlen  $< 0,20$  durchweg mäßig, gemessen am Betrag der erklärten Varianz. Für die Froudezahlen 0,17 und 0,18 erzeugte die in Tabelle 16 angegebene Variablenkombination eine merklich bessere Varianzerklärung als die Kombination, die sich in den darunter und darüber liegenden Froudezahlbereichen als überlegen erwies (Tabelle 15). Da die Kleinheit des abweichenden Intervalls Ausnahmecharakter trägt, erschien es nicht zweckmäßig, die Kontinuität der beiden angrenzenden Intervalle zu durchbrechen. Für die Froudezahlen 0,17 und 0,18 werden daher zwei Ergebnisse mitgeteilt, wobei die Ergebnisse der Variablenkombination nach Tabelle 16 erst nach der letzten Froudezahl der durchlaufenden Variablenkombination nach Tabelle 15 erscheinen. Es empfiehlt sich, im Anwendungsfall für  $F_n = 0,17$  und 0,18 beide Ergebnisse zu berechnen und das sich dem vermittelnden Kurvenzug besser anpassende zu verwenden.

Bei dem Ballasttiefgang ergab sich eine klare Trennung in zwei Variablenkombinationen, von denen die in Tabelle 17 abgedruckte für die Froudezahlen 0,12 bis 0,22, die in Tabelle 18 angegebene für die Froudezahlen 0,23 bis 0,33 die besten Varianzerklärungen lieferte.

## 6.2. Darstellung der Ergebnisse

Die Ergebnisse der Regressionsanalysen sind in den Tabellen 19.1.1. bis 19.1.22. für den Entwurfstiefgang ( $B/T < 3,7$ ) und 20.1.1. bis 20.1.22. für den Ballasttiefgang ( $B/T > 3,7$ ) wiedergegeben. Die Tabellen enthalten folgende Angaben: 1) Korrelationskoeffizienten zwischen den unabhängigen Variablen  $x_j$  und der abhängigen Variablen  $y$ , Mittelwerte, Minimal- und Maximalwerte sämtlicher Variablen, die den endgültigen Kombinationen der beiden Tiefgangsgruppen angehören, 2) Regressionskoeffizienten  $b_i$  mit Standardabweichungen und Vertrauensbereichen derjenigen unabhängigen Variablen, die in die jeweiligen Regressionsgleichungen aufgenommen wurden, sowie die Konstante  $b_0$ , 3) die  $t^*$  Werte der Signifikanztests zur Beurteilung der Rangfolge der Regressionskoeffizienten und damit der unabhängigen Variablen, und 4) die Standardabweichungen der gegebenen  $y$ -Werte (hier:  $y = C_R$ ) sowie ihrer Näherungen durch die Regressionsgleichung ( $y$ -berechnet). Letztere Standardabweichung entspricht  $s_{y;x_1 \dots x_p}$  in den Gln. (30a) und (30b), sie ist die Standardabweichung des Fehlergliedes.

Am Fuße der Tabellen finden sich weitere Angaben zur statistischen Beurteilung des Regressionsergebnisses. Der multiple Korrelationskoeffizient bzw. sein Quadrat sind ein Maß für die Güte der näherungsweise Berechnung der gegebenen  $C_R$ -Werte mit Hilfe der Regressionsgleichung. Die Varianz der gegebenen Stichproben sollte möglichst vollständig erklärt sein. Dann ist auch die größte Gewähr dafür gegeben, daß für einen Schiffsentwurf, dessen Formvariable innerhalb der Grenzen der Stichprobenvariablen liegen, gut zutreffende Restwiderstandsbeiwerte berechnet werden.

Anhand des Tests auf Normalverteilung des Fehlergliedes kann die Verlässlichkeit des Tests auf Signifikanz der Regressionskoeffizienten geprüft werden. Der  $\alpha$ -Wert wurde mit dem  $\chi^2$ -Test unter Verwendung von zehn Klassen aus gleichgroßen Flächen unter der Normalverteilungskurve berechnet. Ist  $\alpha$  klein, so gibt der Signifikanztest der Regressionskoeffizienten ein nur qualitativ, nicht jedoch quantitativ richtiges Bild von der Rangfolge der unabhängigen Variablen untereinander in ihrer Bedeutung bei der Beschreibung der abhängigen Variablen  $C_R$ .

Durch Vergleich mit den an letzter Stelle abgedruckten, tabellierten  $t$ -Werten läßt sich feststellen, daß kein Regressionswert auf der 95%-Wahrscheinlich-

keintsebene unsignifikant ist (hierfür müßte  $|t^*| < t(95\%)$  sein), im Gegenteil sind sehr viele sogar auf der 99%-Wahrscheinlichkeitsebene signifikant ( $|t^*| > t(99\%)$ ).

Es ist möglich, daß sowohl in den Werten  $\alpha$  aus dem  $\chi^2$ -Test als auch in den Werten  $t(95\%)$  und  $t(99\%)$  kleine Abweichungen von den tabellierten Werten  $\chi^2_{7,\alpha}$  bzw.  $t_{\nu,95\%}$ ,  $t_{\nu,99\%}$  auftreten, da zu ihrer Berechnung nur Näherungen mit Hilfe numerischer Interpolationsformeln verwendet wurden. Der Freiheitsgrad  $\nu$  berechnet sich aus  $\nu = n - p - 1$ , wo  $n$  = Stichprobengröße und  $p$  = Anzahl der in die Regressionsgleichung aufgenommenen unabhängigen Variablen ist.

In den Tabellen 19.2.1. bis 19.2.22. für den Entwurfstiefgang und 20.2.1. bis 20.2.22 für den Ballasttiefgang sind die vollständigen Regressionsgleichungen und die Koeffizienten zur Berechnung der Vertrauensgrenzen nach den Gl.n. (30a) oder (30b), nach aufsteigenden Froudezahlen geordnet, zusammengestellt. Hiermit können die Restwiderstandsbeiwerte  $C_R$  für den Modellzustand eines neuen Entwurfs und seine Vertrauensgrenzen  $\pm V$  berechnet werden, wobei anstelle der Werte  $x_k$ ,  $x_l$  in den Gl.n. (30a) und (30b) die Werte  $X_k$ ,  $X_l$  für die unabhängigen Variablen zu verwenden sind, die der neue Entwurf besitzt. Für die Mittelwerte  $\bar{x}_k$ ,  $\bar{x}_l$  sind die in den Tabellen der Koeffizienten für die Vertrauensgrenzen angegebenen Werte einzusetzen. Der Gesamtwiderstandsbeiwert  $C_T$  ergibt sich aus

$$C_T = C_{F0} + (C_R \pm V), \quad (56)$$

worin  $C_{F0}$  nach Gl. (55) zu verwenden ist (ITTC 1957).

Über die statistischen Untersuchungen an der abhängigen Variablen  $C_R$  informieren die Tabellen 19.3.1. bis 19.3.20. für den Entwurfstiefgang und die Tabellen 20.3.1. bis 20.3.22. für den Ballasttiefgang. Am Fuß dieser Blätter sind Diagramme der jeweiligen  $C_R$ -Verteilung und der Summenhäufigkeit abgedruckt. Eine tabellarische Übersicht (die Diagrammform hätte zu viel Platz beansprucht) über die Verteilung der in den endgültigen Regressionsanalysen verwendeten unabhängigen Variablen vermitteln die Tabellen 19.4.1. bis 19.4.22. für den Entwurfstiefgang und die Tabellen 20.4.1. bis 20.4.22. für den Ballasttiefgang. Die Aussage dieser Tabellen wird mit Hilfe der Statistik-Abschnitte dieses Berichtes verständlich werden. Sie zeigen

unter anderem die Häufigkeit jeder Variablen innerhalb ihres in zehn Klassen eingeteilten Wertevorrates. Die Randwerte des Wertebereichs können aus den Tabellen 19.1. bzw. 20.1. entnommen werden.

Im Anwendungsfall kann sich der Gebrauch dieser Tabellen auf folgendes Minimum beschränken: Die für die Regressionsgleichungen des gewünschten Froudezahlbereichs benötigten unabhängigen Variablen  $x_j$  werden zusammengestellt, und es werden nur die zugehörigen Verteilungen in den Tabellen 19.4. bzw. 20.4. bei den gewünschten Froudezahlen betrachtet. In erster Näherung genügt hierzu ein Blick auf die Werte für Schiefe und Überhöhung. Sind diese Werte kleiner als die im Kopf der Tabelle abgedruckten Standardabweichungen für Schiefe und Überhöhung, so gleicht die Verteilung der Variablen in den Regressionsstichproben einigermaßen der Normalverteilung, und man kann einen normalen, nicht übergewichtigen Einfluß dieser Variablen auch bei dem Anwendungsfall erwarten, wenn die Variablenwerte des Anwendungsfalls nicht etwa die Stichprobengrenzen überschreiten. Sind aber Schiefe und Überhöhung größer als ihre Standardabweichungen, so war die Stichprobenverteilung mit Schwerpunkten versehen, die das Ergebnis der Regressionsgleichung vorbelasten. In diesem Fall empfiehlt es sich, die Verteilung der betreffenden Variablen aufzuzeichnen und zu prüfen, ob der Variablenwert des Anwendungsfalls in den Schwerpunktsbereich fällt oder nicht. Ist ersteres der Fall, dann wird das Ergebnis der Regressionsgleichung zweifellos genau im Rahmen der Regressionsnäherung sein, im zweiten Fall wird es, beeinflusst durch die fragliche Variable, weniger genau sein. Diese Tatsache wird sich auch in der Größe der Vertrauensgrenzen widerspiegeln, nur ist es mangels Vergleichs schwer zu beurteilen, ob man die Vertrauensgrenzen als weit oder eng ansehen muß.

Unabhängig davon, ob die Hauptabmessungsvariablen  $L/B$ ,  $B/T$  und  $C_B$  in den benötigten Regressionsgleichungen vorkommen oder nicht, empfiehlt sich die Anlage einer graphischen Darstellung der Stichprobenverteilung dieser Variablen mit Hilfe der Tabellen 19.4. bzw. 20.4. Durch Einordnen des Anwendungsfalls in diese Verteilungen gewinnt man ein erstes, sehr wichtiges Kriterium für die Beurteilung seiner Lage innerhalb der Regressionsstichproben. Randlagen lassen Ergebnisse mit großen Vertrauensgrenzen erwarten, das heißt vereinfacht relativ ungenaue Ergebnisse.

### 6.3. Diskussion der Ergebnisse

Eine eingehende Betrachtung der Tabellen 19.2. bis 19.4. bzw. 20.2. bis 20.4. an dieser Stelle würde zu weit führen, sie bleibt dem gewissenhaften Anwender überlassen. Es sollen lediglich die eigentlichen Regressionsergebnisse (Tabellen 19.1. und 20.1.) diskutiert und bisher nicht gegebene Hinweise nachgetragen werden.

Bei dem Entwurfstiefgang liegt der Betrag der durch die Regressionsgleichung erklärten Varianz bis zu der Froudezahl  $F_n = 0,21$  unterhalb 85 % der Stichprobenvarianz, jenseits  $F_n = 0,21$  erreicht er überwiegend mehr als 90 %. Ähnliche Ausprägungen waren für alle untersuchten Variablenkombinationen typisch. Sicherlich hängen sie mit einem Wechsel der Schiffstypenverteilung zusammen. Unterhalb der genannten Froudezahlschwelle fanden sich in der Mehrzahl völlige Fracht- und Tankschiffe, oberhalb vornehmlich Container- und Passagierschiffe. Aber dies allein begründet nicht die "bessere Analysierbarkeit" bei höheren Froudezahlen, und auch der durchweg größte Stichprobenumfang bei höheren Froudezahlen ist kein ausreichend triftiger Grund, denn bei Froudezahlen  $>0,28$  nimmt dieser bei guter Varianzerklärung wieder stark ab. Eine der möglichen Erklärungen ist natürlich, daß bei den kleineren Froudezahlen die bestgeeignete Variablenkombination nicht gefunden wurde, oder auch, daß die bereitgestellten unabhängigen Formvariablen sich nicht bewährten. Eine andere Hypothese verfolgt einen ganz anderen Aspekt und lautet, daß die Meßungenauigkeit des Widerstandes bis hinauf zu einer Froudezahl von  $F_n = 0,20$  relativ groß sein dürfte. Wenn das so ist, enthalten die  $C_R$ -Werte Einflüsse, die mit Formvariablen allein nicht zu analysieren sind. Für diese Hypothese spricht ein Regressionszwischenergebnis, bei dem die probenhalber als Variable aufgenommene Modelllänge sich bei kleinen Froudezahlen fast stets, bei großen Froudezahlen nur selten als signifikant erwies.

Möglicherweise ist das völlig unzulängliche Regressionsergebnis für die Froudezahl  $F_n = 0,12$  auf eben die vermuteten Meßungenauigkeiten zurückzuführen. Keine der untersuchten Variablenkombinationen steigerte die Varianzerklärung auf mehr als 45 % (abgedruckt: 41,2 %). Daß das Ergebnis für  $F_n = 0,12$ , das nicht benutzt werden sollte, hier dennoch mitgeteilt wird,

hat im wesentlichen drei Gründe. Erstens liegt der Stichprobenumfang mit 56  $C_R$ -Werten weit oberhalb der gesteckten Mindestzahl von 40. Es wäre unehrlich, einen unter diesen Umständen erzielten Mißerfolg zu verschweigen. Zweitens scheint das Ergebnis dahingehend interpretiert werden zu können, daß die Meßgenauigkeit bei sehr kleinen Froudezahlen (gemeint sind  $F_n = 0,12$  und kleiner) außerordentlich hoch sein muß. Drittens fällt das Ergebnis soweit aus dem Rahmen der übrigen, daß vielleicht mehr dahinter steht als das, was zu seiner Deutung bisher gesagt wurde, die richtige Erklärung aber erst noch gefunden werden muß.

Bei dem Ballasttiefgang ist die Varianzerklärung schon bei kleinen Froudezahlen hoch und unterschreitet im gesamten Froudezahlenintervall nicht 82 %. Die "gute Analysierbarkeit" bei kleinen Froudezahlen kann unter Bezug auf das vorstehend Gesagte erklärt werden mit der gegenüber dem Entwurfstiefgang schlankeren Unterwasserform in Ballast, an der die Strömung weniger leicht zu Instabilitäten neigt, z.B. Abreißen im Hinterschiff, so daß höhere Meßgenauigkeit zu erreichen ist. Ein Blick auf die Korrelationskoeffizienten, d.h. die Meßzahl für die Größe einer linearen Beziehung zwischen dem Restwiderstandsbeiwert  $C_R$  und den unabhängigen Variablen, zeigt, daß einige von ihnen im Unterschied zum Entwurfstiefgang bereits bei kleinen Froudezahlen höhere Werte annehmen. Dies kann als ein Zeichen für die von vornherein in den  $C_R$ -Werten angelegte Eigenschaft angesehen werden, den verwendeten Formvariablen zuordnungsfähig zu sein. Es spricht für fast reine Formabhängigkeit und Abwesenheit formfremder Einflüsse.

Es ist schwer zu sagen, welche Größe die Varianzerklärung mindestens haben sollte, um sichere Anwendbarkeit der Regressionsgleichungen zu gewährleisten. Die Statistik macht eine solche Angabe nicht. Das einzelne Ergebnis ist durch seinen aus der Regressionsgleichung berechneten Mittelwert und seine Vertrauensgrenzen statistisch völlig ausreichend charakterisiert. Die Vertrauensgrenzen wachsen mit geringer werdendem Betrag der Varianzerklärung, was darauf beruht, daß beide Größen unter Verwendung der gleichen Standardabweichungen  $s_{y;x_1 \dots x_p}$  berechnet werden. Für den Anwender ist allerdings das Ergebnis mit den größeren Vertrauensgrenzen weniger angenehm, aber er verfügt ja immerhin über ein Maß für die Größe des Vertrauens, das er in die Richtigkeit seiner Berechnung setzen darf. Nach Meinung des Verfassers lassen Regressionsgleichungen

mit etwa 90 % erklärter Varianz sehr ordentliche Ergebnisse für Schiffe erwarten, deren unabhängige Variablen innerhalb der Stichprobengrenzen liegen und möglichst auch deren eventuell vorhandenen Schwerpunktbereichen angehören. Bei Froudezahlen mit niedrigeren Varianzerklärungen dürfte man sich wachsenden Schwierigkeiten gegenübersehen, eine strakend verlaufende Kurve durch die berechneten Punkte zu legen.

Die Bilder 24 bis 26 zeigen Anwendungen der Regressionsergebnisse auf drei Schiffen auf Entwurfstiefgang, die zum Kollektiv der Datenschiffe gehören. Sie wurden zufällig gezogen mit der Einschränkung, daß eins von ihnen den unteren und eins den oberen Froudezahlbereich überdecken sollte, ein drittes den gesamten Bereich. Zur Abschätzung des Fehlers wurden die Kurven des Rest- und Gesamtwiderstandes mit eingetragen, und zusätzlich die Plattenreibung nach der ITTC-Linie von 1957. Die beiden Vertrauensbereiche sind unterschiedlich schraffiert eingezeichnet. Man erhält einen Eindruck von der Streuung oder dem "Springen" der Regressionspunkte, und aufgrund der gemessenen Kurven einen Anhalt dafür, mit welchem glatten Verlauf die Interpolationskurve hindurchzulegen ist. Die größten auf den Gesamtwiderstand bezogenen Abweichungen der berechneten  $C_R$ -Werte betragen in den drei Fällen 3,6; 11,7 und 5,2 %. Bei dem Schiff von Bild 25 nimmt die gemessene Widerstandskurve einen bei den hohen gemessenen Froudezahlen außergewöhnlich flachen Verlauf, den die Regressionskurve fehlerhaft wiedergibt.

