

Anwendungs-Programmsystem DAPROA

Verarbeitungsprogrammbaustein LIREV

Version 5 .1

Stand: 31.10.2011

Leistung:

Schätzung der linearen Koeffizienten A_j eines mathematischen Modells in dem die Modellterme linear auftreten mittels linearer Regressionsanalyse,
mit: Y Zielgröße, X_i Einflussgrößen

$$Y = A_0 + A_1 \cdot X_1 + A_2 \cdot X_2 + \dots + A_n \cdot X_n$$

mit der Methode der kleinsten Fehlerquadrate:

Die Summe { der Quadrate_(berechneter_Wert minus gemessener Wert) }
wird zum Minimum geführt

Summe der Fehlerquadrate \rightarrow Minimum

Das Programm LIREV ist besonders geeignet für die Bearbeitung großer Datenmengen.

Mit ihm kann der lineare Zusammenhang zwischen Messgrößen, Datenmengen (Y , X_1 , X_2, \dots, X_n) in einem n -dimensionalen Koordinatensystem bzw. abhangigen und unabhängigen Variable untersucht werden.

Die Terme X_1, \dots, X_n können natürlich auch durch Datentransformation aus ursprünglichen Werten von unabhängigen Variablen entstanden sein.

Eine Verarbeitung unvollständiger Datensätze wird realisiert, indem bezüglich des jeweiligen Modellansatzes nur die gegenseitigen Korrelationskoeffizienten einbezogen werden.

Die Methode der Kamm- oder Verzerrungsregression kann angewendet werden.

Die Kammregression ist geeignet für Anwendungsfälle, bei denen die Einflussgrößen (unabhängige Variable) untereinander stark korreliert sind. Um die dann oft auftretenden Überschätzungen der gegenseitigen Wirkung der Einflussgrößen zu vermeiden und die korrelierten Einflussgrößen möglichst zu trennen, wird auf die erwartungstreue Schätzung verzichtet und das Schätzproblem verzerrt. Das wird erreicht durch Modifikation der Hauptdiagonale der Informationsmatrix durch Addition mit einem Verzerrungsparameter aus dem Intervall $(0,1)$. Die Folge ist aber, dass die Bestimmtheit und die Reststreuung des Ergebnisses ungünstiger werden.

Das Programm LIREV wird normal in 2 und maximal in 12 Zyklen abgearbeitet.

- Im 1.Zyklus: lineare Regression, t-Test, Verzerrung = 0
- Im 2.Zyklus: falls nach t-Test gesicherte **und** ungesicherte Koeffizienten vorliegen lineare Regression, Ergebnis mit nur gesicherten Koeff., Verzerrung = 0
- ab 3. bis max. 12. Zyklus: falls Anzahl der Verzerrungsvarianten > 0 lineare Regression, t-Test, Verzerrung > 0

Der erste und zweite Zyklus wird immer abgearbeitet. Der zweite allerdings nur dann, wenn im Ergebnis des 1.Zyklus gleichzeitig gesicherte **und** ungesicherte Koeffizienten vorliegen

Die Nummer des Zyklus, dessen Modellergebnis abgespeichert werden soll, kann mit den Steuereingabedaten festgelegt werden.

Zu den Voraussetzungen für die Anwendung der Methode der linearen Regression, dem mathematisch-statistischem Vorgehen und der Bewertung der berechneten Ergebnisse siehe unter der Aufgabenstellung zum Verarbeitungsprogramm Baustein POLMOD. Das dort gesagte gilt auch für den rein linearen Fall.

LIREV behandelt mathematisch gesehen, den in den unabhängigen Variablen linearen Spezialfall von POLMOD.

Eingabedateien:

- **unterprojekt_nqd.txt** geordnete Zwischendatei im Textformat
- **unterprojekt_mtx.txt** als Datenmatrix.
- **unterprojekt_std.txt** Statistik-Steuerdatei mit von NORM berechneten statistische Kenngrößen des Datenbestandes:
Mittelwert, Streuung, kleinster und größter Wert der Variablen, einfache Korrelationskoeffizienten zwischen den Variablen.'
- **unterprojekt_abd.txt** Projektsteuerdatei

Parameter, Grenzen:

- Es werden maximal 50 Variable/Messgrößen in einem Modellansatz verarbeitet.
- Die Anzahl der Datensätze ist faktisch unbegrenzt.
- Fehlstellen bei Messwerten/Einzeldaten werden berücksichtigt

Steuerparameter

statistische Sicherheit	99.8 , 99.0 , 95.0 , 90.0 , ohne
maximaler Grad der unabhängigen Variablen	1
Anzahl der Variablen in der Auswahl	100
Maximale Anzahl der Variablen im linearen Modell	50

Ergebnisdateien:

- **modellname_eg0.txt** Protokolldatei der Modellbildung
- **modellname_mod.txt** Darstellung des gebildeten Modells
- **modellname_tsx.txt** Zwischendatei berechnet - gemessen
- **modellname_sw1.txt** Ergebnis des Vergleiches berechnet – gemessen, Eingabedatei für die grafische Darstellung
- **modellname_bg.typ** Grafikdatei berechnet - gemessen
- **lirev_mld.txt** Programmnachrichten(Ablauf, Fehler, ...)

Aufbau der Datei **projekt_abd.txt** für den Baustein LIREV:

Steuermerkmal	Spalte	Typ	Parameter	Inhalt des Parameters
<Variablen >	01-03 05-07 09-11 ...	Int	ausvar(n)	Index der Modellvariablen, der letzte Index gehört der abhängigen Variablen
<Steuerparameter >	01-01	Int	itest	gewählte statist. Sicherheit
	03-04	Int	ilv	Anzahl Verzerrungsvarianten
	06-10	Real	svp	Schrittweite Verzerrung
	12-12	Char	ckoeff	Koeffizient zu Rechnung mit gesicherten Termen

DAPROA: Daten- und Prozessanalyse
Bausteine LIREV / LIREV_e

	14-14	Char	cprot	Protokollparameter
	16-16	Char	zmod	Modell abspeichern
	18-19	Int	imod	Zyklusnummer Modell speichern
	21-28	Char	modnam	Name Modelldatei
	30-30	Char	cbrgm	berechnet-gemessen vornehmen
	32-39	Char	zugrd	Namenszusatz Grafikdatei
	41-43	Int	jsvx	Index der unabhängigen Variablen für Grafikdarstellung, Abzisse
	45-52	Char	uprnam	Unterprojektname
	54-66	Real	ugvx	Untergrenze der Abszissenvariablen
	68-80	Real	ogvx	Obergrenze der Abszissenvariablen
	82-84	Int	jxsch	Anzahl Variationsschritte

Das Bausteinmerkmal <lirev > muss genau 10 Zeichen lang sein (Leerstellen und spitze Klammern mitgezählt) .

Die Steuermerkmale müssen genau 21 Zeichen lang sein (Leerstellen und spitze Klammern mitgezählt) .