Am Beispiel der Tabelle 19.1.11 (Entwurfstiefgang,  $F_n = 0,22$ ) sei noch einmal zu einer Betrachtung der Korrelationskoeffizienten zurückgekehrt. Bei der Wanderung durch die betreffende Spalte fällt auf, daß eine Reihe von Variablen mit hohen Korrelationskoeffizienten nahe  $\pm 1$  nicht in die Regressionsgleichung aufgenommen wurde, hingegen mit  $x_{54}$  und  $x_{50}$  Variable mit außerordentlich kleinen Korrelationskoeffizienten nahe Null.

Dem Regressionsergebnis selbst ist keine vertretbare Erklärung für diese verwunderliche Erscheinung zu entnehmen. Die Vermutung ist richtig, daß die Regressionsanalyse miteinander stark korrelierende Variable gefunden und dies berücksichtigt haben muß. Aber welche Variablen korrelieren denn stark miteinander, oder welche Einzelvariable korreliert denn mit welcher Gruppe von anderen Variablen? Es ist nicht ersichtlich. Man könnte sich

natürlich die Korrelationsmatrix der 55 Variablen ansehen, wird dies aber wegen der Fülle von Zahlen nach kurzer Zeit aufgeben müssen. Die optische Analyse der Korrelationsmatrix ist aussichtslos, sie gibt ja auch keinen Aufschluß über Gruppenkorrelationen. Angenommen, sie hätte dennoch Erfolg, dann würde man unter Benutzung der naheliegenden Voraussetzung, nur noch hoch mit der Zielgröße korrelierende unabhängige Variable in die engere Wahl ziehen zu sollen, ganz offenbar Schiffbruch erleiden.

Dies ist das Problem der anscheinend unbegrenzten Zahl von notwendigerweise auszuprobierenden Variablen und Variablenkombinationen in der Regressionsanalyse. Es wurde hier mit Hilfe der Faktorenanalyse methodisch zu lösen versucht, und der Lösungssatz soll an oben gewähltem Beispiel überprüft werden.

Der leitende Gesichtspunkt, unter dem die Faktorenanalyse eingesetzt wurde, ist es, möglichst viele Variablen mit differierendem, voneinander unabhängigem Informationswert zu finden, weil die Regressionsanalyse gerade solche "unabhängigen" Variablen braucht. Hierbei kommt es zunächst gar nicht darauf an, daß die untersuchten Variablen viel mit der Zielgröße der Regressionsanalyse zu tun haben, sie sollen in erster Linie nur unabhängig voneinander sein (man wird sie aus anderen Überlegungen trotzdem so wählen, daß der Zusammenhang mit der Zielgröße zumindest mutmaßlich ist). Die Faktorenanalyse vermittelt nur eben jenes Wissen über die wahre Unabhängigkeit der "unabhängigen" Variablen, das die Regressionsanalyse zwar implizit verwendet, aber explizit nicht darzustellen vermag. Man kann vor der ersten Regressionsanalyse die Variablenliste solange erweitern, bis man über genügend viele unabhängige Variable zu verfügen glaubt. Wenn das Regressionsergebnis dann doch nicht befriedigt, liegt es mit Bestimmtheit daran, daß Anzahl und Art der unterschiedlichen Informationswerte der Variablen noch nicht ausreichen, und man wird mit der Faktorenanalyse weiter suchen (wenn nicht, was aber durch andere Überlegungen zu klären ist, die Analysierbarkeit der Stichprobe, z.B. aufgrund von unabwendbaren Meßfehlern, überhaupt infrage gestellt ist). Weiterhin läßt sich aber auch die Merkwürdigkeit manches Regressionsergebnisses erst mit Hilfe der Ergebnisse einer Faktorenanalyse verstehen.

Mit der Kenntnis über gleichwertige Informationsgehalte der Variablen von Tabelle 19.1.11., die die Faktorenanalyse der vierten Variablen-Gruppe (Tabelle 11) geliefert hat, soll das als Beispiel gewählte Regressionsergebnis untersucht werden. Die Variablen  $x_7$  und  $x_9$  ( $10^3(C_B^E)^3$  und  $x_{F/L}^E$ ) gehören dem Faktor 1 von Tabelle 11 an, enthalten mithin weitgehend dieselbe Information wie die Variablen 1 bis 13 und einige weitere. Daß zwei Variable dieses Faktors gewählt wurden, deutet einerseits auf die Stärke der Information bezüglich  $C_R$  hin, und besagt zweitens fast trivial, daß die Beschreibung der gegebenen Zielgrößenwerte eben besser mit zwei Angehörigen dieser Information gelingt als mit nur einer.  $x_{48}$  ( $10B/L \cdot (y_{2,3} - y_{1,3})$ ), ein Wasserlinienwinkel im Heck, gehört nicht zu den Variablen der Tabelle 11, ist aber wohl eindeutig eine Information, die die anderen Variablen nicht enthalten.  $x_{50}$  (B/T) bildet den Faktor 7, hat also selbständigen Informationswert.  $x_{54}$  ( $\text{SQRT}(L^B B^B H^B / LBT)$ ) ist eine Wulstvariable, die zwar in Tabelle 11 nicht auftaucht, aber doch wohl dem Faktor 5 zuzuordnen ist. Die Variable ist somit wiederum eine unabhängige Informationsquelle.

Die Regressionsanalyse hat also fast ausschließlich Variable ausgewählt, die verschiedenen Faktoren angehören, ist also genau nach dem Schema der voneinander unabhängigen Informationen vorgegangen. Nun verwundert die Aufnahme von  $x_{50}$  und  $x_{54}$  nicht mehr und auch das Auslassen z.B. der Variablen  $x_{25}$  und  $x_{26}$  ( $y_{2,3}$  und  $y_{18,3}$ ), die dem Faktor 1 angehören, wird erklärlich. Letztere Variablen wurden deshalb nicht benötigt, weil  $x_7$  und  $x_9$  sie bereits repräsentieren. Welche Variable eines Faktors wichtig ist, kann die Faktorenanalyse natürlich nicht vorhersagen. Denn das ist ein Problem der Beschreibung der Zielgröße, wofür die Regressionsanalyse zuständig ist.

Wenn die Variablenkombinationen, die hier für die Regressionsanalysen gewählt wurden, häufig mehrere Angehörige desselben Faktors enthalten, dann deshalb, weil ja nie bekannt ist, welche Variable eines Faktors gebraucht wird. Die Zusammenstellung der Variablen eines Faktors für die endgültige Regression kann z.B. in der Weise erfolgen, daß alle seine in den vorbereitenden Regressionsanalysen jemals, das heißt bei irgendeiner Froudezahl, ausgewählten Variablen berücksichtigt werden. So entstanden die hier wiedergegebenen Variablenkombinationen. Es ist durchaus nicht notwendig, daß jede Variable mindestens einmal bei irgendeiner Froudezahl erneut gewählt werden muß, weil möglicherweise eine besser geeignete Variable sie ersetzt.

## 7. Formfaktor k

### 7.1. Theoretische Ableitung

Nach der von W. FROUDE formulierten Hypothese kann der Schiffswiderstand in einen von der Schwerkraft und einen von der Zähigkeit des Wassers bedingten Anteil aufgespalten werden, und beide sind voneinander unabhängig. Sichtbare Wirkung der Schwerkraft beim bewegten Schiff sind die Wellen, bekannte Wirkung der Zähigkeit ist die Grenzschicht.

Die beiden Widerstandskomponenten folgen verschiedenen Maßstabsgesetzen. Um im Modellversuch Ähnlichkeit mit den Widerstandskräften des Schiffes einzurichten, muß für den schwerkraftbedingten Wellenwiderstand das Froudesche Gesetz und für den Zähigkeitswiderstand das Reynoldssche Gesetz erfüllt sein. Bekanntlich ist gleichzeitige Erfüllung beider Gesetze am Modell nicht möglich, man erfüllt nur das Froudesche Gesetz. Damit sind die Wellenwiderstände von Modell und Schiff einander ähnlich, und der dimensionslose Wellenwiderstandsbeiwert  $C_W = R_W / 0,5\rho V^2 S$  hat für Modell und Schiff die gleiche Größe. Die Bedingung für Ähnlichkeit des Zähigkeitswiderstandes ist aber nicht erfüllt.

Es ist üblich, den Zähigkeitswiderstand in den Reibungswiderstand einer ebenen Platte und den zähigkeitsbedingten Druckwiderstand zu zerlegen. Der Reibungswiderstand der ebenen Platte konnte in Versuchen über einen großen Reynoldszahlbereich bestimmt werden, und in Anlehnung daran wurden mehrere nicht ganz übereinstimmende Korrelationslinien festgelegt. Hieraus können die Reibungsbeiwerte  $C_{FO}$  von Modell bzw. Schiff bei den jeweils geltenden Reynoldszahlen berechnet werden.

Die Übertragung des zähigkeitsbedingten Druckwiderstandes erfolgt seit FROUDE nach einer ingenieurmäßigen Methode. Es war zunächst nicht möglich, die Größe dieses Widerstandsanteils experimentell oder theoretisch zu bestimmen. Wenn man aber annimmt, der Formwiderstand sei unabhängig von der Reynoldszahl, dann kann man ihn dem Wellenwiderstand zuschlagen und erhält

$$C_T = C_{FO} + C_R, \quad (57)$$

wo  $C_R$  die Summe aus den Beiwerten von Wellenwiderstand und Druckwiderstand

und  $C_T$  der Gesamtwiderstandsbeiwert ist. In Gl. (57) ist lediglich  $C_{F0}$  maßstabsabhängig. Unter Berücksichtigung der verschiedenen Reynoldszahlen läßt sich der Schiffswiderstand mit Gl. (57) aus dem Modellwiderstand berechnen. Die Methode liefert praktisch zufriedenstellende Ergebnisse.

Nach einem anderen Ansatz werden Reibungswiderstand und Zähigkeitsbedingter Druckwiderstand zum viskosen Widerstand  $R_V$  zusammengefaßt, in Beiwertschreibweise

$$C_V = (1 + k) C_{F0} \quad (58)$$

Dies führt zu folgendem Ausdruck für den Gesamtwiderstandsbeiwert:

$$C_T = C_V + C_W = (1+k) C_{F0} + C_W, \quad (59)$$

wo  $C_W$  der Beiwert des Wellenwiderstandes ist;  $k$  wird mit Formfaktor des Zähigkeitswiderstandes bezeichnet. Angenommen der Formfaktor ist bekannt, dann zerlegt man den im Modellversuch gemessenen Gesamtwiderstand gemäß Gl.(59) in seine Bestandteile und überträgt  $C_W$  nach dem Froudeschen Gesetz, das heißt, es hat gleiche Größe für Modell und Schiff, und  $C_{F0}$  unter Berücksichtigung der unterschiedlichen Reynoldszahlen. Der Formfaktor wird als maßstabsunabhängig angesehen. Diese Methode ist theoretisch der anderen überlegen. Ihre Anwendung stößt jedoch auf praktische Schwierigkeiten, da die Ermittlung des Formfaktors im Modellversuch ein diffiziles Problem ist. In der Literatur ist außerdem noch die Verwendung der geeigneten Plattenreibungslinie für  $C_{F0}$  umstritten (z.B. SCHOENHERR, HUGHES, ITTC 1957). Jede Reibungslinie führt zu einem anderen Wert von  $k$ . Im folgenden wird die durch Gl.(55) gegebene Korrelationslinie der ITTC 1957 verwendet.

## 7.2. Ermittlung des Formfaktors

Zur Ermittlung des Formfaktors gingen HUGHES [21] und etwa gleichzeitig PROHASKA [22] von wellentheoretischen Überlegungen aus. Ihr Ergebnis bestand im wesentlichen aus einer Formel für den wellenbedingten Widerstand, wie sie in 4.2.1.2. durch Gl. (49) dargestellt ist. Für kleines  $F_n$  ( $F_n < 0.20$ ) kann der Einfluß der sin- und cos-Glieder vernachlässigt werden, und in

$$C_T = (1+k)C_{F0} + yF_n^4 \quad (57)$$

wurde eine einfach auszuwertende Formel für den Gesamtwiderstand gefunden, die den Formfaktor  $k$ , den Plattenreibungswiderstand  $C_{F0}$  und die Froudesche Zahl  $F_n$  enthält. Trägt man Widerstandsmeßpunkte bei kleinen Froudezahlen in einem Diagramm mit der Ordinate  $C_T/C_{F0}$  und der Abzisse  $F_n^4/C_{F0}$  auf, so sollten sie durch eine Gerade verbunden werden können, die die Ordinatenachse bei dem Wert  $(1+k)$  schneidet und die Steigung  $y$  besitzt.

Mittels Auftragungen dieser Art wurde die Eignung von 207 Widerstandsversuchen mit den Datenshippen für die Ermittlung von  $k$  und  $y$  geprüft. In dieser Zahl sind 32 vom SFB 98 in Auftrag gegebene Ergänzungsmessungen bei kleinen Froudezahlen enthalten, die durchgeführt wurden, wenn der Geschwindigkeitsbereich eines industrieseitig vorgesehenen Widerstandsversuchs bis zu einer Froudezahl von etwa  $F_n = 0.15$  hinunterreichte. Der generellen Verwendbarkeit der Messungen standen Hindernisse entgegen: 1) die starke Streuung der Meßpunkte bei Froudezahlen um  $F_n = 0.10$ , verursacht durch Effekte wie Laminarströmung, Ablösung im Hinterschiff und die bei sehr kleinen Gesamtwiderständen relativ wachsende Ungenauigkeit der Meßgeräte, 2) die Neigung zu konkavem, wellenförmigem oder sogar konvexem Verlauf der Verbindungslinie bei manchen Modellformen. Solche Erscheinungen, auf die auch PROHASKA hinweist, sind durch einen stärkeren als mittels  $yF_n^4$  erfaßbaren Wellenwiderstandseinfluß schon bei kleinen Froudezahlen zu erklären. Die Unsicherheit in der Auswertung selbst relativ eindeutiger Meßpunktfolgen wird durch den unvermeidlichen subjektiven Einfluß des Bearbeiters noch erhöht. Ein Beispiel für die Art der Auswertung zeigt Bild 10.

Bei nur 112 der ursprünglich 207 Widerstandsmessungen gelang eine für vertretbar gehaltene Bestimmung der Werte  $k$  und  $y$ . Hiervon gehörten je 65 der Entwurfstiefgangsgruppe und je 47 der Ballasttiefgangsgruppe an. Sämtliche Beiträge stammten von Modellen der Typen Frachter, Tanker und Erz/Oboschiff mit großer Völligkeit.

Unter Benutzung derselben Variablen und Variablenkombinationen wie bei der Regression der Restwiderstandsbeiwerte wurden beide Gruppen zunächst getrennt analysiert. Die besten Ergebnisse, berechnet mit verschiedenen Variablenkombinationen, waren: für  $k$ (Entwurf) 68%,  $k$ (Ballast) 94%,  $y$ (Entwurf) 35%,  $y$ (Ballast) 63% erklärte Varianz. Kein Resultat außer dem für  $k$ (Ballast) konnte zufriedenstellen.

In all diesen Regressionen ließen sich Ballasttiefgangswerte besser analysieren als Entwurfstiefgangswerte. Schon bei der Auswertung der Widerstandsversuche war aufgefallen, daß die nach PROHASKA umgerechneten Meßwerte der Ballasttiefgänge wesentlich besser auf einer Geraden lagen als die Meßwerte der Entwurfstiefgänge. Der mögliche Fehler bei der Bestimmung des  $k$ -Faktors wird für die Ballastfälle somit geringer sein.

Besonders groß sind die Unsicherheiten offenbar bei dem Steigungsmaß  $y$ . Daher erschien es nicht zu verantworten, Interpolationsformeln für  $y$  anzugeben.

Nachträglich können Einflüsse der Versuchstechnik und des Auswerteverfahrens zwar nicht mehr beseitigt werden, aber man kann doch versuchen, wenigstens statistisch das Ergebnis für  $k$ (Entwurf) aufzubessern. Eine Möglichkeit hierzu bestand in der gemeinsamen Analyse der  $k$ -Werte von Entwurfs- und Ballasttiefgang. Es ist dies ein Kompromiß, der leidlichen Erfolg hatte. Die zwar fehlerbehafteten, aber doch nach bestem Wissen ermittelten Formfaktoren des Entwurfstiefganges werden durch die gemeinsame Regressionsgleichung im Mittel besser angenähert als in der Einzelanalyse, für die Formfaktoren des Ballasttiefganges gilt jedoch das Gegenteil. Die Ergebnisse dieser Berechnung werden hier mitgeteilt.

### 7.3. Darstellung und Diskussion der Ergebnisse

Tabelle 21 enthält Numerierung und Bezeichnung der unabhängigen Variablen in der endgültigen Formfaktorenanalyse. Zusätzlich zu den schon bekannten Variablen wurden unter Nummer 46 bis 48 drei Formelausdrücke aufgenommen, die in der Literatur als Interpolatoren für den Formfaktor angegeben werden (Tabelle 26). Gegenüber weiteren derartigen Formeln [23] wurden sie in vorbereitenden Regressionen mehrfach ausgewählt.

Die Ergebnisdarstellung entspricht der Darstellung für den Restwiderstandsbeiwert. Tabelle 22 enthält die Ergebnisübersicht, Tabelle 23 die explizit geschriebene Regressionsgleichung und die Koeffiziententabelle zur Berechnung der Vertrauensgrenzen. Aus Tabelle 24 ist das Resultat der statistischen Untersuchung der gegebenen Formfaktorwerte zu entnehmen, Tabelle 25 zeigt die Verteilung und Verteilungskennwerte der unabhängigen Variablen. Der Gebrauch der Tabellen kann sinngemäß nach der im Zusammenhang mit der  $C_R$ -Analyse gegebenen Beschreibung erfolgen.

Die erklärte Varianz der gegebenen k-Faktorwerte beläuft sich auf nicht ganz befriedigende 80% (Tabelle 21). Aber immerhin liefert der Test auf Normalverteilung mit  $\alpha = 5\%$  einen relativ hohen Wert, und die Aussagen des Signifikanztestes für die Regressionskoeffizienten (t\*-Test) lassen die Bedeutung der unabhängigen Variablen für die interpolatorische Darstellung der Formfaktoren ziemlich sicher beurteilen. Danach sind Winkel im Hinterschiff (Wasserlinienwinkel  $x_{24}$  und Spantwinkel  $x_{10}$ ) besonders wichtig, aber auch der mehrere Variable zusammenfassende Formelausdruck  $x_{48}$  (ITTC-Formel, Tabelle 26),  $C_M$  ( $x_{17}$ ) und das Heckspantaufmaß  $x_{50}$  treten hervor. Die übrigen ausgewählten Variablen sind zwei Wulstparameter und ein weiteres Heckspantaufmaß. Die Betonung liegt also auf Hinterschiffs- und Wulstkennwerten. Beides kann hydrodynamisch betrachtet nicht überraschen.

Der Vergleich mit den Ergebnissen der Faktorenanalyse zeigt, daß die in der Formfaktorregression ausgewählten Variablen weitgehend voneinander unabhängig sind.

Die Verteilung der gegebenen k-Werte weicht stark von der Normalverteilung ab (Tabelle 24), und Zufälligkeit der Stichprobe ist nicht gewährleistet. Tabelle 25 zeigt denn auch, daß viele Verteilungen der gewählten, mit Schwerpunkten behaftet sind. Bei der Anwendung der Regressionsgleichung ist demzufolge große Vorsicht geboten. Man wird nur für Schiffe, deren Variablenwerte nahe den in der Stichprobe auftretenden Schwerpunkten liegen, einigermaßen verlässliche k-Faktoren interpolieren können. Eine Anwendung der Regressionsgleichung auf alle Datenschiffe zeigt, daß für nicht zu den analysierten Schiffstypen gehörende Schiffe sogar negative Formfaktoren berechnet werden.

Die Analyse macht deutlich, daß große Schwierigkeiten bei der Ermittlung des k-Faktors vor allem bei auf vollem Tiefgang geschleppten Modellen auftreten. Dies ist insbesondere auf die überwiegende Verwendung von Daten aus Routinemessungen zurückzuführen. Solche Messungen orientieren sich natürlich an ökonomischen Gesichtspunkten und nicht an den Erfordernissen der Formfaktorbestimmung. Diese laufen generell auf eine dichtere Meßpunktfolge und speziell auf die eingehende Untersuchung der Umgebung von auffälligen Meßpunkten hinaus, aus der gegebenenfalls der Einfluß von Strömungsinstabilitäten abgelesen werden kann. Das verfügbare Datenmaterial war in dieser Hinsicht naturgemäß unzulänglich, und es ist daher tatsächlich eine schwer zu entscheidende Frage, ob sich die Verwendung der hier aufgestellten Regressionsformel zum Zweck der Übertragung von Modellwiderständen auf das Schiff empfiehlt. Einzig die Zwangslage, daß viel besseres nicht in Sicht ist und auch andere Versuchsanstalten offenbar vor denselben Schwierigkeiten stehen, könnte einen solchen Schritt rechtfertigen.

## 8. Zusammenfassung

In jüngster Vergangenheit waren, um einem fast schlagartig einsetzenden kommerziellen Interesse Genüge zu leisten, neuartige oder doch aus dem Rahmen des gewohnten fallende Schiffsformen zu entwickeln. In erster Linie sind hier die völligen, langsam fahrenden Schiffe für trockene und flüssige Ladung sowie die schnellen Containerschiffe zu nennen. Der Reife-prozeß der Unterwasserformen fand vornehmlich in den Versuchsanstalten statt, da empirische Entwurfshilfsmittel entweder gar nicht zur Verfügung standen oder vorhandene sich nicht mit der notwendigen Zuverlässigkeit extrapolieren ließen. Mittlerweile jedoch ist der angesammelte Vorrat an Modellmeßdaten groß genug, so daß an die Ergänzung empirischer Entwurfsunterlagen gedacht werden kann.

Die vorliegende Arbeit will die auf dem wichtigen Gebiet des Schiffswiderstandes entstandene Lücke in den empirischen Entwurfsunterlagen schließen. Sie stützt sich auf Modellmeßdaten der Hamburgischen Schiffbau-Versuchsanstalt aus den Jahren 1967 bis 1973. Das gewählte Mittel der Datenerschließung ist die Regressionsanalyse. Mit ihrer Hilfe berechnete Interpolationsformeln für den Restwiderstandsbeiwert  $C_R$  an festen Froudezahlen in den Bereichen  $0,13 \leq F_n \leq 0,31$  bzw.  $0,12 \leq F_n \leq 0,33$  für Modelle in voll beladenem bzw. Ballastzustand bilden den Kern der Ergebnisse. Parameter (unabhängige Variable) der Formeln sind ausschließlich Kenngrößen der Schiffsform sowie die Schraubenzahl. Der Gesamtwiderstandsbeiwert läßt sich aus den Gleichungen mittels Addition des Beiwertes des Plattenreibungswiderstandes nach der ITTC 1957-Korrelationslinie errechnen:  $C_T = C_{F0} + C_R$ . Die Ergebnisgenauigkeit wird durch Angabe von Vertrauensgrenzen für das Interpolationsergebnis überwacht.

In den bisher veröffentlichten Arbeiten über Regressionsanalysen des Widerstandes ist ein Mangel an gründlicher Diskussion der zu betrachtenden statistischen Techniken festzustellen. Desgleichen sind kaum Ansätze zu einer methodischen Auswahl der Formvariablen vorhanden, mit denen der Widerstand beschrieben werden soll. Untersuchungen über die bestgeeignete Funktion, in der eine Formvariable mit dem Widerstand korreliert, fehlen ebenfalls.

Diese drei Aspekte bilden Schwerpunkte der vorliegenden Arbeit. Der erste Punkt wurde ausführlich unter Benutzung der umfangreich vorhandenen Literatur behandelt. Für den zweiten wurde durch Anwendung der Faktorenanalyse ein Ansatz gefunden, der sich als sehr tragfähig erwies. Für den dritten Punkt wurde eine Lösung auf graphischem Wege angestrebt, hier jedoch ist eine elegantere, den Computer benutzende Methode denkbar.

Frühzeitig ließen die durchgeführten Faktorenanalysen erkennen, daß lediglich aus den Hauptabmessungen gebildete Parameter nur sehr wenig differenzierte Information über die Schiffsform enthalten. Sie können die beobachtete große Streuung der Restwiderstandsbeiwerte nicht ausreichend erklären. Aus diesem Grunde wurde die Zahl der unabhängigen Variablen erheblich über das in vergleichbaren Arbeiten übliche Maß hinaus gesteigert. Hinzugefügt wurden Kenngrößen wie Längen- und Höhenschwerpunkt der Verdrängung, Wasserlinienswerpunkt und -völligkeit des gesamten Schiffes und von charakteristischen Schiffsabschnitten wie Heck, Bug und Wulst. Ferner wurden Aufmaße und Konturwinkel von je zwei Spanten in Vor- und Hinterschiff sowie geometrisch einfache, aus den Spantaufmaßen berechnete Wasserlinienwinkel berücksichtigt.

Die Faktorenanalysen zeigten, daß in all diesen Formvariablen (rund 50) nur dreizehn voneinander unabhängigen Grundinformationen enthalten sind. Ausgenommen sind hierbei die Wasserlinienwinkel und die erst im Laufe der Untersuchung hinzugefügte Schraubenzahl, die weitere unabhängige Informationen darstellen. Zahlreiche Formvariable bedingen sich demnach gegenseitig, das heißt mehrere von ihnen besagen in jeweils nur anderer, uns zugänglicher Form dasselbe. Im Prinzip wies die Faktorenanalyse nach, daß eine Grundinformation oft von einem ganzen Bereich des Schiffskörpers geprägt wird. Z.B. enthalten Längenschwerpunkt der Verdrängung, Wasserlinienswerpunkt und Wasserlinienvölligkeit weitgehend identische Forminformation über das Gesamtschiff, oder das Hinterschiff, je nachdem von welchem Schiffsteil sie stammen.

Die Faktorenanalyse läßt völlig offen, ob eine Grundinformation zum Zweck der Widerstandsbeschreibung nützlich oder unbrauchbar ist, weil sie lediglich Korrelationen unter den unabhängigen Variablen untersucht. Aber sie ordnet die vorgesehenen Variablen nach Gemeinsamkeiten, sagt etwas aus über die Menge der wirklich vorhandenen unabhängigen Forminformationen und

erleichtert wesentlich das Aufstellen erfolgversprechender Variablenkombinationen für die Regressionsanalyse. Mit den von den Faktorenanalysen vermittelten Kenntnissen konnte die fast unendlich große Zahl der denkbaren Variablenkombinationen sehr rasch auf ein erträgliches Maß reduziert werden.

Die Regressionsanalyse wählt automatisch voneinander unabhängige Informationswerte zur Aufnahme in die Regressionsgleichung aus, und solche sollten ihr von Anfang an in ausreichender Anzahl vorgegeben werden. Von mehreren, dieselbe Grundinformation vertretenden Variablen verwendet die Regressionsanalyse immer diejenige, die den größten Nutzen zur Erzielung der kleinsten Fehlerquadratsumme mit sich bringt.

Beträchtlichen Zeitaufwand erforderte die zur Berechnung der zusätzlichen Formvariablen benötigte Aufmessung der Modelle. Zweifellos wurde hierdurch die Auswertung einer noch größeren Zahl von Versuchsergebnissen verhindert, doch wäre der Verzicht auf Aufmessung einem nur mäßigen Analysenergebnis gleichgekommen. Insgesamt konnten 554 Versuche mit 308 Modellen in Betracht gezogen werden, von denen nach dem Ausscheiden von vom Nutzen her bezweifelten Messungen 250 Versuche auf voll beladenem Tiefgang und 175 Versuche auf Ballasttiefgang die endgültigen Kollektive bildeten. Bei den einzelnen Froudezahlen ergaben sich hiermit Stichproben von einer Größe zwischen 40 (untere festgesetzte Grenze) und 128 analysierten  $C_R$ -Werten. Statistisch gesehen sind dies kleine Stichproben. Diese Tatsache hat unmittelbare Auswirkung auf die Größe der Vertrauensbereiche.

Die auf graphischem Wege betriebene Suche nach der bestgeeigneten Funktion für den Zusammenhang zwischen einer Formvariablen und dem Restwiderstandsbeiwert beschränkte sich auf die Verwendung des einfachen Potenzansatzes. Das erstaunlichste Ergebnis war, daß dem gesamten Komplex in bisher erschienenen Arbeiten über Widerstandsregressionen offenbar keinerlei Beachtung geschenkt worden ist. Anders ist schwer zu erklären, warum ein so vertrauter und übersichtlicher Parameter wie  $L/B$  gewöhnlich in den Funktionen  $(L/B)$ ,  $(L/B)^2$  oder  $(L/B)^3$  auftaucht anstelle der doch eher einleuchtenden Form  $1/(L/B)$  oder  $1/(L/B)^2$ . Mehrere ähnliche Fälle wurden gefunden.

Am häufigsten traten folgende Variablen, hier ohne Berücksichtigung der funktionalen Form genannt, in den Regressionsgleichungen auf: Als Kennwerte des vollbeladenen Zustandes Blockkoeffizient, Wasserlinienschwerpunkt und -völligkeit des 20% L langen Vorschiffs, der Längenschwerpunkt des ebenso langen Hinterschiffes, Wasserlinienwinkel im Heck und eine aus Länge, Breite und Höhe gebildete Kenngröße des Wulstes; als Kennwerte des Ballasttiefganges der Längenschwerpunkt des Gesamtschiffes, die Größen  $\nabla/L^2T$  sowie  $(L/B) \cdot \nabla/L^3$ , mehrere Spantaufmaße und Spantkonturwinkel in Vor- und Hinterschiff, Wasserlinienwinkel im Hinterschiff sowie der Blockkoeffizient des Wulstes und seine mit Länge, Breite und Höhe gebildete Kenngröße.

Aus dieser Übersicht geht hervor, daß der Restwiderstandsbeiwert sehr stark von Formdetails abhängt. Konturwinkel z.B. erwiesen sich in den Regressionen als äußerst wirkungsvolle Hilfsmittel zur Beschreibung lokaler Formen.

Fast alle genannten Formvariablen enthalten voneinander unabhängige Grundinformationen. Nur bisweilen, z.B. bei den Wulstkenngößen, traten mehrere weitgehend gleichbedeutende Variable zugleich in den Regressionsgleichungen auf. Das war offenbar zur Erfassung einer besonders starken Abhängigkeit zwischen Restwiderstandsbeiwert und Formgrundinformation notwendig.

Nicht alle Grundinformationen waren für die Beschreibung des Widerstandsbeiwertes von gleichem Nutzen. Doch sie wurden sämtlichst wenigstens einmal bei irgendeiner Froudezahl in den Regressionen angefordert.

Der Erfolg der Analyse wurde durch die herkunftsbedingt unsystematische, in manchen Froudezahlbereichen durch kleine Modellserien auch schwerpunktbehaftete Zusammensetzung des Gesamtdatenkollektivs beeinträchtigt, andererseits durch die genannten Begleituntersuchungen wirkungsvoll gefördert. Die Güte der Annäherung der gegebenen  $C_R$ -Werte durch die Regressionsgleichungen wechselte von Froudezahlbereich zu Froudezahlbereich. Restwiderstandsbeiwerte des Ballastzustandes ließen sich, erklärlich aus der weniger störanfälligen Umströmung der vergleichsweise schlankeren Schiffsform, besser analysieren als solche des beladenen Zustandes. Die schlechtesten Ergebnisse wurden bei Froudezahlen kleiner als 0,22 für den Entwurfstiefgang erhalten. Zum Teil ist dies auf modelltechnisch bedingte Einflüsse zurückzuführen, z.B. auf den von der Modellgröße her notwendigen Verzicht auf ausreichend hohe

Reynoldszahlen. Ein weiterer Grund liegt aber auch darin, daß nur eine begrenzte Anzahl von Formvariablen zur Verfügung stand, deren widerstandsbeschreibende Funktion nicht unbedingt die bestmögliche gewesen sein muß.

Als Maß für die Güte einer Interpolationsformel wurde der multiple Korrelationskoeffizient herangezogen, der die Streuung des Fehlergliedes der Regressionsgleichung und die Streuung der gegebenen Stichprobenwerte zueinander ins Verhältnis setzt. Es ist das Ziel der Regressionsanalyse, das Fehlerglied zu minimieren, und je besser dies gelingt, desto mehr nähert sich der Korrelationskoeffizient dem größtmöglichen Wert 1. In diesem Fall wird die Varianz (Streuung) der gegebenen Stichprobenwerte durch die Regressionsgleichung restlos erklärt. Die Zahlenwerte der erklärten Varianz betragen für den vollbeladenen Zustand bei Froudezahlen bis 0,21 zwischen 70 und 82%, ab 0,22 zwischen 85 und 98%. Für den Ballasttiefgang ergaben sich über das gesamte Froudezahlintervall Werte zwischen 82 und 95%. Als Erklärung für die Differenz zu 100% können die oben erwähnten Gründe dienen.

Die von Froudezahl zu Froudezahl stets unterschiedliche Zusammensetzung der analysierten Datengruppe und die damit verbundene Auswahl voneinander abweichender Formvariablen in den Regressionen führten dazu, daß die über einen größeren Geschwindigkeitsbereich aus den Regressionsgleichungen berechneten Widerstandspunkte einen sprunghaften Verlauf aufweisen. Zur Ermittlung strakender Werte ist man deshalb auf die Hilfe einer graphischen Darstellung und das Einzeichnen einer mittelnden Kurve angewiesen. Die Vertrauensgrenzen sollten hierbei als wertvolle Unterstützung, d.h. als Maß für den Strakspielraum, angesehen werden. Für den Anwendungsfall empfiehlt es sich, wenn möglich auf beiden Seiten des notwendigen Geschwindigkeitsintervalls einige weitere Punkte zu berechnen, um den Kurvenverlauf sicherer festlegen zu können.

Es wird geschätzt, daß man den Gesamtwiderstand eines Schiffes unter Verwendung der Regressionsgleichungen für  $C_R$  im Mittel genauer als mit anderen Verfahren, vorhersagen kann. Der Grund für diese Annahme liegt darin, daß versucht wurde, der Formenvielfalt der Schiffe durch eine Fülle verschiedenster Formparameter gerecht zu werden.