Sie müssen, wie in Beispiel und Aufbaubeschreibung dargestellt, geschrieben, also auch mit der entspr. Anzahl von Leerzeichen aufgefüllt werden.

Dialogbaustein LIREV_e

In dem Dialogbaustein werden die Parameter für die Projektarbeitungsdatei **projekt_abd.txt** abgefragt, eingegeben und geprüft.

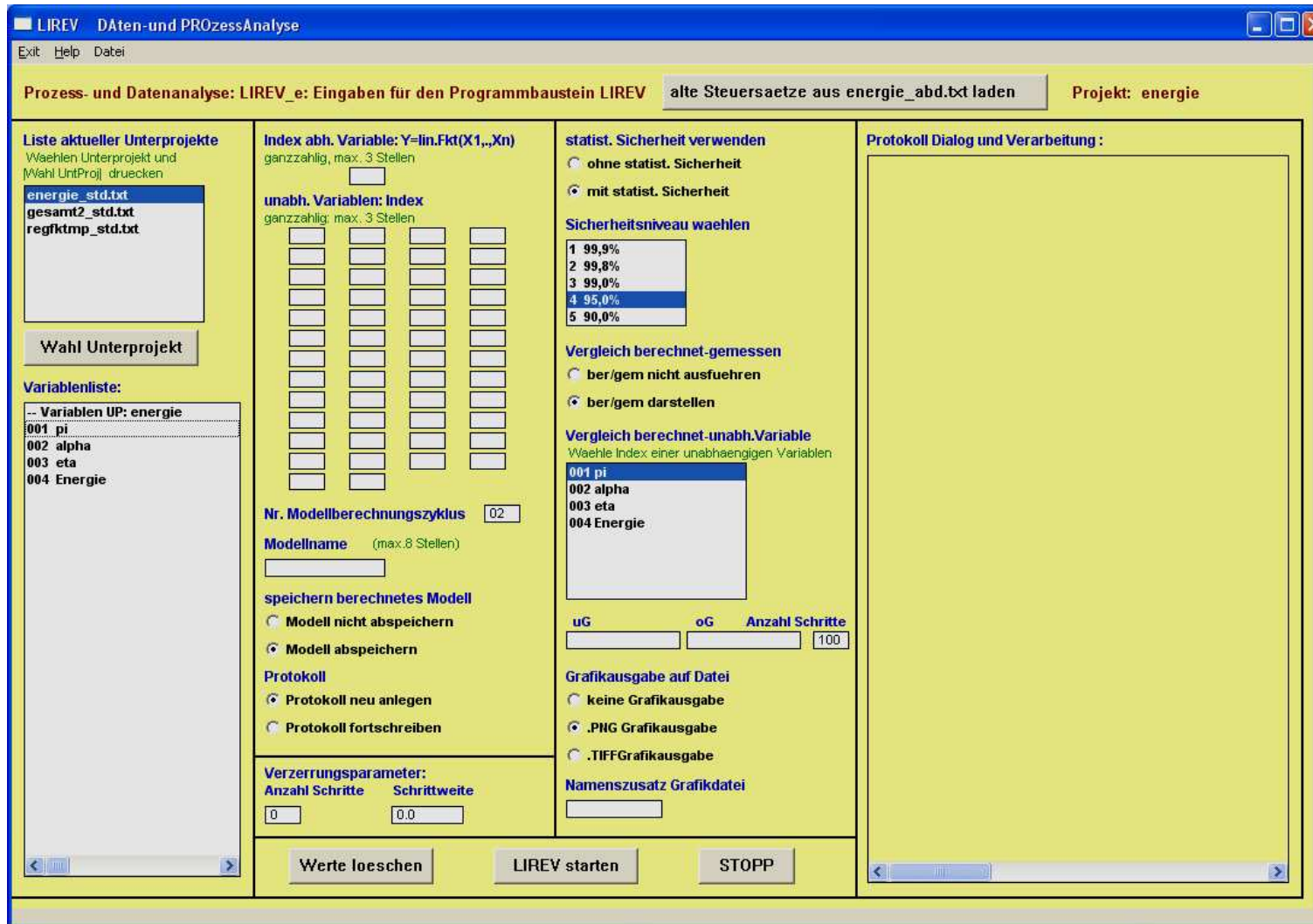


Bild 1:
Startmaske des
Dialogbausteins LIREV_e

Mit dem Dialogbaustein LIREV_e werden die Struktur des linearen Modells, die Steuergrößen für die Abarbeitung und die Parameter zur Ergebnisausgabe festgelegt.

Wahl des Unterprojektes

In der Liste |**Liste aktueller Unterprojekte**| kann ein Unterprojekt ausgewählt werden. Die Auswahl ist mit der Taste **Wahl Unterprojekt** zu bestätigen.

In der darunter stehenden |**Variablenliste**| werden nach der Bestätigung die Variablen des Unterprojektes angezeigt. Am Anfang der Variablenliste wird das ausgewählte und zugehörige Unterprojekt genannt.

Beim Start von LIREV_e wird das Originalprojekt vorgelegt, es wird mit seinen Variablen markiert und angezeigt.

Mit der Taste **alte Steuersaetze aus projekt_abd.txt laden** können die alten Steuersätze geladen und in den Dialogmasken vorgelegt werden, soweit welche vorhanden sind.

Aufbau des linearen Modells

Im Eingabefeld [**Index abh. Variable: Y=lin.Fkt(X1,..,Xn):**] wird die Indexnummer der abhängigen Variablen eingetragen, die in der Liste |**Variablenliste:**| aufgeführt ist.

In den Eingabefeldern[**unabh. Variablen: Index**] werden die Indexnummern der unabhängigen Variablen(Einflussgrößen) eingetragen, die in der Liste |**Variablenliste:**| aufgeführt sind.

Im Eingabefeld [**Nr. Modellberechnungszyklus**] wird der Berechnungszyklus angegeben, dessen Modellergebnis abgespeichert wird. Es gibt maximal 12 Zyklen.

Der Wert 2 wird als Standard vorgelegt.

Es kann passieren, dass der Wert 2 eingetragen ist, aber der zweite Zyklus nicht abgearbeitet wird.

Gründe können in den Termen liegen, welche im ersten Zyklus berechnet wurden:

- alle Terme sind statistisch gesichert
- alle Terme sind statistisch nicht gesichert

Wenn in einem solchen Fall aber als Zyklusnummer, dessen Modellergebnis abgespeichert werden soll, die zwei gesetzt wurde, wird u.U. je nach Parameterlage das im 1.Zyklus berechnete Modell abgespeichert. Eine entsprechende Meldung erscheint im Dialog bzw. steht in der Datei **lirev_mld.txt**.

Im Eingabefeld [[Modellname\(max.8 Stellen\)](#)] wird der Name der Modell- und Ergebnisdateien festgelegt. Somit können zu einem Projekt verschiedene Modellvarianten dokumentiert werden.