Neben der Analyse des Restwiderstandsbeiwertes wurde noch eine weitere Untersuchung durchgeführt. Der Widerstand eines Schiffes oder seines Modells hat einen schwerkraftbedingten und einen Zähigkeitsbedingten Anteil. Für die Übertragung des im Modellversuch gemessenen Widerstandes auf das Schiff müssen verschiedene Ähnlichkeitsgesetze auf die beiden Anteile angewendet werden. Durch eine Zerlegung des Modellwiderstandes in  $C_T = C_{F0} + C_R$ , wo  $C_{F0}$  dem Reynoldsschen Gesetz und  $C_R$  dem Froudeschen Gesetz unterliegt, erhält man zwar praktisch brauchbare Ergebnisse, doch theoretisch begeht man einen Fehler.  $C_R$  erhält mit dem Formwiderstand, der ein Zähigkeitsbedingter Druckwiderstand ist, einen reynoldszahlenabhängigen Anteil, und ein physikalisch konsistenter Ansatz sieht daher eine Widerstandsaufspaltung in viskosen und Wellenwiderstand vor,  $C_T = C_V + C_W$ . Für den viskosen Widerstand wird der einfache Ansatz  $C_V = (1+k)C_{F0}$  gemacht, und man versucht, den sog. Formfaktor  $k$  aus Modellversuchen bei kleinen Froudezahlen zu bestimmen.

Einige Anteile des  $C_R$ -Meßdatenkollektivs waren für die entsprechende Auswertung geeignet. Mit insgesamt 112 aus Widerstandsversuchen auf voll beladenem und Ballasttiefgang ermittelten Zahlenwerten  $k$  konnte so mittels Regressionsanalyse eine Interpolationsformel für den Formfaktor aufgestellt werden. Die darin vorkommenden Parameter der Schiffsförmung lassen einen starken Einfluß streng lokaler Rumpfbereiche erkennen, besonders des Hinterschiffs und des Bugwulstes. Ein weiterer kombinierter Parameter, der von der ITTC 1972 als Gesamtinterpolator für den Formfaktor vorgeschlagen wurde, berücksichtigt den Einfluß der Hauptabmessungen. Die neue Formel gilt nur für völlige, langsamfahrende Schiffe.

Die graphisch ermittelten Formfaktoren für den vollen Tiefgang wiesen eine durch die zur Verfügung stehenden Schiffsförmparameter nur schwer erklärbare Streuung auf. Dies wurde im wesentlichen auf versuchstechnisch bedingte Einflüsse, z.B. Strömungsstabilitäten durch zu kleine Reynoldszahlen, zurückgeführt. Zugleich wurde festgestellt, daß, offenbar bedingt durch die in diesem Fall schlankere Unterwasserform, Formfaktoren des Ballasttiefganges derartigen Einflüssen in weit geringerem Ausmaß unterliegen.

In der als Kompromiß durchgeführten gemeinsamen Analyse beider Tiefgangsgruppen überwogen die Störfaktoren des vollen Tiefganges. Die Regressionsgleichung erklärt die Varianz der gegebenen  $k$ -Werte nur zu 80%. Dieses Erwendung der Formel zum Zweck der Modell-Schiff-Korrelation zu gestatten.

Abschließend sei darauf hingewiesen, daß sämtliche Ergebnisse der Regressionsanalysen den Gesetzen der statistischen Methoden unterliegen. Sie sollten nur auf Schiffsentwürfe angewendet werden, deren Formmerkmale aus den Grenzen der hier analysierten Stichproben stammen. Von Extrapolationen ist abzuraten, weil die funktionalen Zusammenhänge zwischen Restwiderstandsbeiwert bzw. Formfaktor und Formvariablen weder theoretisch noch aus Erfahrung ausreichend bekannt sind. Zur Information des Anwenders wurden den Ergebnisgleichungen umfangreiche Tabellen beigefügt, aus den alle benötigten Angaben, z.B. über kleinste und größte Werte der Stichprobenvariablen sowie die Stichprobenverteilungen, entnommen werden können.

Diese Arbeit wurde aus Mitteln des Sonderforschungsbereiches 98 "Schiffstechnik und Schiffbau" der Deutschen Forschungsgemeinschaft bezahlt. Den zweitweisen Mitarbeitern, Herrn H. Lochte-Holtgreven, Frä. S. Reimlinger und Frau I. Schweeren sei aufrichtig gedankt. Alle Berechnungen wurden auf der TR440 des Rechenzentrums der Universität Hamburg mit elfstelliger Genauigkeit durchgeführt.

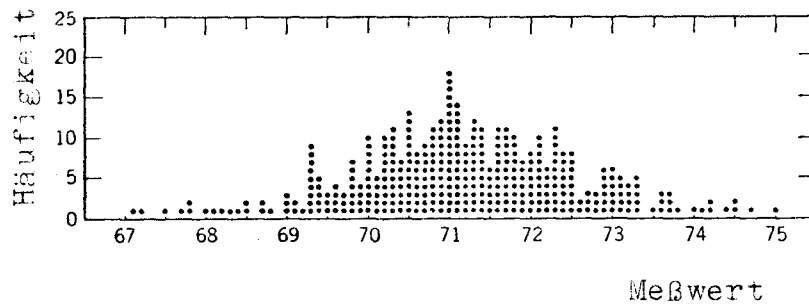


Bild 1: Punktediagramm der Verteilung der Meßwerte eines Merkmals

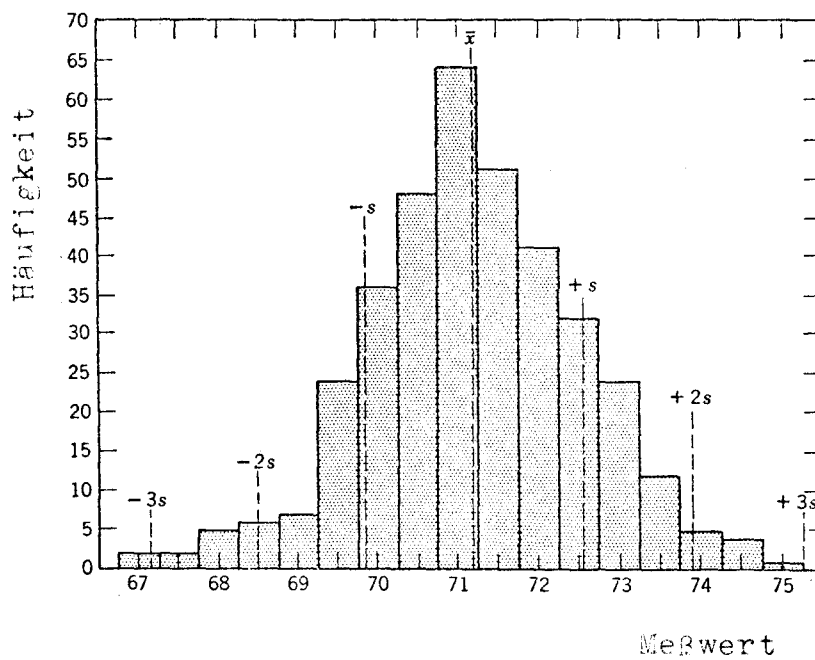


Bild 2: Häufigkeitsdiagramm der Daten von Bild 1 mit 17 Klassen. Angabe des Mittelwertes  $\bar{x}$  und der Bereiche  $\bar{x} \pm 1, 2, 3s$ ,  $s$  = Standardabweichung

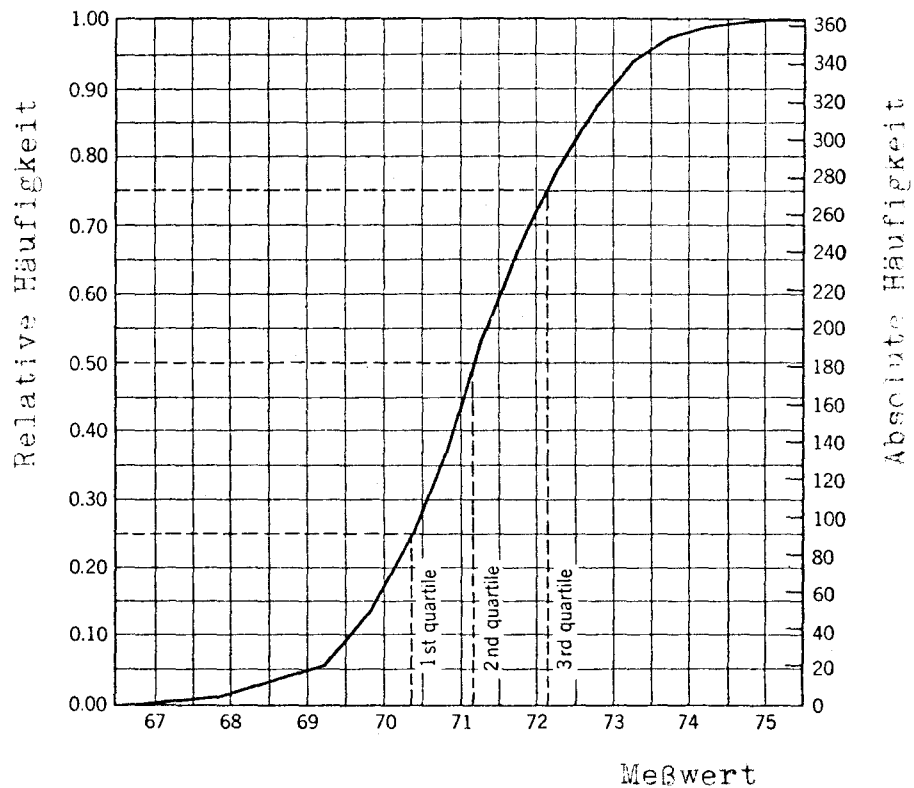


Bild 3: Summenhäufigkeit der Daten von Bild 1

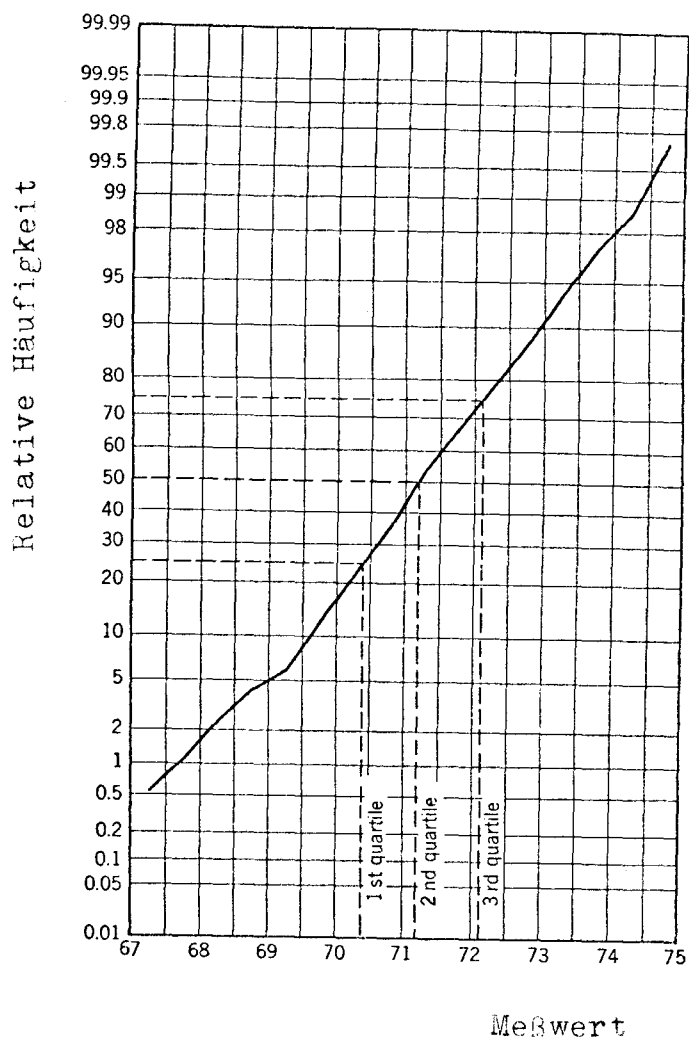


Bild 4: Summenhäufigkeit der Daten von Bild 1 auf Wahrscheinlichkeitspapier

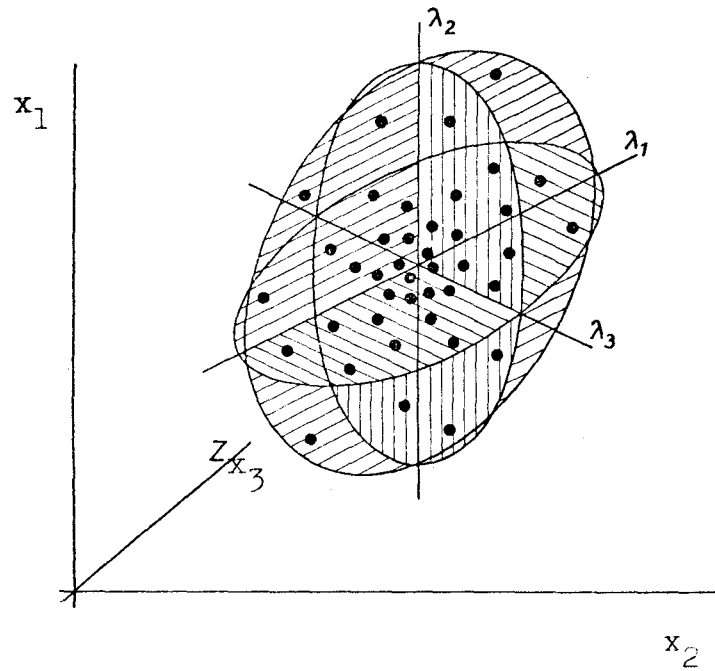


Bild 5: Veranschaulichung der Hauptachsenmethode.  $x_1, x_2, x_3$  sind die ursprünglichen Koordinatenachsen,  $\lambda_1, \lambda_2, \lambda_3$  das Koordinatensystem der Hauptachsen. Der die Punktwolke umhüllende Körper ist in drei Ebenen geschnitten, die senkrecht zueinander stehen und durch den Schwerpunkt gehen

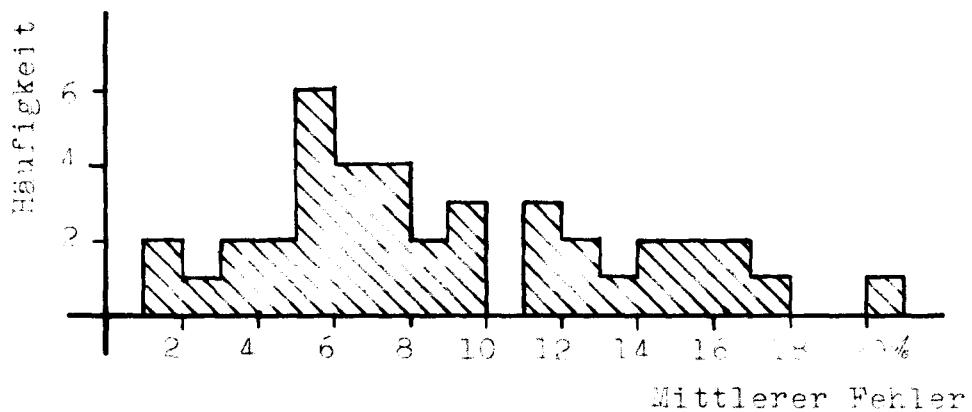


Bild 6: Verteilung des gemittelten Fehlers aus dem Vergleich von Ergebnissen nach GULDHAMMER/HARVALD mit Messungen von 40 Widerstandsversuchen

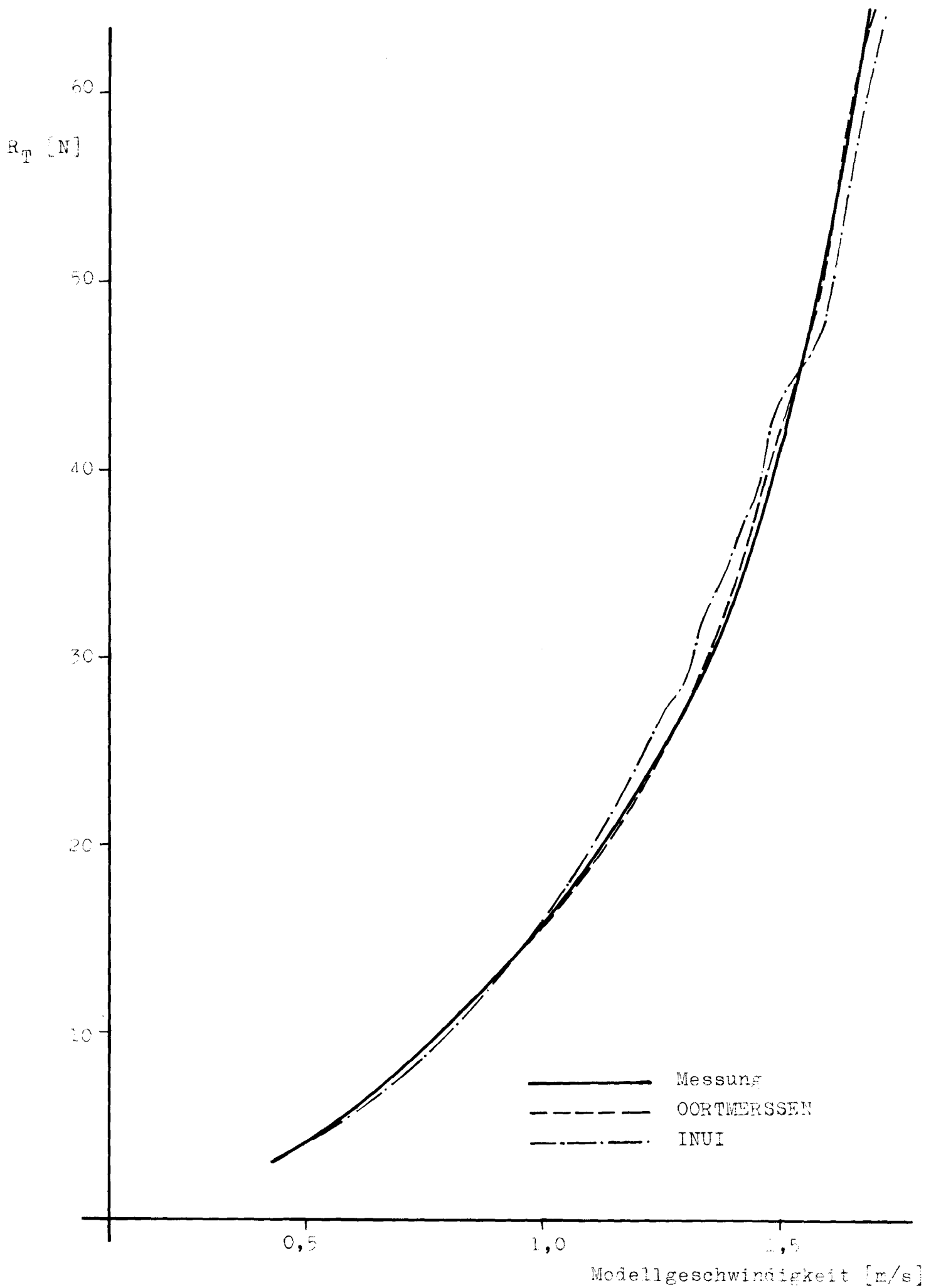


Bild 7: Angleichung einer gemessenen Widerstandskurve durch Widerstandsformeln nach OORTMERSSEN und INUI

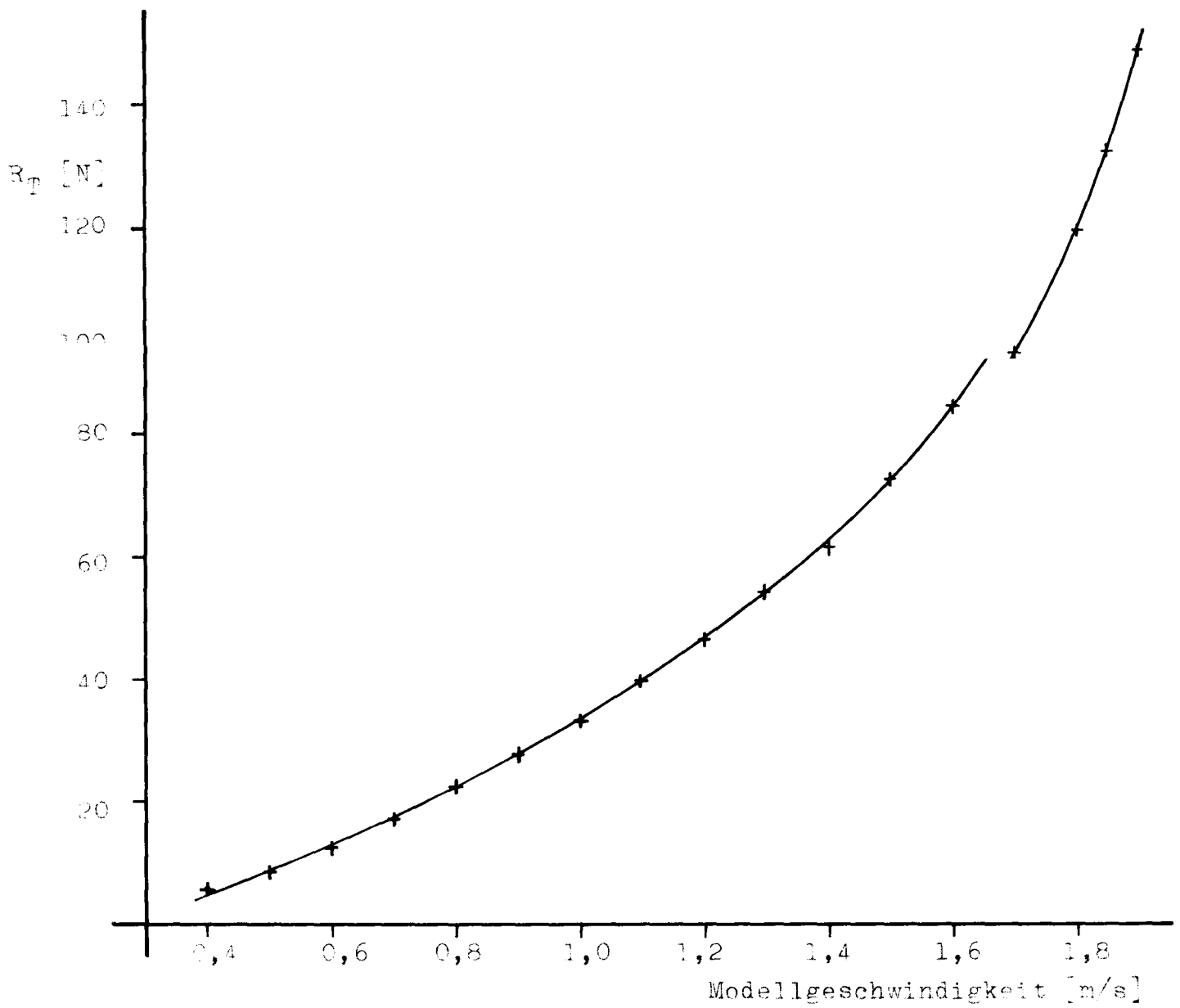


Bild 8: Widerstandsmeßkurve eines Schiffmodells

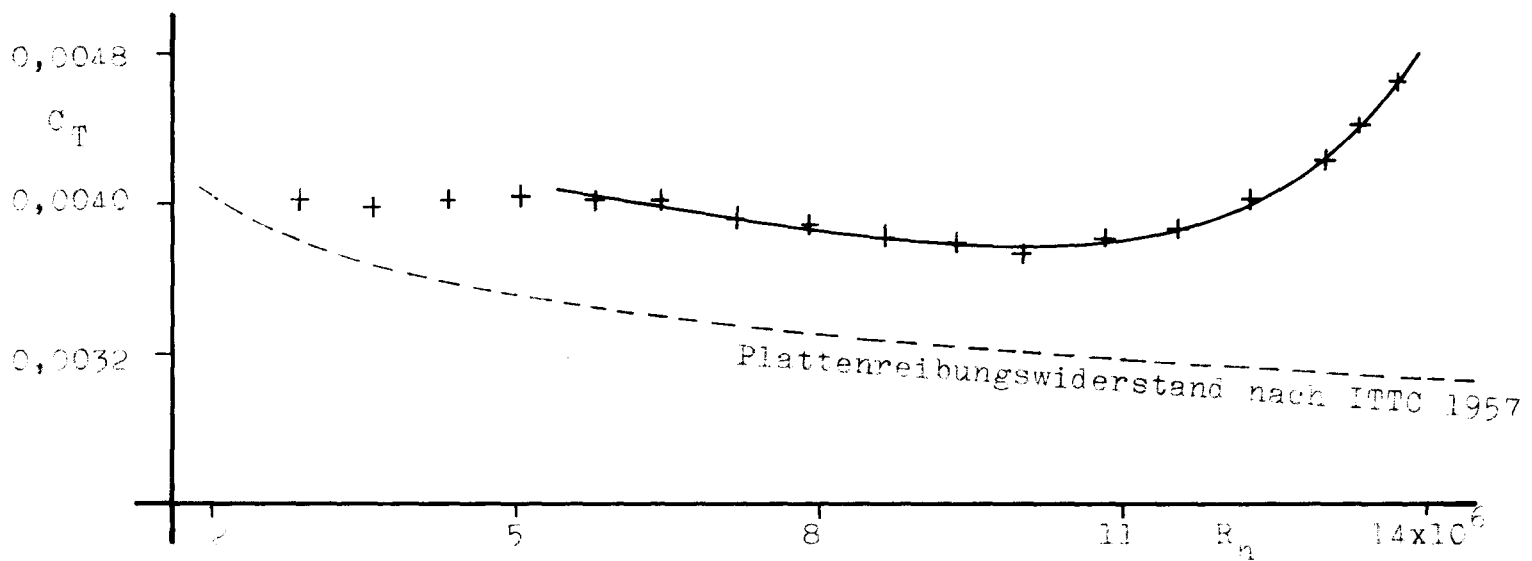


Bild 9: Gesamtwiderstandsbeiwert über der Reynoldszahl für die Messung von Bild 8

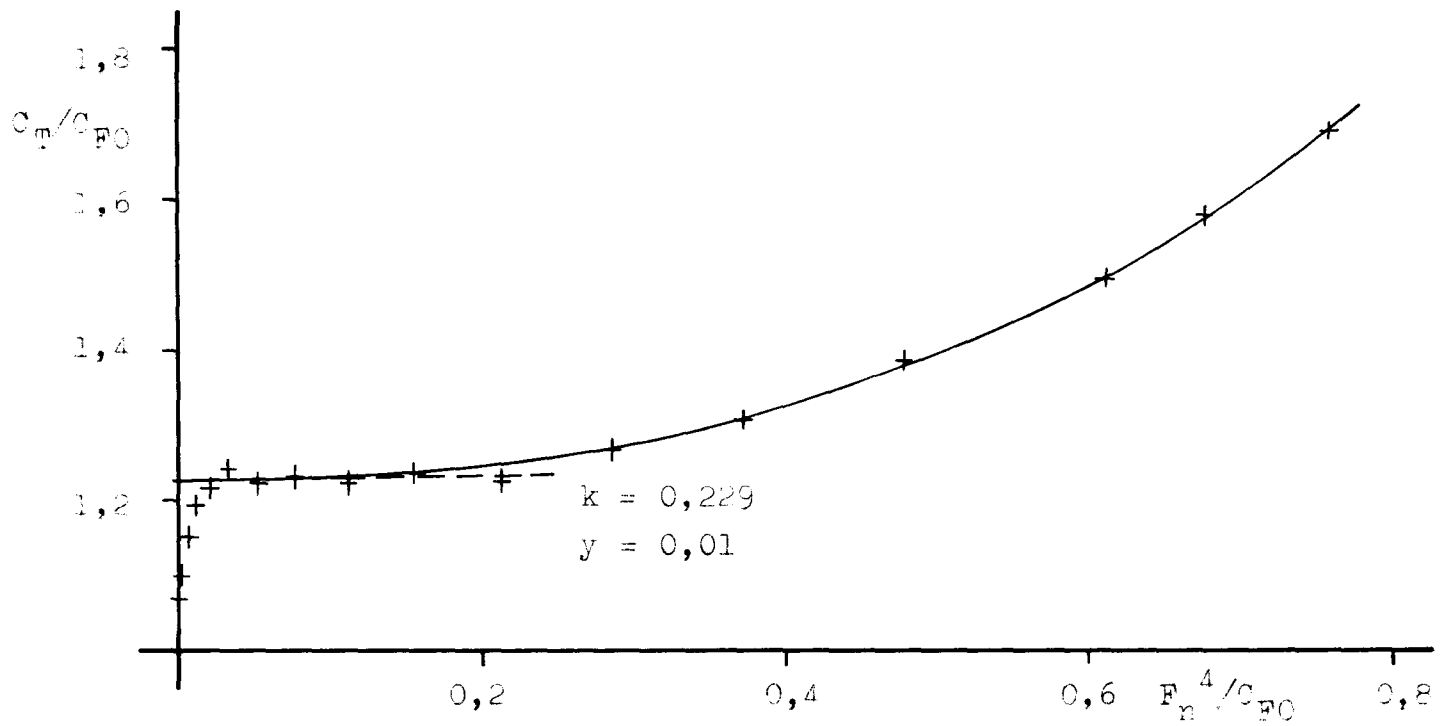


Bild 10:  $C_T/C_{FO}$  über  $F_n^4/C_{FO}$  (Auftragung nach HUGHES/PROHASKA) für die Messung von Bild 8