Steuergrößen für die Abarbeitung

In den folgenden Auswahlboxen werden Steuergrößen für die Abarbeitung und Parameter zur Ergebnisausgabe festgelegt.

°speichern berechnetes Modell°

"Modell nicht abspeichern" oder
"Modell abspeichern"

°Protokoll°

"Protokoll neu anlegen" oder
"Protokoll fortschreiben"

°statistische Sicherheit verwenden°

"ohne statist. Sicherheit" oder
"mit statist. Sicherheit"

In der Liste [[Sicherheitsniveau waehlen](#)] wird des Niveau der statistischen Sicherheit für die Signifikanzprüfung der linearen Terme ausgewählt

°Vergleich berechnet-gemessen°

"ber/gem nicht ausfuehren" oder
"ber/gem darstellen"

Falls "ber/gem darstellen" gewählt wurde kann in der Variablenliste

[[Vergleich berechnet-unabh.Variable](#)] der Index eine Einflussgröße (unabh. Variable) und ihr Index markiert werden. Diese Variable muss Modellbestandteil und darf nicht die abhängige Variable sein.

Nach der erfolgreichen Markierung werden der Index und die Minimal- bzw. Maximalwert dieser Variablen aus den Datensätzen angezeigt.

In den Eingabefeldern [[uG](#)] , [[oG](#)] und [[Anzahl Schritte](#)] können die Werte verändert werden.

Diese Werte dienen zur Darstellung der Grafik "berechnet/gemessen gegen die gewählte unabhängige Variable".

In der Grafik berechnet-gemessen werden diese Werte über die eingetragene unabhängige Variable als Abszisse dargestellt.

°Grafikausgabe°

"keine Grafikausgabe" oder

"Grafikausgabe" auf einen genannten Dateityp

Im Eingabefeld [[Namenszusatz Grafikdatei](#)] kann ein maximal 8-stelliger Namenszusatz für die Grafikausgabedatei gewählt werden. Zu einem Modell können so mehrere Grafiken aufbewahrt werden.

Im Eingabefeld [[Verzerrungsparameter Anzahl der Schritte](#)] wird eine Anzahl von 0 bis 10 eingetragen. Der Wert 0 wird als Standard vorgelegt.

Im Eingabefeld [[Verzerrungsparameter Weite](#)] wird die Schrittweite für die Verzerrung eingetragen. Der Wert 0.0 wird als Standard vorgelegt.

Beispiel: Anzahl der Schritte = 4, Schrittweite 0.1

Es wird die Verzerrung mit den Parametern 0.1 , 0.2 , 0.3 und 0.4 durchgeführt.

Mit der Taste [Werte loeschen](#) wird die Eingabewerte auf der Dialogmaske in ihren "Leerzustand" zurückgesetzt

Mit der Taste [STOPP](#) wird der Dialog LIREV_e beendet.

Mit der Taste [LIREV starten](#) wird der Verarbeitungsbaustein LIREV ausgeführt.

Bevor LIREV wirklich startet, werden im Dialogfenster rechts die gewählten Parameter/Eingabewerte angezeigt. Sie müssen bestätigt werden. Bei Nichtbestätigung wird zum Eingabedialog zurückgekehrt.

Bei Fehlern während der Verarbeitung im Baustein LIREV bricht der Vorgang ab. Der Fehler wird in der Datei "[lirev_mld.txt](#)" protokolliert.

Nach Beendigung von LIREV kehrt die Steuerung zum Dialogbaustein LIREV_e zurück. Der Inhalt der Datei "[lirev_mld.txt](#)" , das Abarbeitungsprotokoll, wird nach der Beendigung von LIREV im Dialogfenster rechts angezeigt, hier stehen auch die Fehlermeldungen, falls LIREV wegen auftretenden Fehlern abbricht.

Das Verarbeitungsprotokoll von LIREV befindet sich auch im Verzeichnis \meldung\

Beispiel: Es soll ein lineares Modell **Energie**= b0 + b1*pi + b2*alpha + b3*eta durch Schätzung der Koeffizienten b0 bis b3 der linear im Modellansatz auftretenden Variablen mittels Regressionsanalyse bestimmt werden.

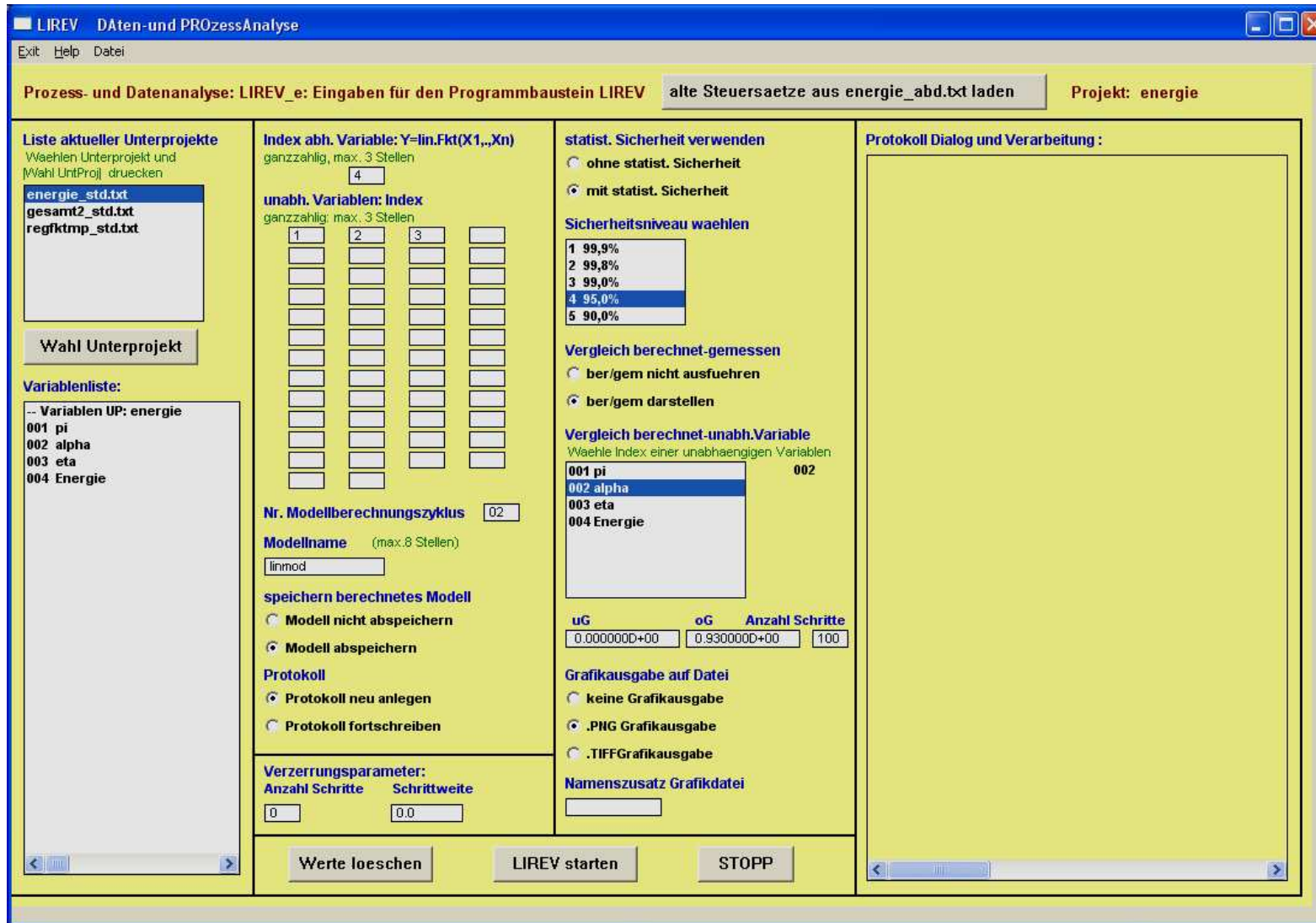


Bild 2:
Projekt energie ,
Unterprojekt energie

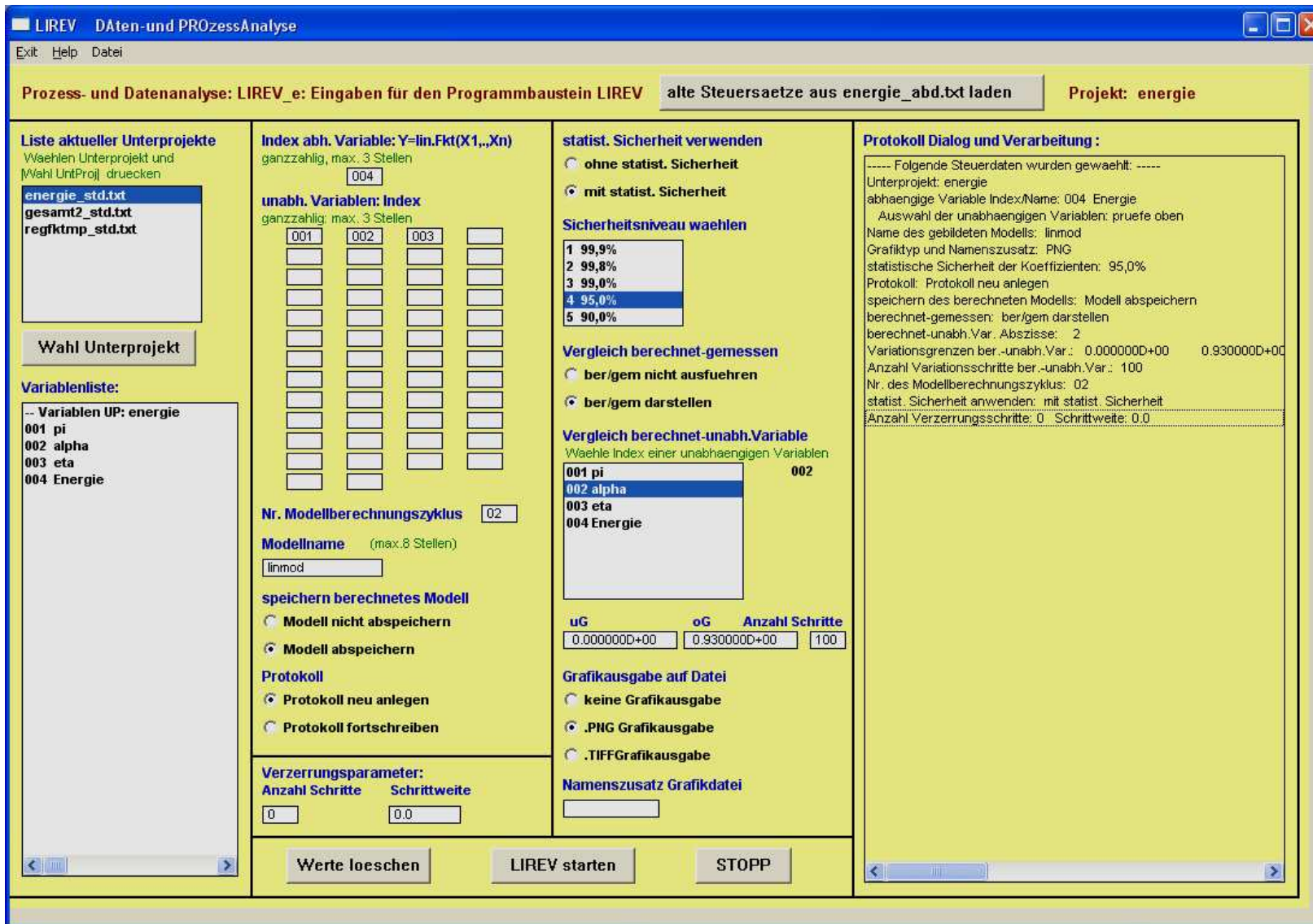


Bild 3:
Steuerparameter für
die Bearbeitung der
Aufgabe mit LIREV

- Im Projekt **energie** wurde das Unterprojekt **energie** mit der Datensteuerdatei **energie_std.txt** gewählt.
- Als abhängige Variable wurde die mit dem Index 4 (**Energie**) bestimmt
- Als unabhängige Variablen wurde die mit dem Indizes 1 , 2 , 3 (**pi, alpha, eta**) bestimmt
- Das Modell des Berechnungszyklus 2 wird unter dem Namen **linmod** gespeichert
- Der Modellkoeffiziententest wird mit einer statistischen Sicherheit von 95% durchgeführt.
- Die grafische Darstellung von berechneten gegen gemessene Werte wird gewählt.
- Dadurch muss eine unabhängige Modellvariable für die grafische Darstellung festgelegt werden, es ist "**alpha**" Index 2
- Die Grafik wird neben dem Bildschirm auch auf die Datei **linmod_bg.PNG** ausgegeben.

Nach betätigen der Taste **LIREV starten** erscheint die Abfrage



Falls ja, startet die Verarbeitung mit LIREV

Da die Grafikdarstellung berechnet/gemessen gewählt wurde, erscheint der Dialogbildschirm des Programmbausteins GRBEGED.