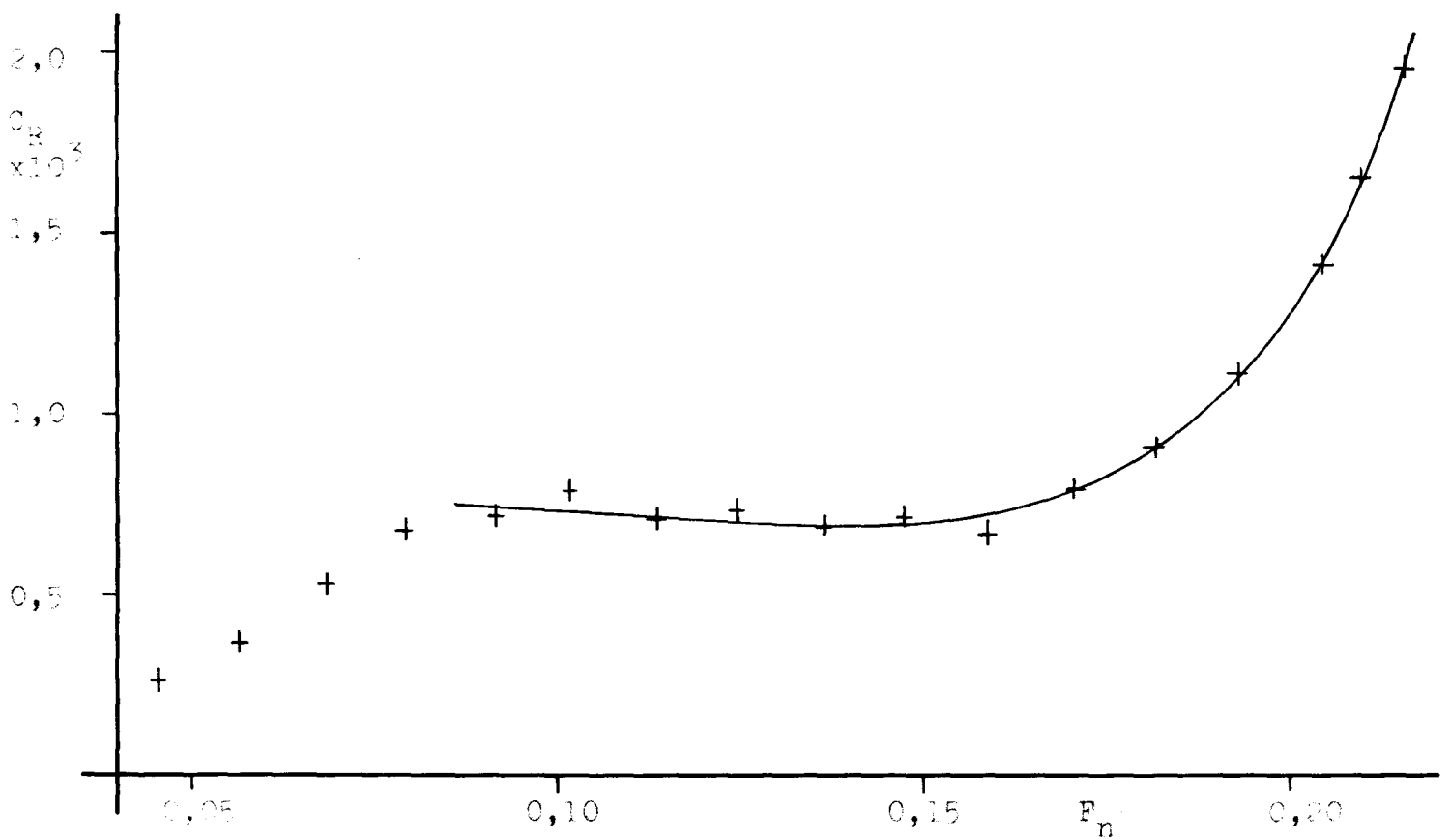


Bild 11: Restwiderstandsbeiwert über der Froudezahl für die Messung von Bild 8

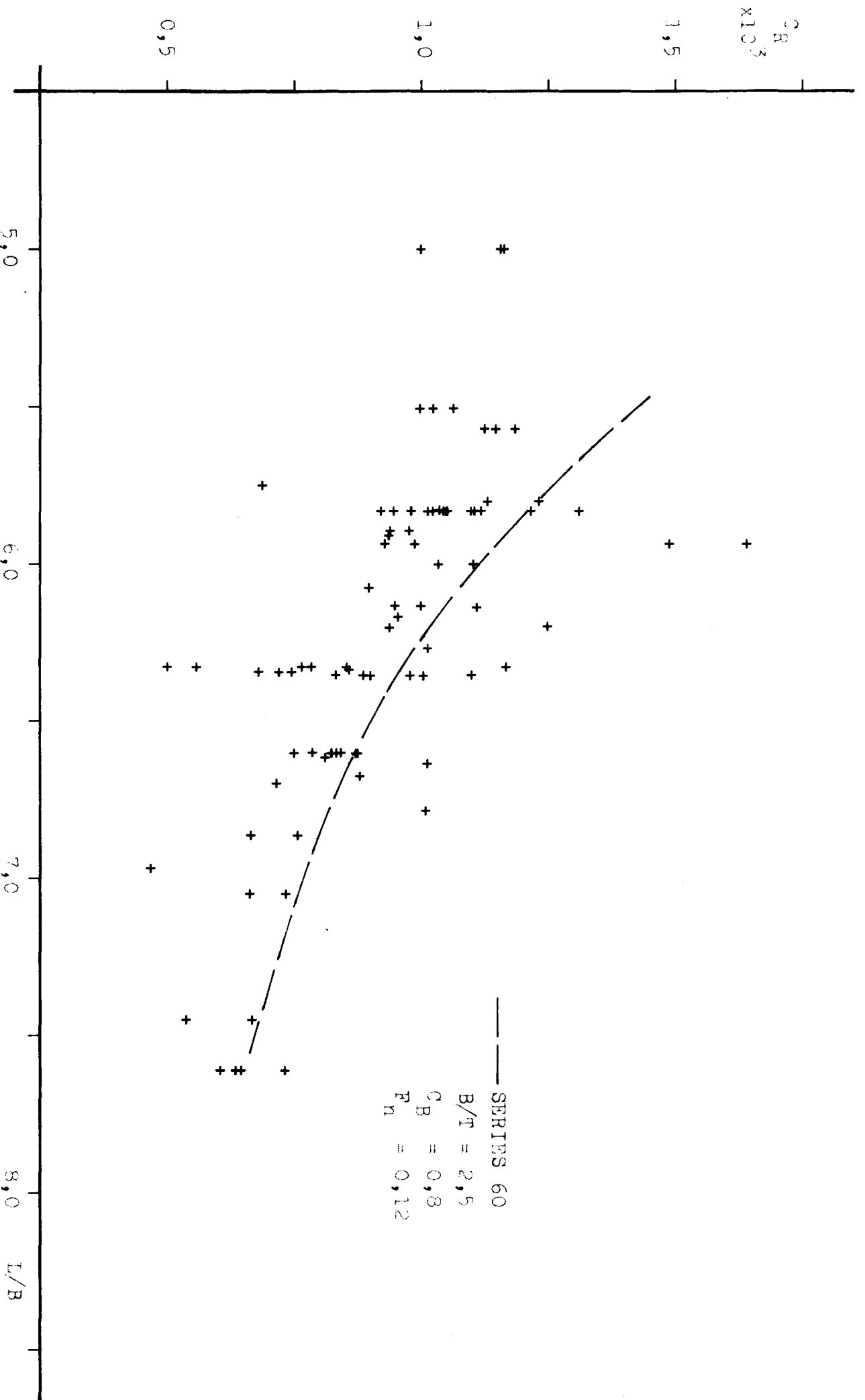
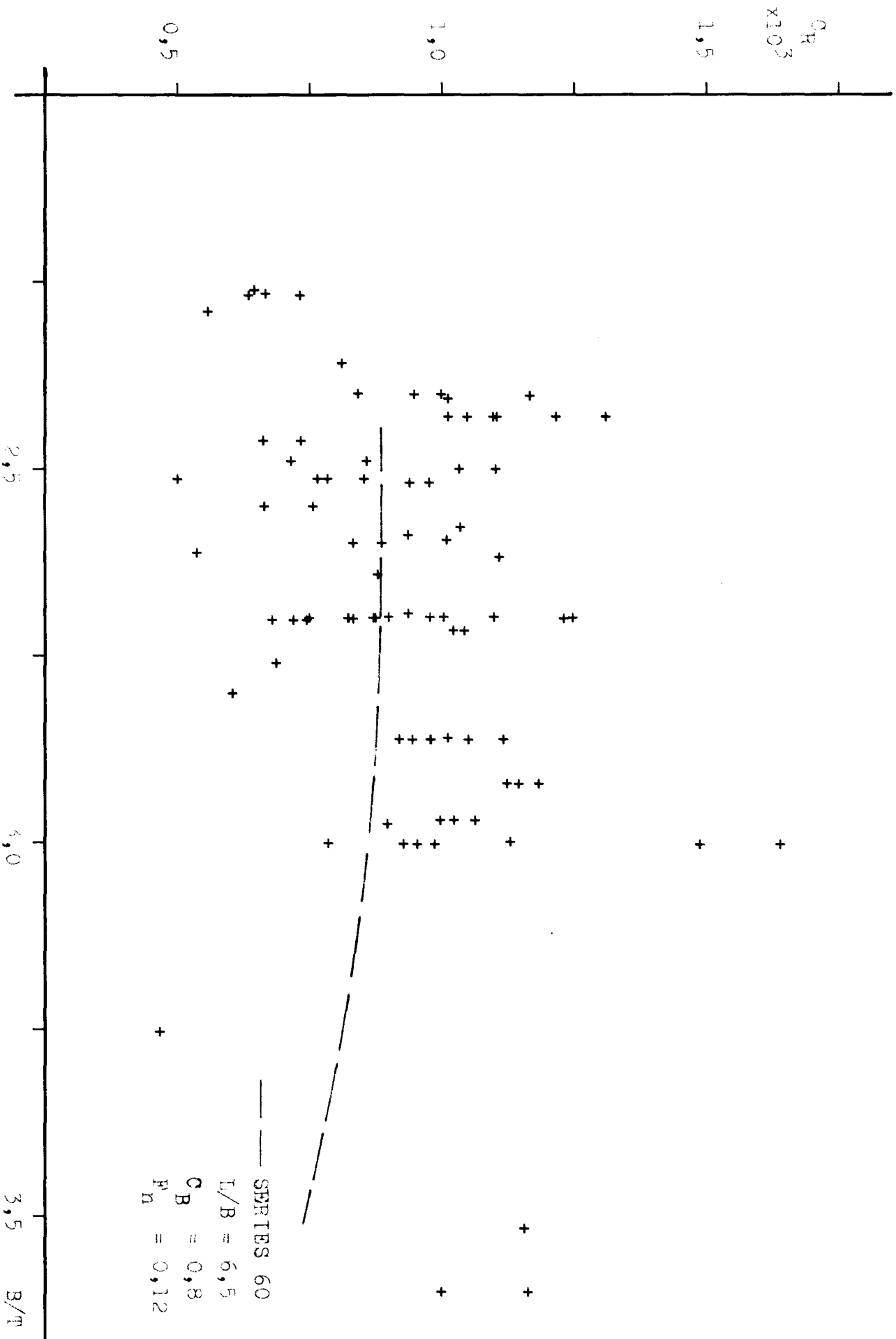


Bild 12:  $C_R$  über  $L/B$ ,  $F_n = 0,14$ , für einschraubige Frachter, Tanker, Erz/Ober-Schiffe, beladen



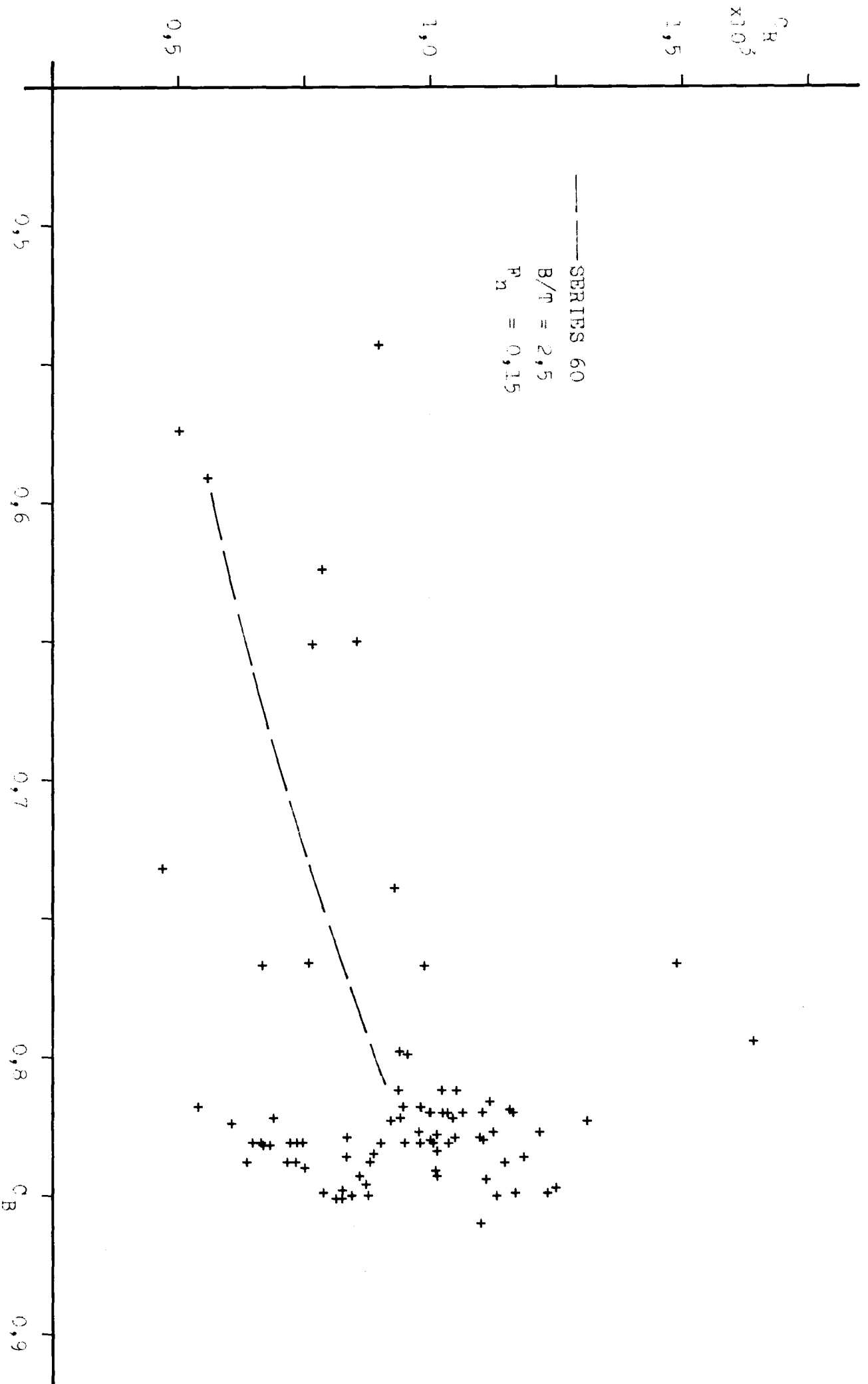


Bild 14:  $C_R$  über  $C_B$ ,  $F_n = 0,14$ , für einschraubige Frachter, Tanker, Erz/Obo-Schiffe, beladen

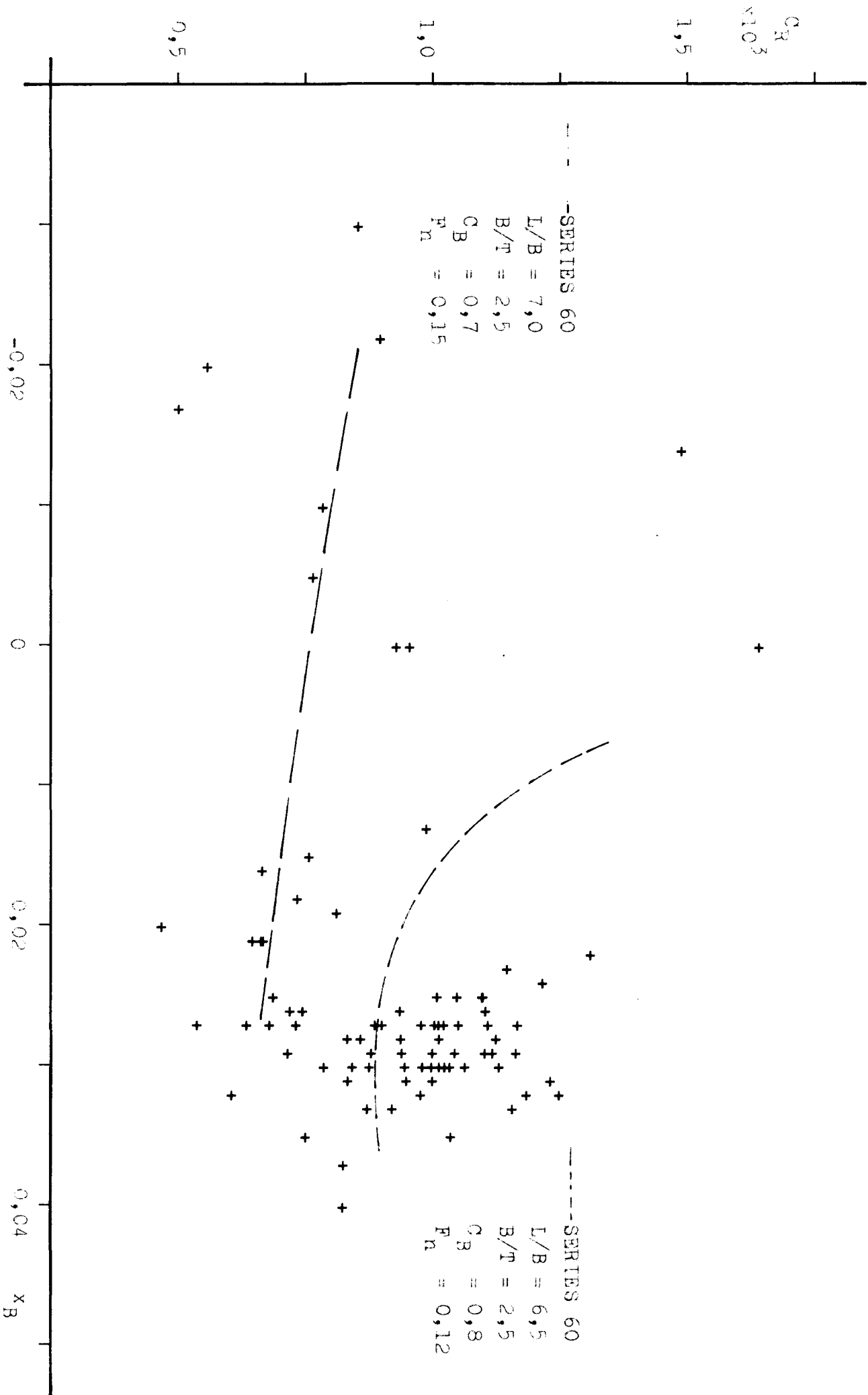


Bild 15:  $C_R$  über  $x_B$ ,  $F_n = 0,14$ , für einschraubige Frachter, Tanker, Frz/Obb-Schiffe, beladen

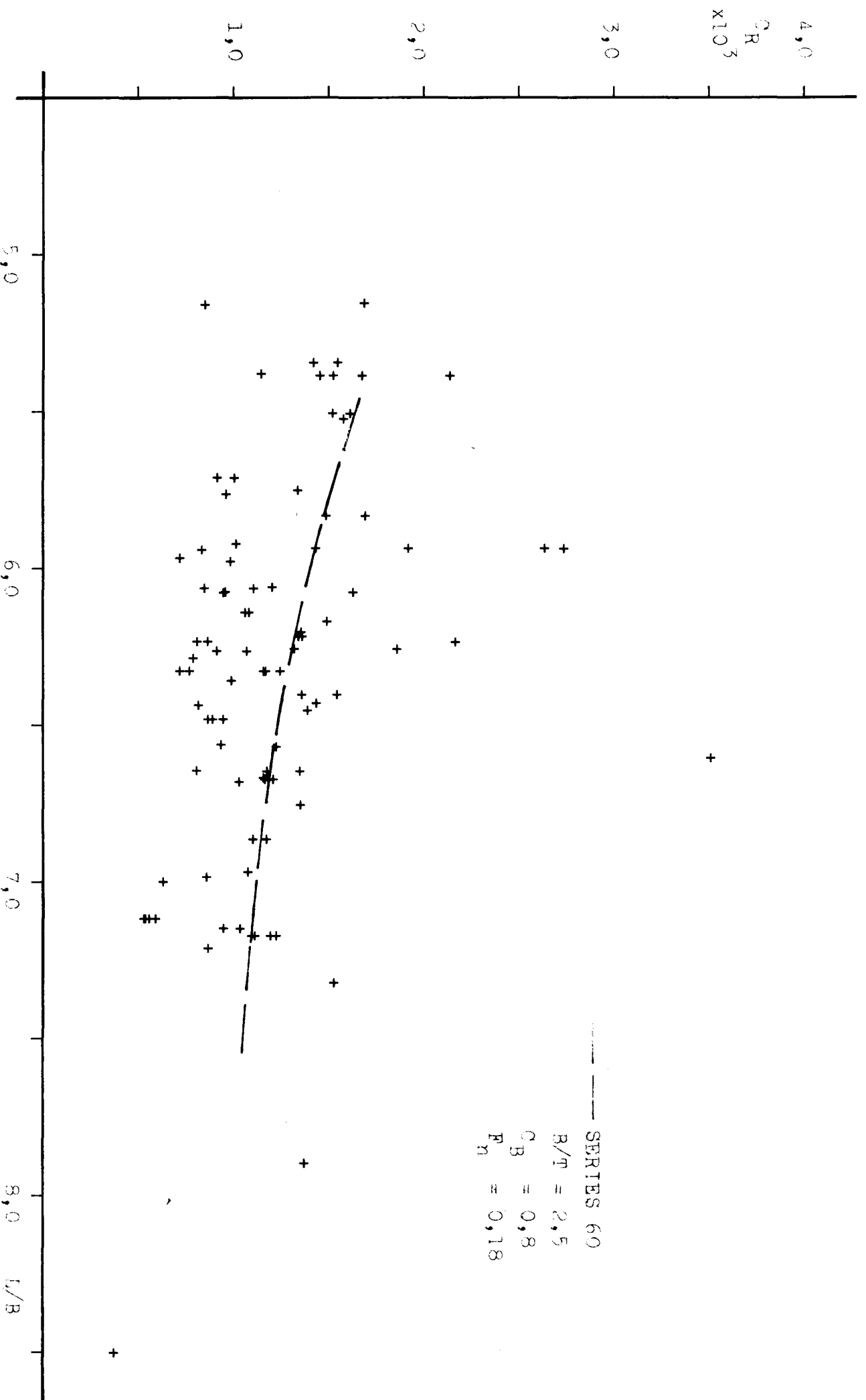


Bild 16:  $C_R$  über  $L/B$ ,  $F_n = 0,20$ , für einschraubige Frachter, Tanker, Erz/Obo-Schiffe, beladen