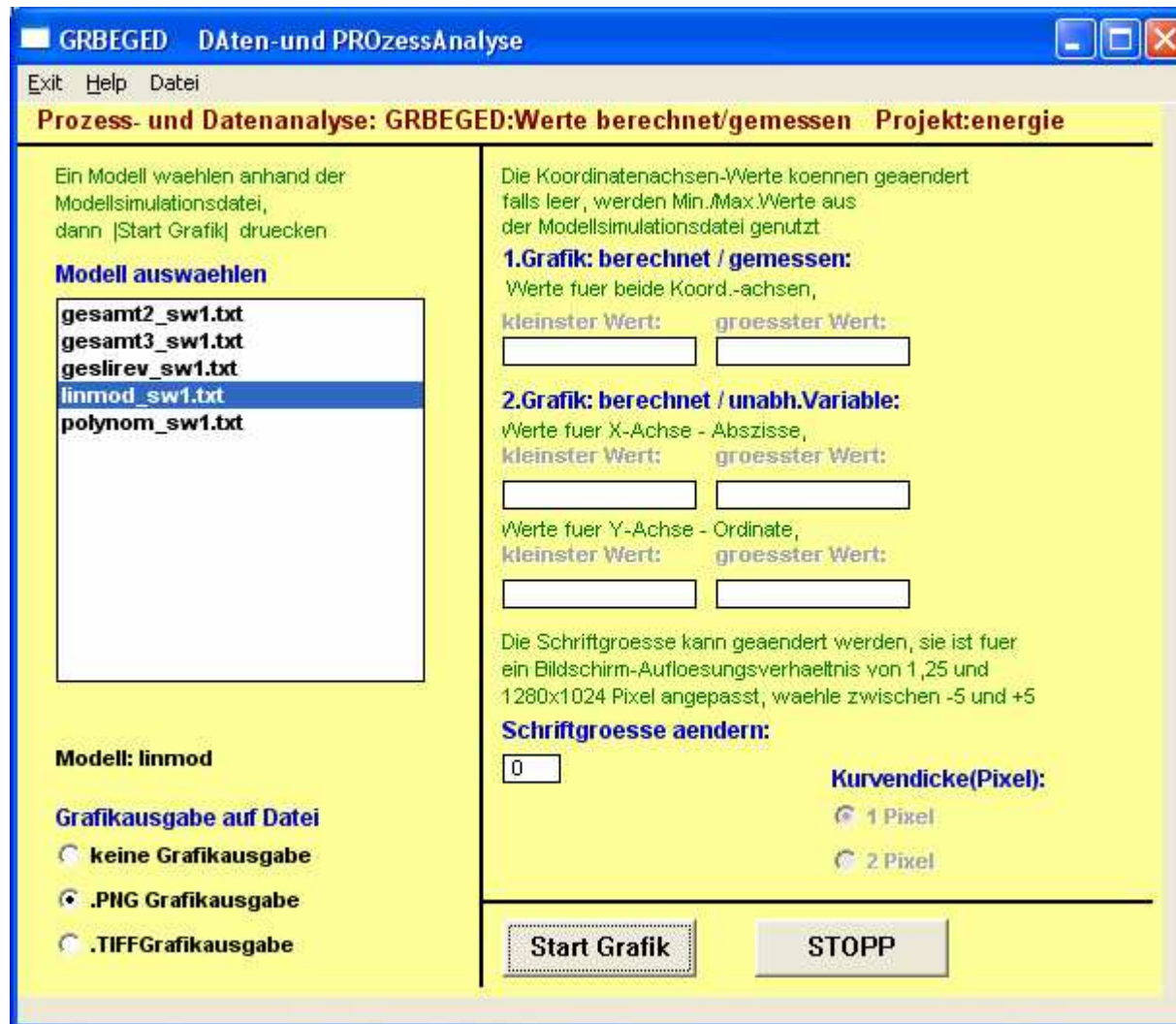


Bild 4:
Startbildschirm des Bausteins GRBEGED

Mit der Taste **Start Grafik** wird die Grafik angefordert. Die Grafik erscheint im eigenen Fenster und kann mit den Tastaturlisten [Alt][Druck] in den systeminternen Speicher übernommen werden. Die Grafik wird geschlossen mit [Enter] oder [rechte Maustaste].
Nachstehend die Grafik berechnete Werte gegen gemessene Werte

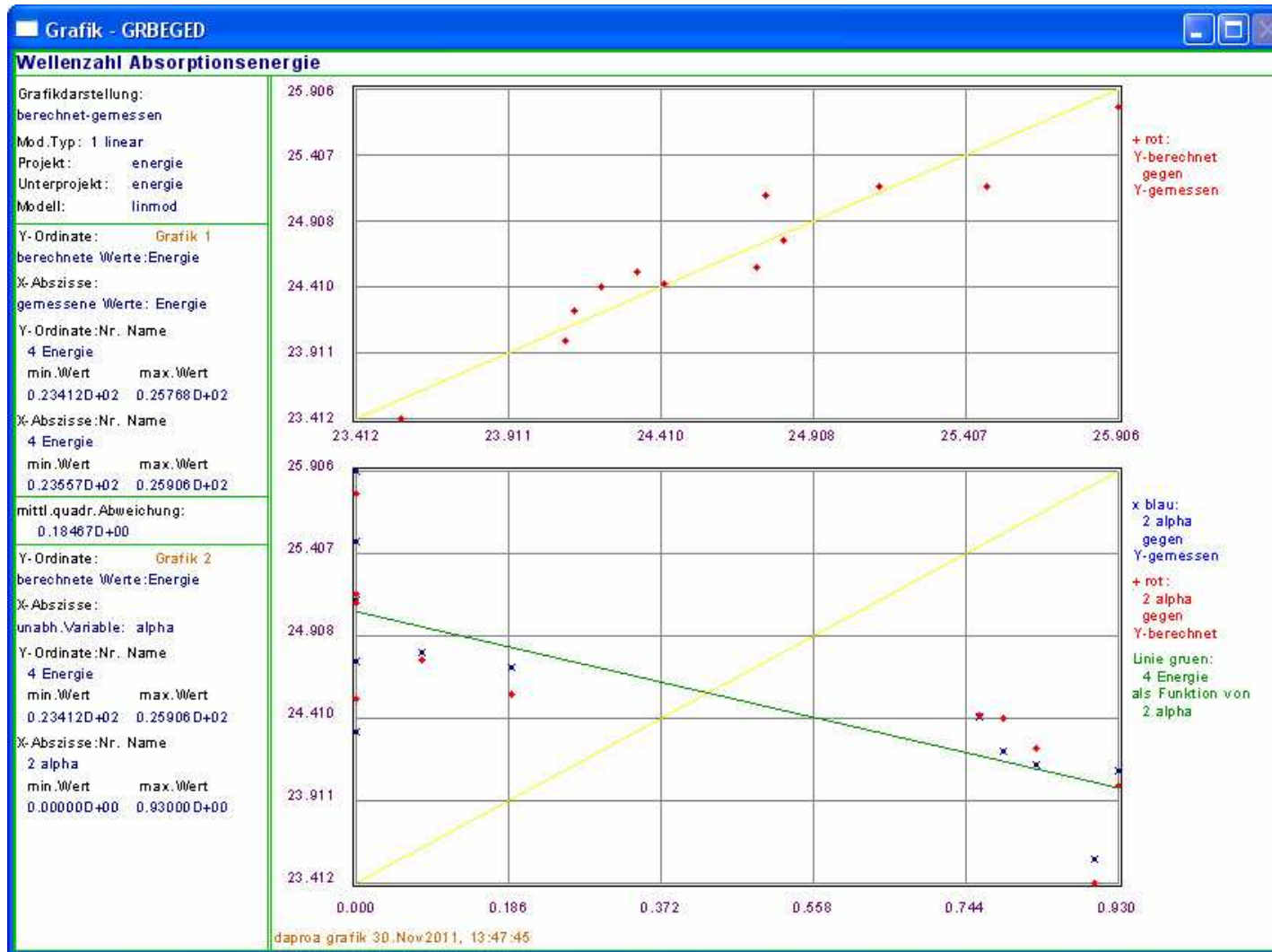


Bild 5:
Darstellung von
berechnete gegen gemessene
Werte von **Energie**
und
Energie als Funktion von **alpha**

1. Koordinatensystem
rot: **Energie** gemessen
gegen berechnet

2. Koordinatensystem
rot: **alpha** gegen
Energie berechnet
blau: **alpha** gegen
Energie gemessen
grün: **Energie** als funktion
von **alpha**

gelb: Diagonale im Koord.-system

Nach der Grafik in Bild 5, erscheint noch einmal das Dialogfenster von GRBEGED. Über die Eingabefelder kann die Darstellung der Grafik verändert werden.

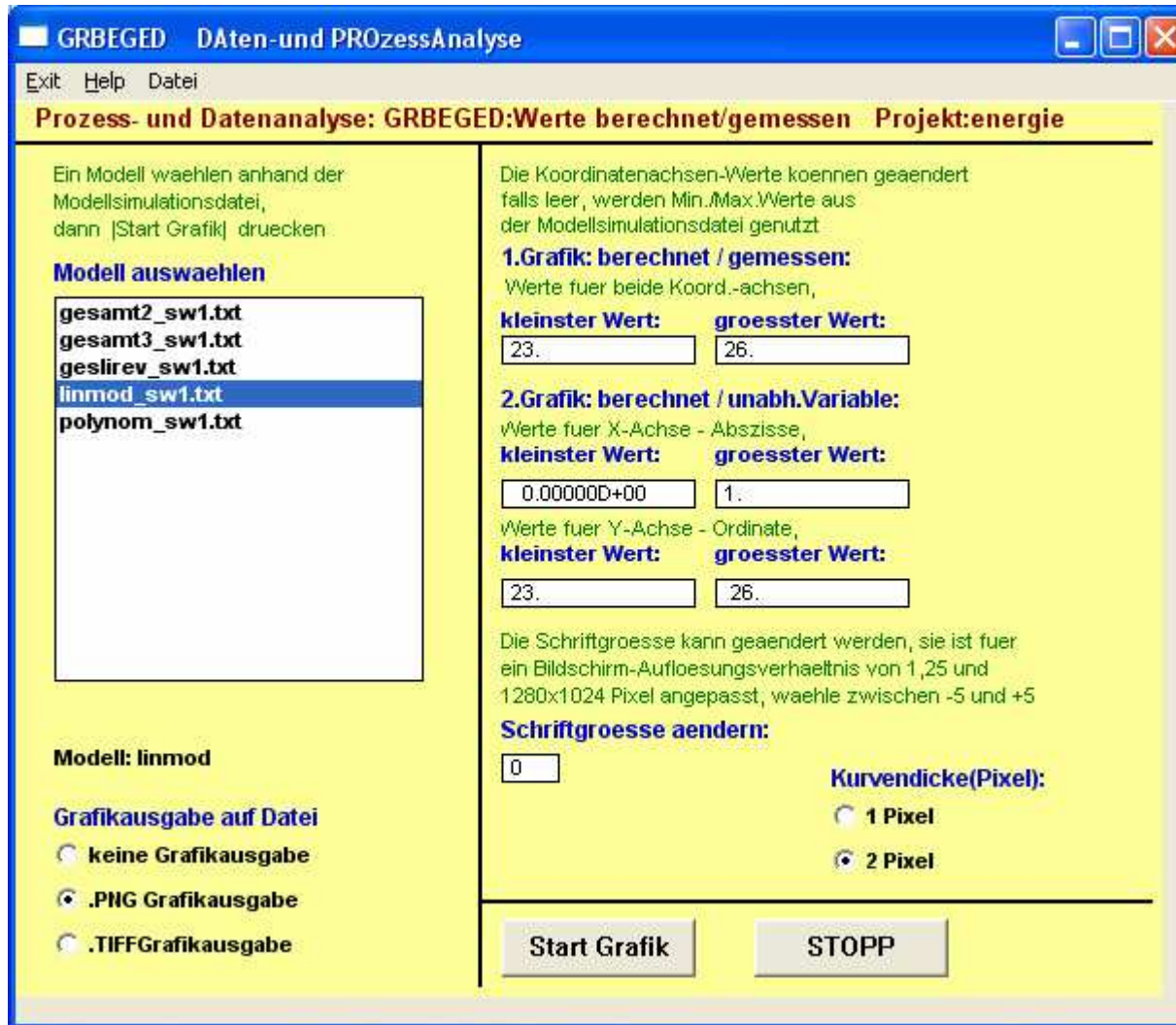


Bild 6:
Im Eingabefenster von GRBEGED können nun die Gestaltungsparameter der Grafik verändert werden.