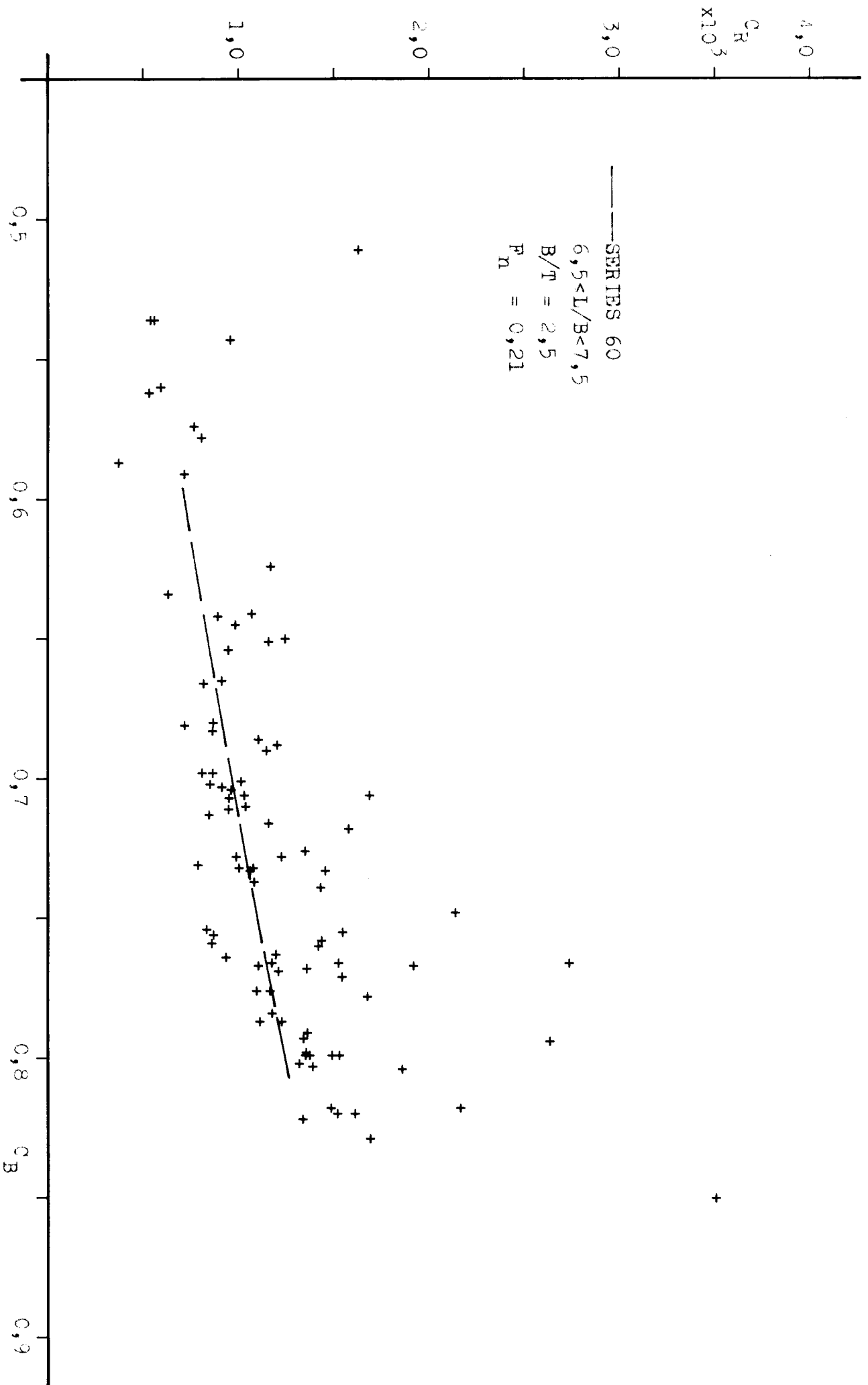


Bild 18:  $C_R$  über  $C_B$ ,  $F_n = 0,20$ , für einschraubige Frachter, Tanker, Erz/Obo-Schiffe, beladen

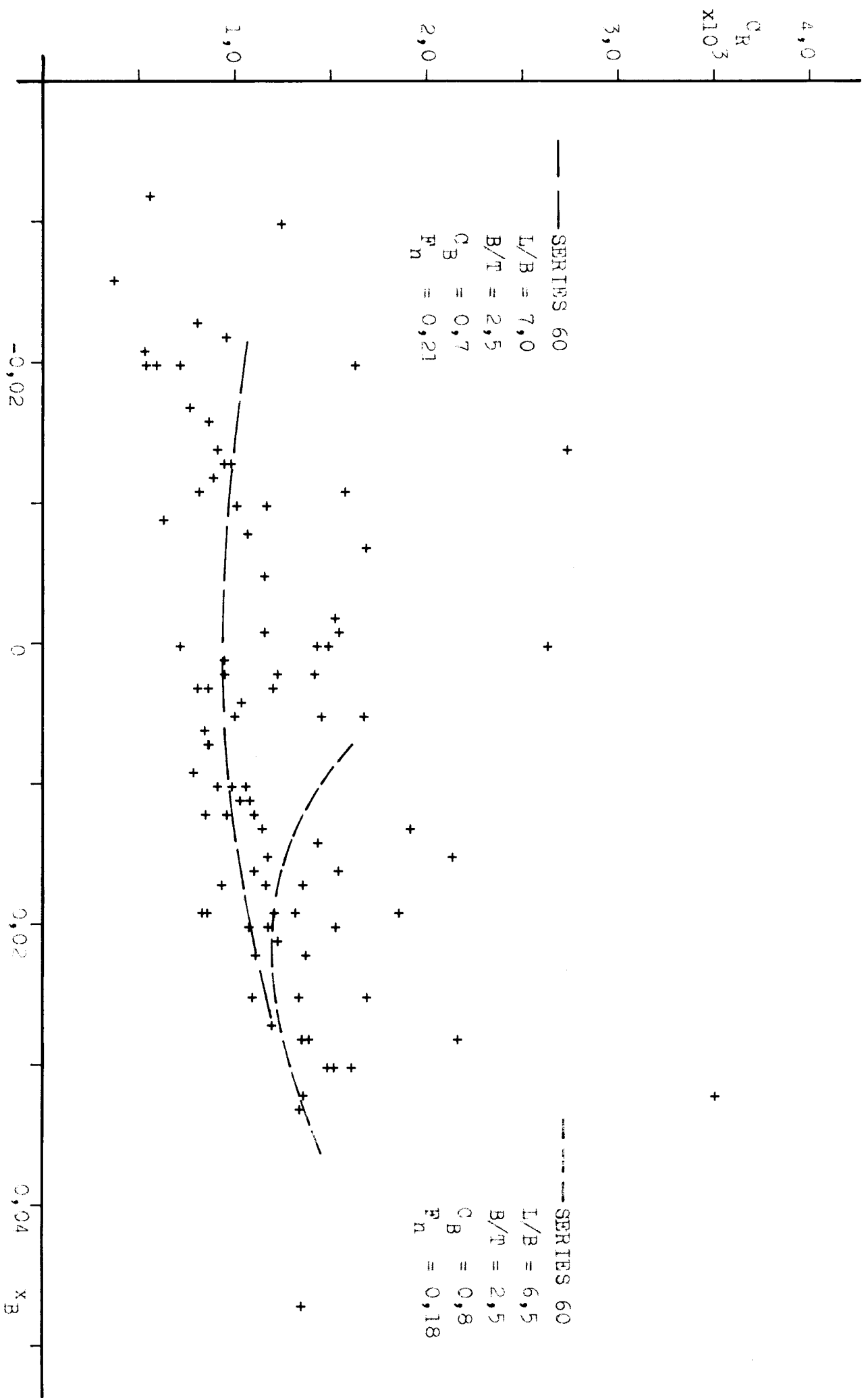
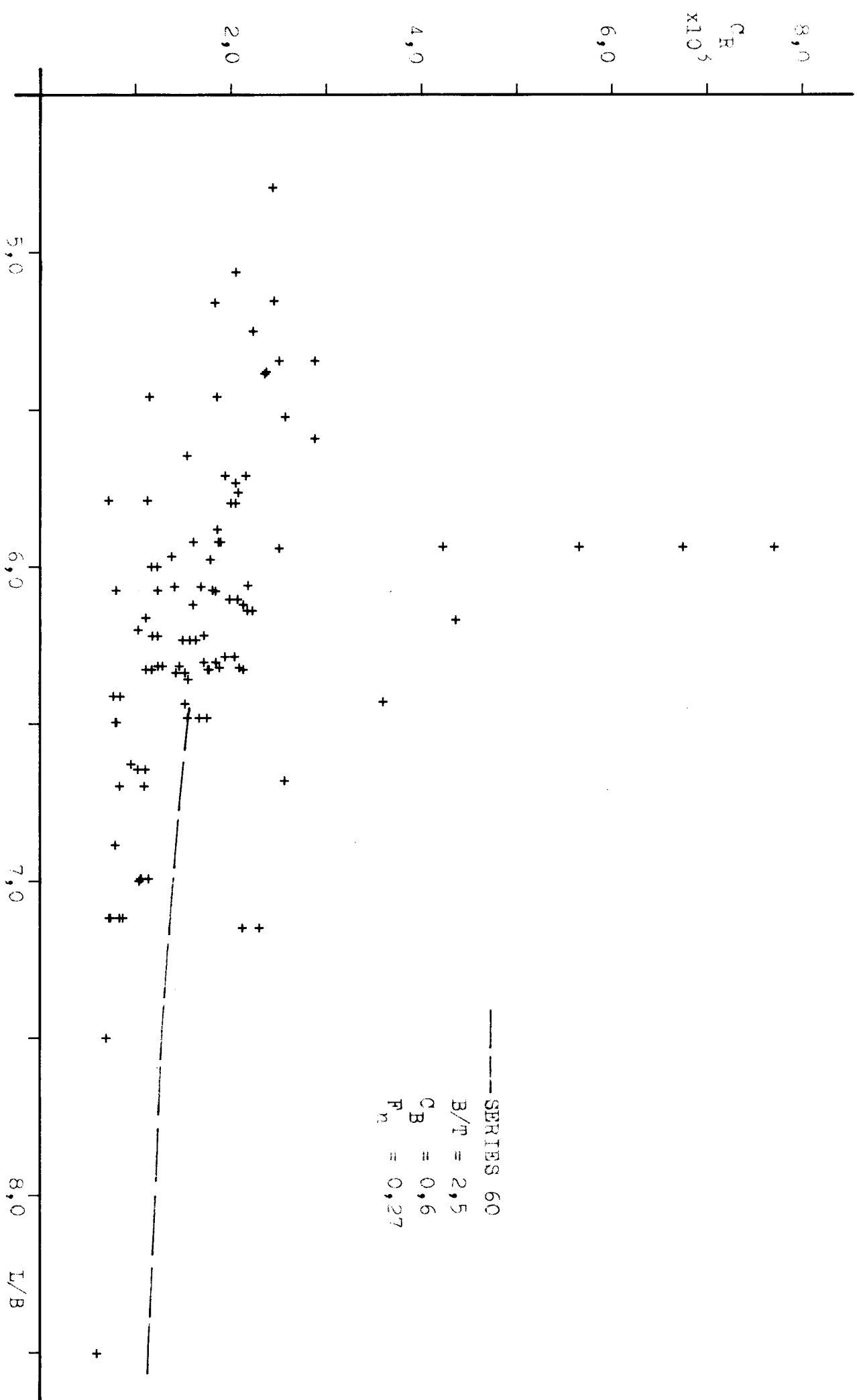


Bild 19:  $C_R$  über  $x_B$ ,  $F_n = 0,20$ , für einschraubige Frachter, Tanker, Frz/Obo-Schiffe, beladen



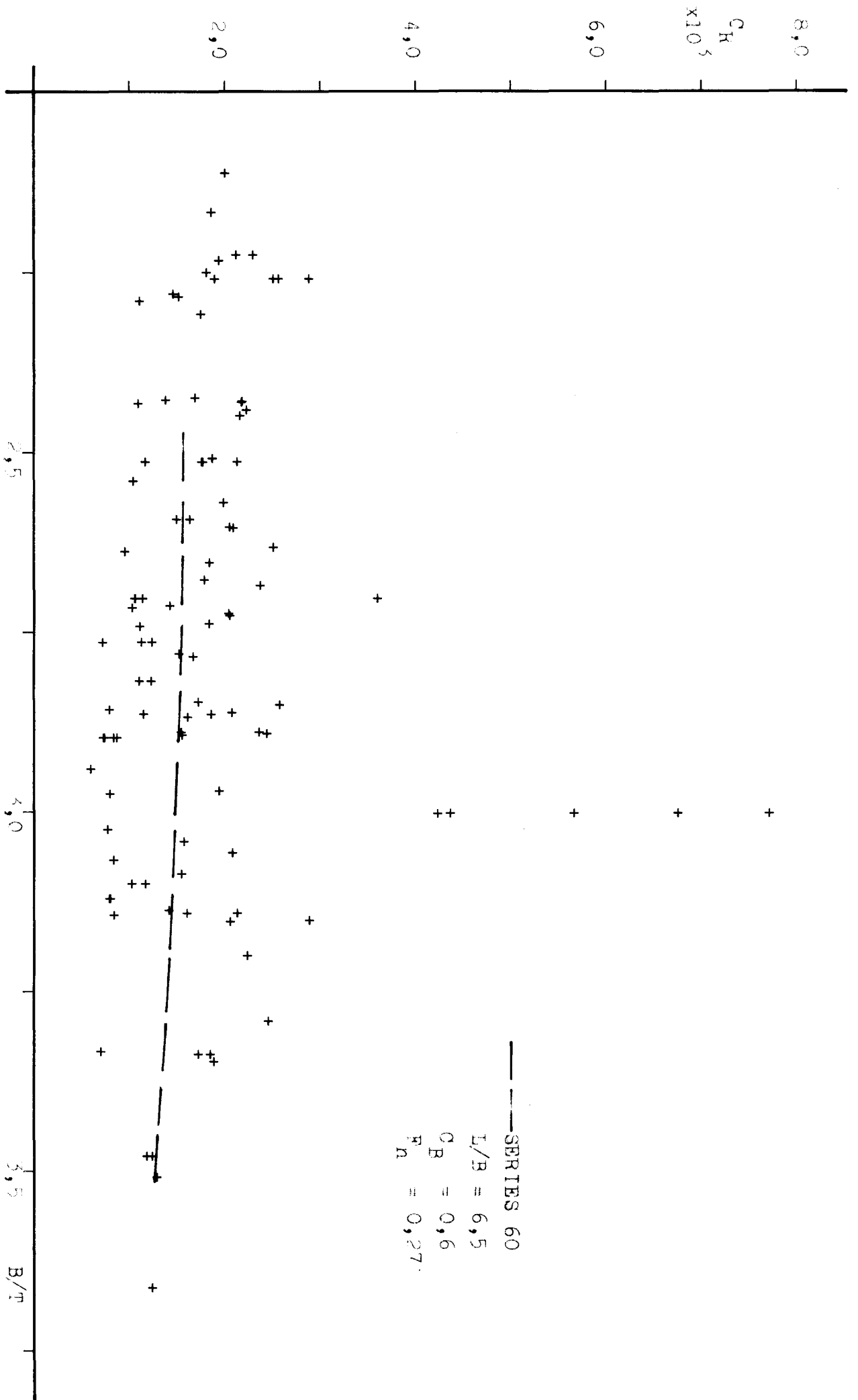


Bild 21:  $C_R$  über  $B/T$ ,  $F_n = 0,26$ , für einschraubige Frachter, Tanker, Erz/Ober-Schiffe, beladen

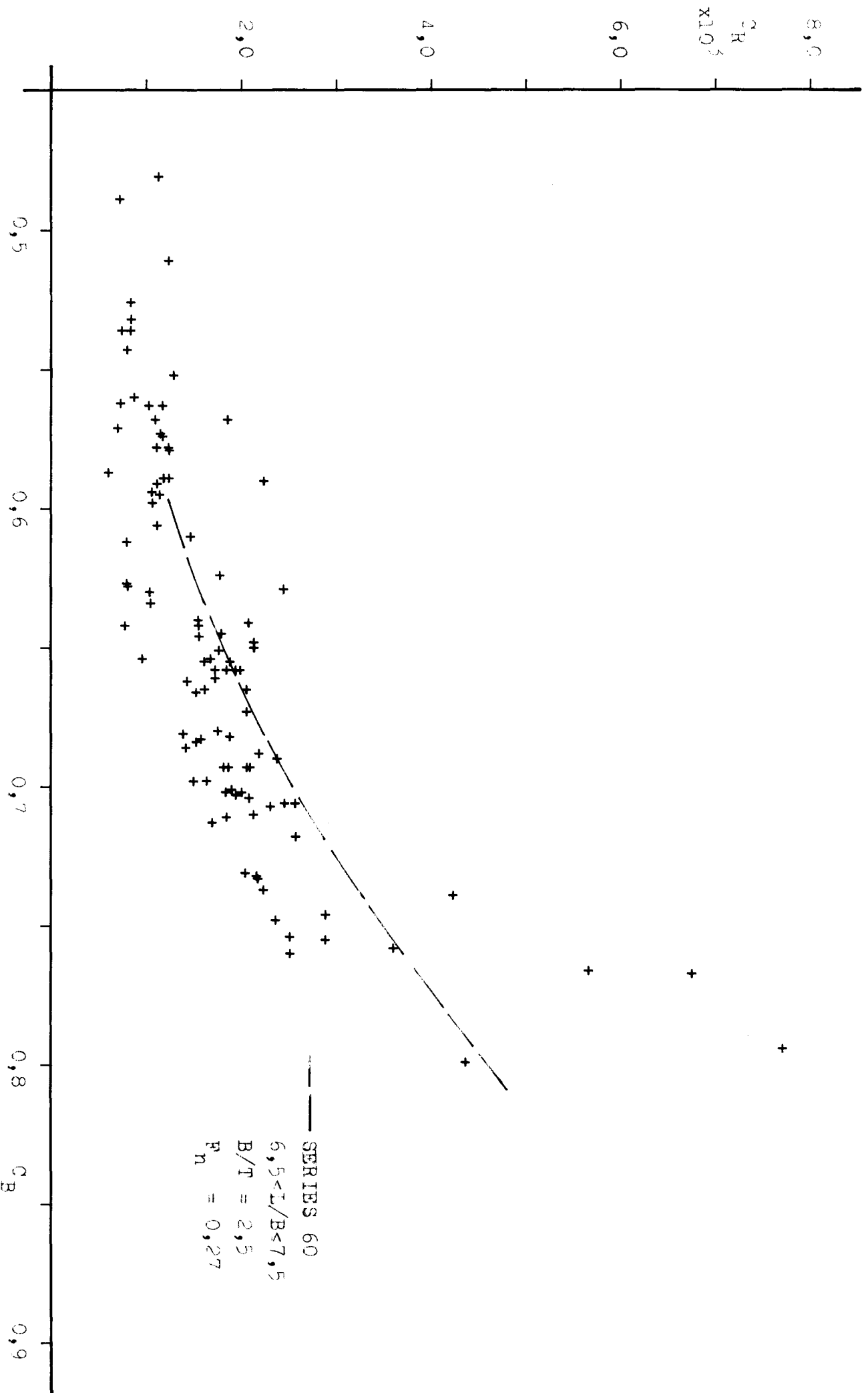


Bild 22:  $C_R$  über  $C_B$ ,  $R_n = 0,26$ , für einschraubige Prachtter, Tanker, Frz/Obo-Schiffe, beladen