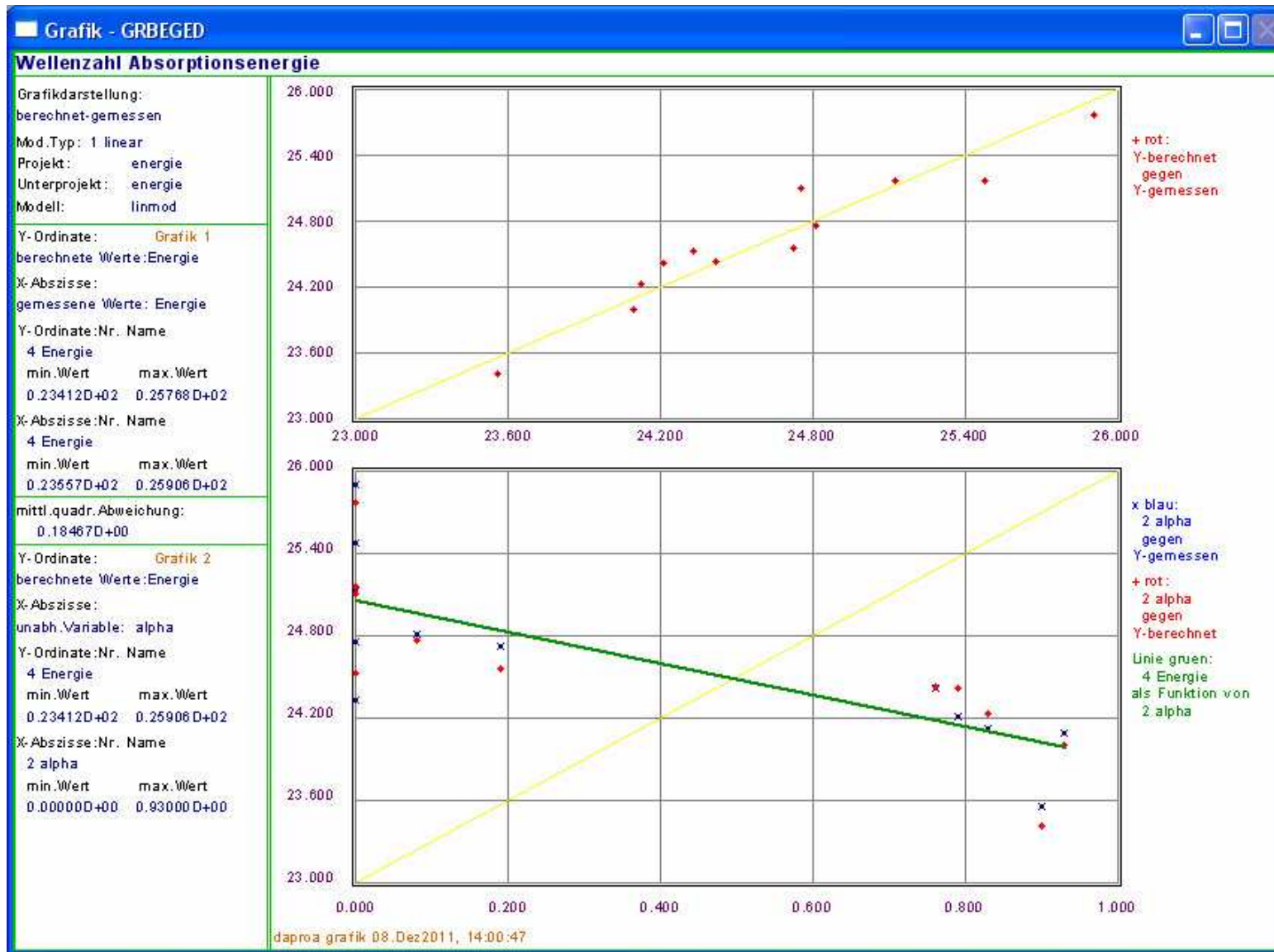


Bild 7:
Darstellung von
berechnete gegen
gemessene Werte
von **Energie**
und
Energie als Funktion von
alpha

1. Koordinatensystem
rot: **Energie** gemessen
gegen berechnet

2. Koordinatensystem
rot: **alpha** gegen
Energie berechnet

blau: **alpha** gegen
Energie gemessen
grün: **Energie** als funktion
von **alpha**

gelb: Diagonale im Koord.-
system

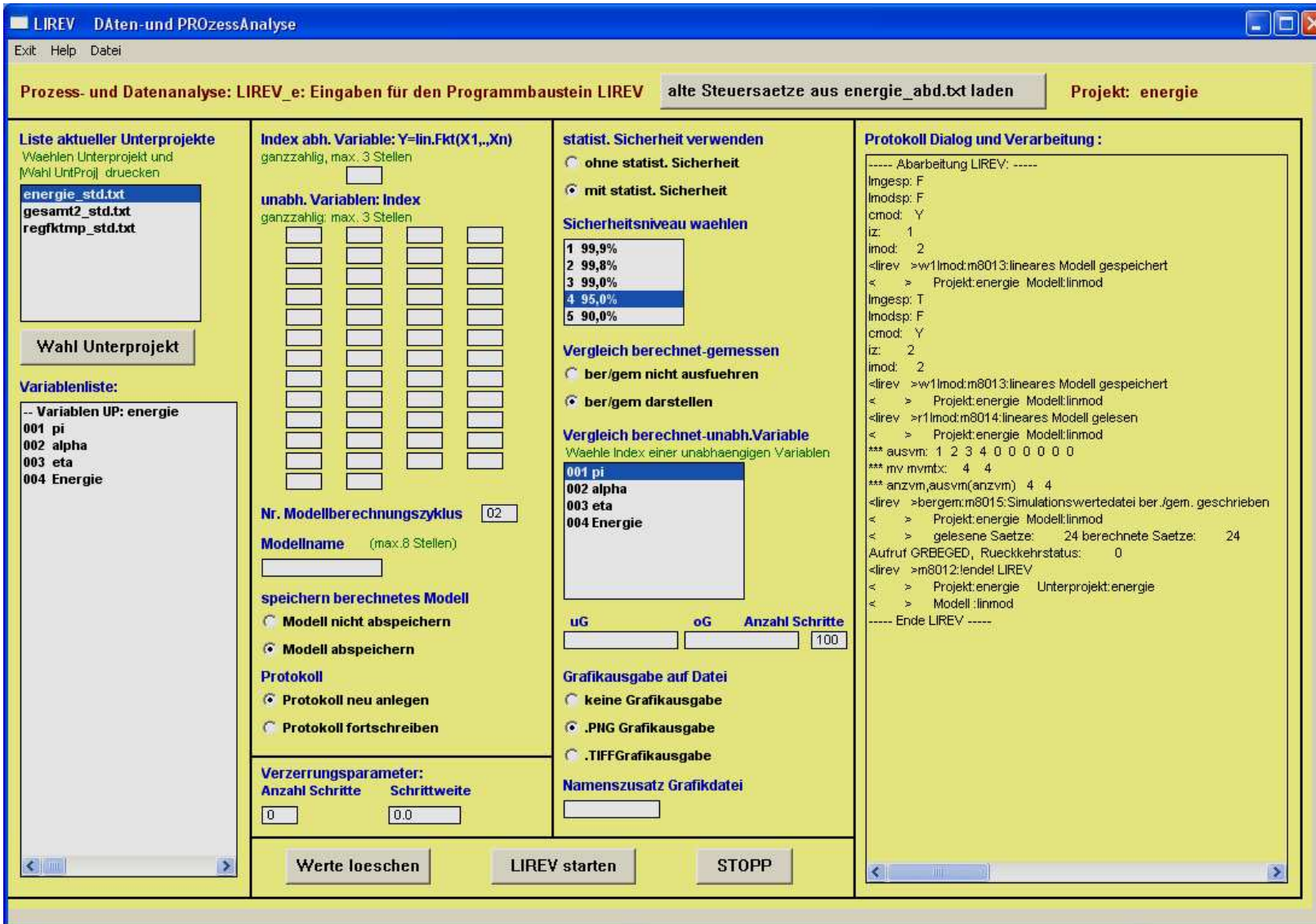


Bild 8:
Nach Beenden von
GRBEGED wird
das LIREV-
Dialogfenster mit
dem Protokoll
aufgeblendet

Nach dem Beenden des Bausteins GRBEGED kehrt die Verarbeitung zum Dialogfenster von LIREV zurück. Im rechten Teil des Fensters erscheint ein Verarbeitungsprotokoll

Im Verzeichnis ..\daproa\Ergebnis\energie sind die Ergebnisse zu finden.

linmod_eg0.txt Verarbeitungsprotokoll mit detailliertem Ergebnis
linmod_mod.txt standardisierte Modelldarstellung, sie dient anderen
 Programmbausteinen als Eingabedatei (Simulation, grafische Darstellungen)
linmod_sw1.txt Datenmatrix für die Darstellung berechnet - gemessen
linmod_bg.png Kopie der Grafikdatei berechnet – gemessen

Die Datei mit der standardisierten Modelldarstellung **linmod_mod.txt** darf nicht verändert werden, da damit auch das Modellergebnis verändert wird.

Es sei denn, man will ein bestimmtes Modell für Simulation, grafische Darstellungen usw. bereitstellen.

Im Verzeichnis ..\daproa\Daten\energie ist in der Projektsteuerdateidatei

Energie_abd.txt
der benutzte Parametersteuersatz für LIREV gespeichert.

```
<lirev    ><Variablen                    >001 002 003 004  
<lirev    ><Steuerparameter            >4 0   0.0    1 N Y 02 linmod    Y 090401    2   energie
```


Nachstehend werden das Ergebnisprotokoll und das berechnete Modell gezeigt.

Ergebnisprotokoll **linmod_eg0.txt**

Programm:lirev Projekt:energie UP:energie 08.Jan2010 21:26:43

Titel:Wellenzahl Absorptionsenergie
unabhaengige Variable: Energie

Eingabewerte:

Schritt: 1 Regression mit allen Einflussgrößen

Anzahl Messwerte: 24
Anzahl Einflussgrößen: 3
Anzahl Modellterme: 3
Freiheitsgrad: 20
statistische Sicherheit: 95.0%
Verzerrungsfaktor: 0.000

Ergebnisse:

Testwert t-Verteilung: 2.086
Bestimmtheitsmass: 0.91782
Reststandardabweichung: 0.19541D+00

Modelltyp: linear

Ind	abhng.Variable	Koeffizient	unabh.Variable	t-Testwert	Niveau	unten	Konfidenzinterv.oben
4	Energie	= 0.26203D+02					
1		-0.19764D+01	* pi	-0.87065D+01	95.0%	-0.24499D+01	-0.15029D+01
2		-0.12189D+01	* alpha	-0.87780D+01	95.0%	-0.15085D+01	-0.92923D+00
3		0.90734D-01	* eta	0.73538D+00		-0.16664D+00	0.34811D+00

Titel:Wellenzahl Absorptionsenergie
unabhaengige Variable: Energie

Eingabewerte:

Schritt: 2 Regression nur mit gesicherten Einflussgrößen

Anzahl Messwerte: 24
Anzahl Einflussgrößen: 2
Anzahl Modellterme: 2
Freiheitsgrad: 21
statistische Sicherheit: 95.0%
Verzerrungsfaktor: 0.000

Ergebnisse:

Testwert t-Verteilung: 2.080
Bestimmtheitsmass: 0.91560
Reststandardabweichung: 0.19326D+00

Modelltyp: linear

Ind	abhng.Variable	Koeffizient	unabh.Variable	t-Testwert	Niveau	unten	Konfidenzinterv.oben
4	Energie	= 0.26235D+02					
1		-0.19459D+01	* pi	-0.88157D+01	95.0%	-0.24050D+01	-0.14868D+01
2		-0.11477D+01	* alpha	-0.11659D+02	95.0%	-0.13524D+01	-0.94294D+00

Der Einfluss der unabhängigen Variablen eta (Index 3) erreicht in dem gestellten linearen Modellansatz nicht das geforderte Signifikanzniveau von 95%. Sie taucht daher nach dem zweiten Berechnungsschritt nicht mehr im Modell auf.

Das berechnete Modell in linmod_mod.txt

```
<Modell >linear
<IndModTyp > 1
<DatumZeit >08.Jan2010 21:26:43
<QueProjekt>energie
<QueModell >linmod
<QueMatrix >energie
<Titel >Wellenzahl Absorptionsenergie
<ModParKoeff> 0 Energie          4  0.26235D+02  0  4  0  0  0  0
<ModParKoeff> 1 pi              1 -0.19459D+01  1  1  0  0  0  0
<ModParKoeff> 2 alpha           2 -0.11477D+01  1  2  0  0  0  0
<StandAbw > 0.19326D+00
<BestMass > 0.91560
<StatSich >95.0%
<AnzMesswrt> 24
<AnzVarMtx > 4
<AnzEG > 2
<FrGrad > 21
<Wert-tTest> 2.080
<VerzPar > 0.000
<VarAbh > 4 Energie          0.23557D+02  0.25906D+02  0.24628D+02  0.63566D+00
<VarUnabh >X 1 pi            0.24000D+00  0.92000D+00  0.60583D+00  0.18294D+00
<VarUnabh >X 2 alpha         0.00000D+00  0.93000D+00  0.37333D+00  0.41025D+00
<VarUnabh > 3 eta           0.41000D+00  0.20000D+01  0.84667D+00  0.46801D+00
```

Das Modell noch einmal extra dargestellt:

$$\begin{aligned} \text{Energie} &= 0.26235D+02 - 0.19459D+01 * \text{pi} - 0.11477D+01 * \text{alpha} \\ &= 26.235 - 1.9459 * \text{pi} - 1.1477 * \text{alpha} \end{aligned}$$

KURZANLEITUNG

- (1) Übernahme des vorgelegten Unterprojektes (dazu ist keine Aktion nötig)
oder
Auswahl eines anderen Unterprojektes aus der Liste
|Liste aktueller Unterprojekte| und bestätigen mit der Taste **Wahl Unterprojekt**
- (2) In den Eingabefeldern [**Index abh. Variable: $Y=lin.Fkt(X1,..,Xn)$**] und
[**unabh. Variablen: Index**] sind die Indizes der Modellvariablen einzutragen.
- (3) In das Eingabefeld [**Nr. Modellberechnungszyklus**] wird der Zyklus eingetragen bei dem das gebildete Modell übernommen wird.
- (4) In das Eingabefeld [**Modellname**] einen Modellnamen eintragen. Unter diesen Namen werden das gebildete Modell und andere Ergebnisdateien abgespeichert. Ist dieses Feld nicht ausgefüllt, wird "mdnoname" als Name vergeben.
- (5) Nun die übrigen Verarbeitungsparameter festlegen. Dabei sind logische Zusammenhänge zu beachten. So muss das gebildete Modell abgespeichert werden, wenn der Vergleich berechnet/gemessen dargestellt werden soll.
- (6) Ist der Vergleich berechnet/ gemessen gewählt worden, dann in der Variablenliste |**Vergleich berechnet-unabh.Variable** | die unabhängige Modellvariable markieren, welche als Abszisse in der grafischen Darstellung dient. Diese Variable muss Modellbestandteil und darf nicht die abhängige Variable sein.
- (7) Wurde die Markierung vom Programm akzeptiert können in den Eingabefeldern [**uG**] , [**oG**] und [**Anzahl Schritte**] die Werte der gewählten unabhängigen Modellvariablen für die grafische Darstellung verändert werden.
- (8) Start des Bausteins LIREV mit der Taste **LIREV starten**