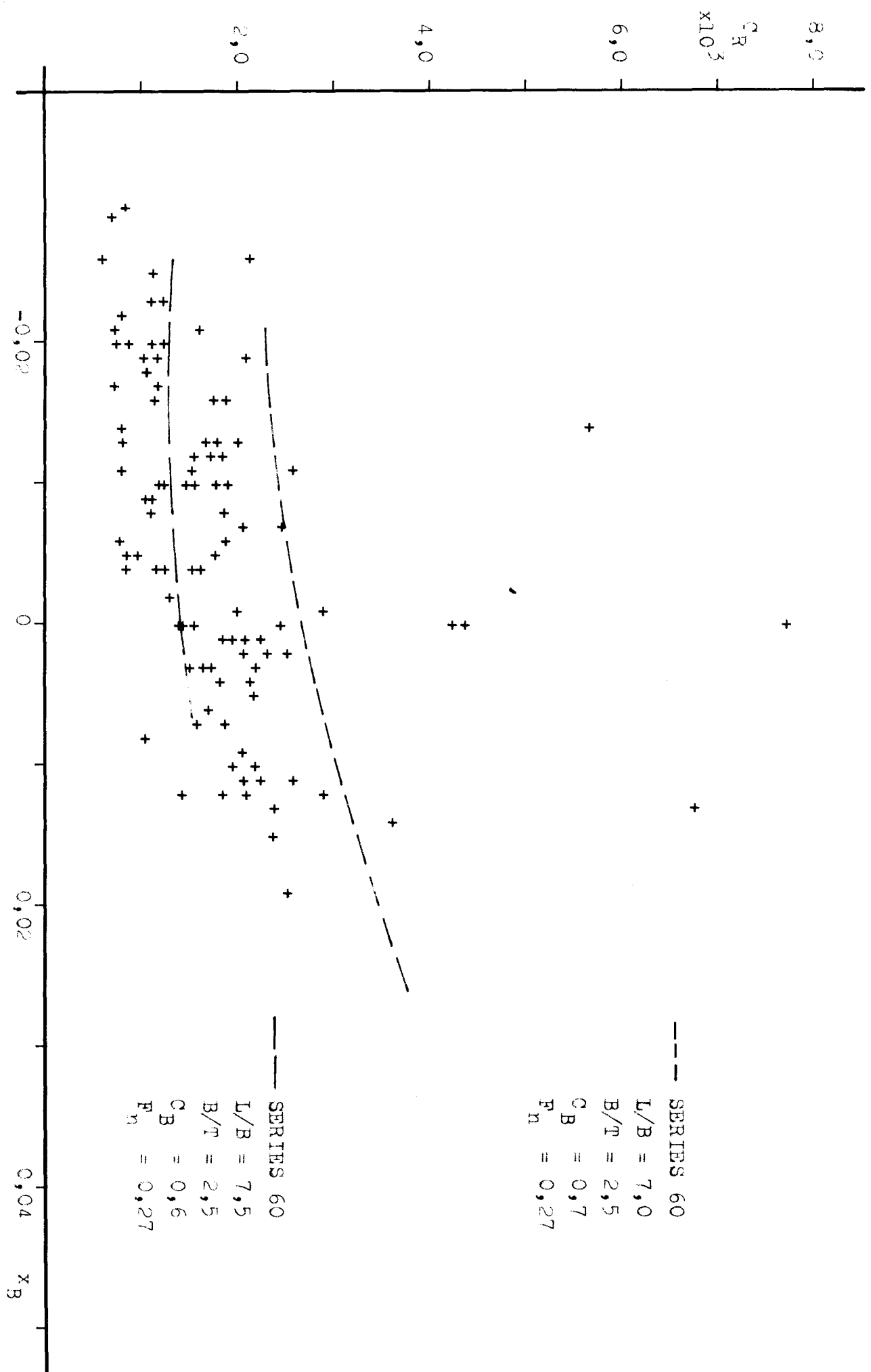


Bild 23: über  $x_B$ ,  $F_n = 0,26$ , für einschränbige Frachter, Tanker, Erz/Obschiffe, beladen

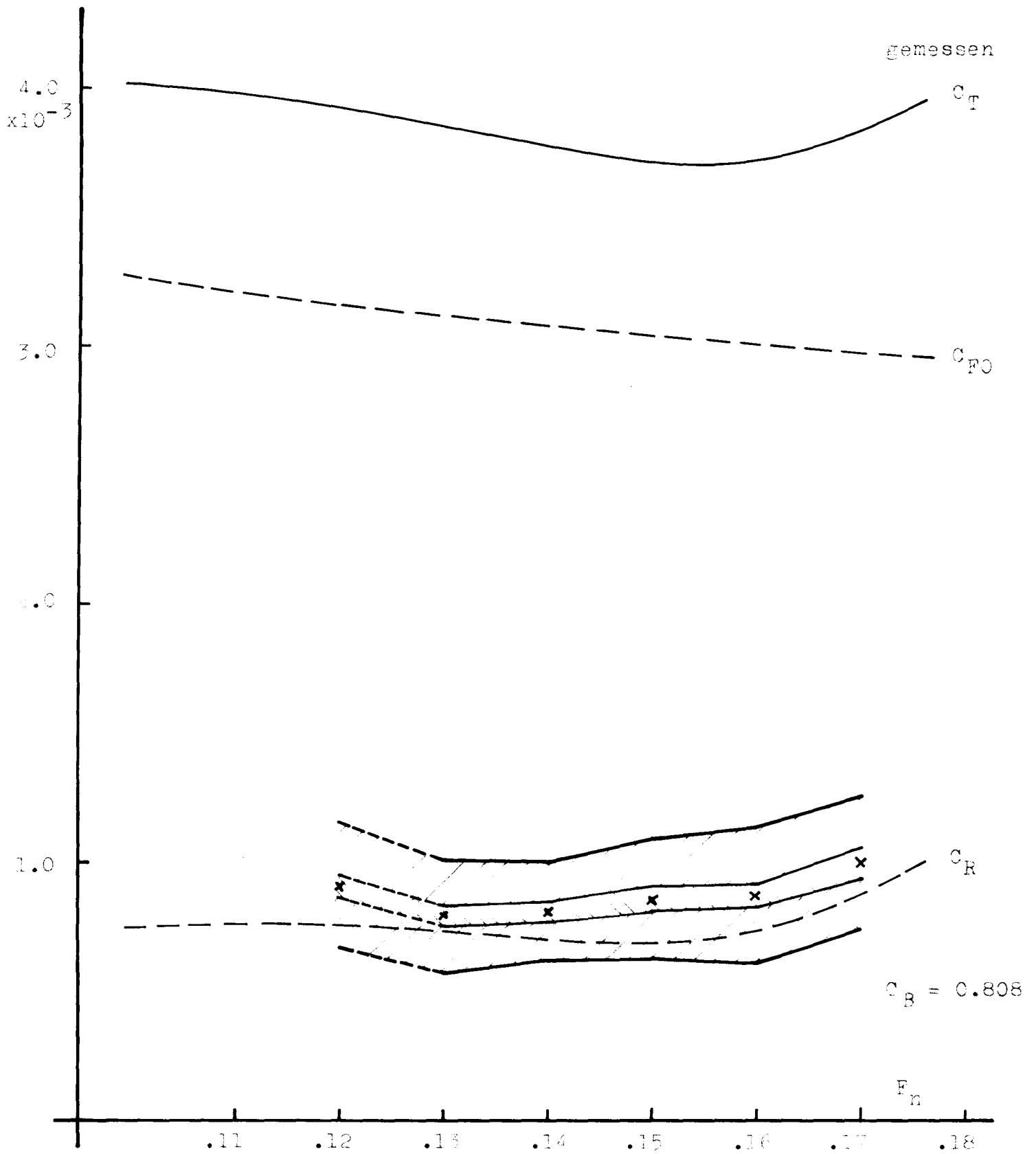


Bild 24: Berechnete  $C_R$ -Werte (x) eines Datenschliffes und Vertrauensgrenzen für die Regressionsgleichungen (innere Schraffur) sowie die Einzelwerte (äußere Schraffur)

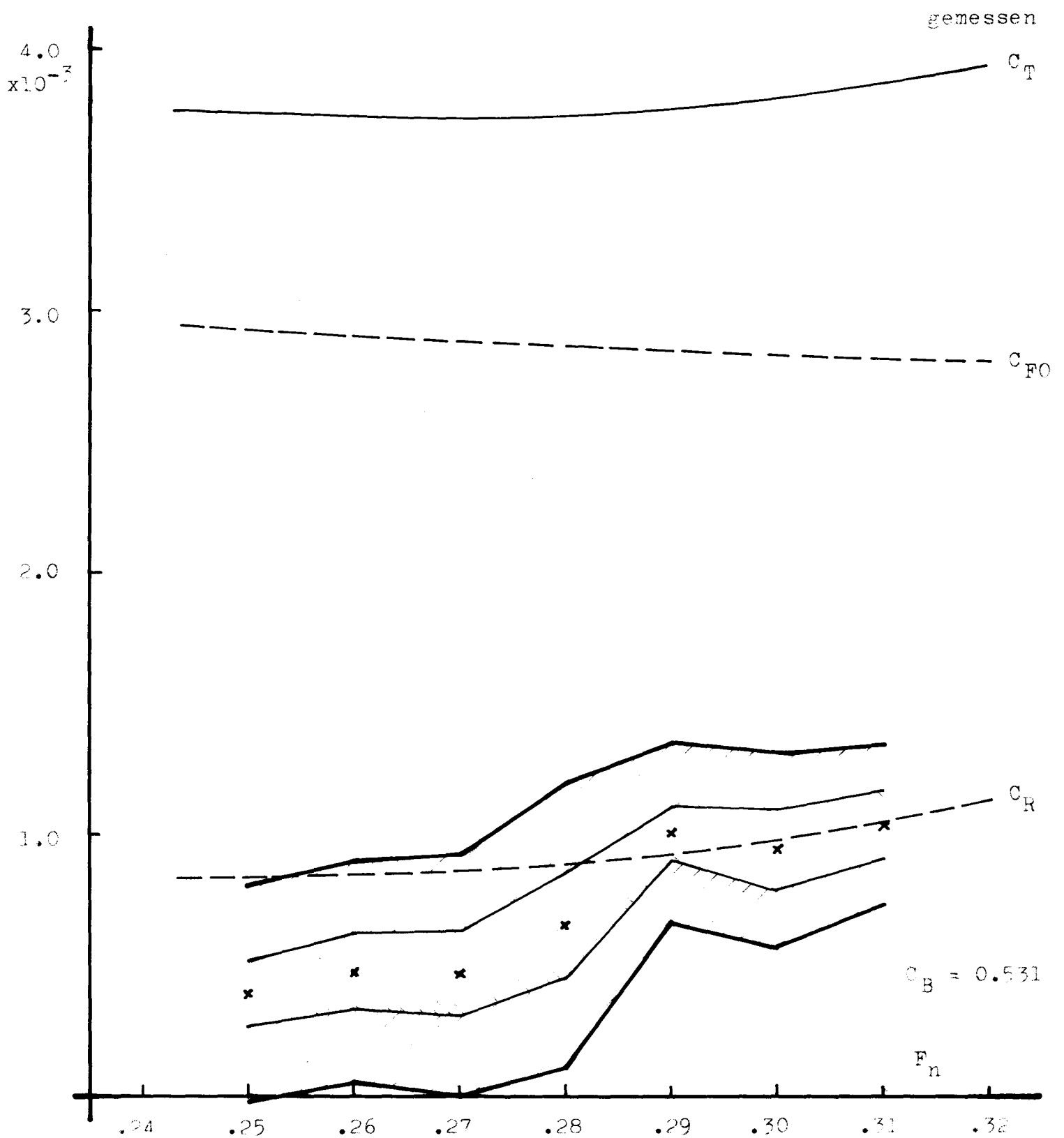


Bild 25: Berechnete  $C_R$ -Werte (x) eines Datenschiffes und Vertrauensgrenzen für die Regressionsgleichungen (innere Schraffur) sowie die Einzelwerte (äußere Schraffur)

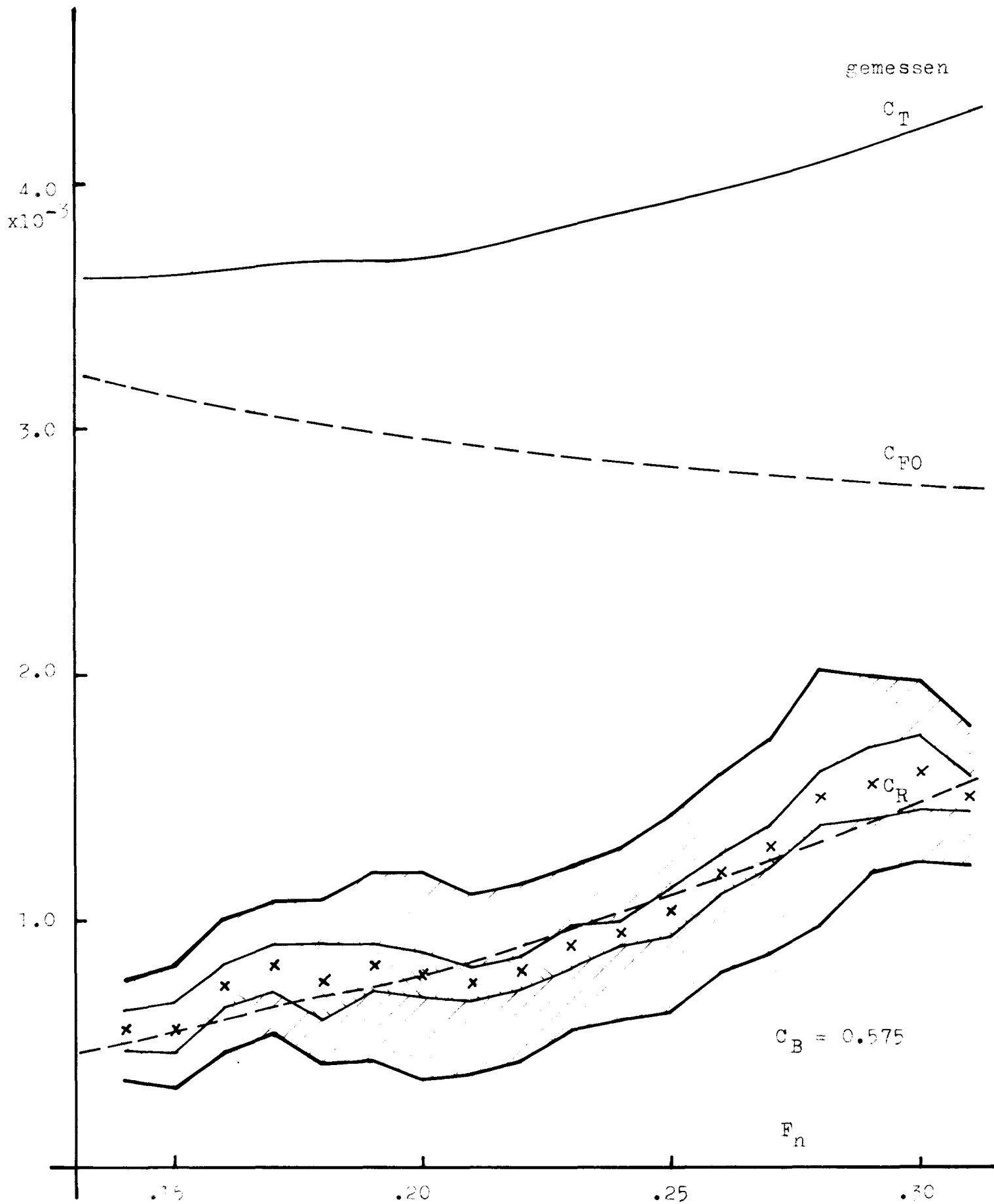


Bild 26: Berechnete  $C_R$ -Werte (x) eines Datenschiefes und Vertrauensgrenzen für die Regressionsgleichungen (innere Schraffur) sowie die Einzelwerte (äußere Schraffur)