

NETI – Network Inference

Version 1.2

User Manual

Copyright (C) 2004-2007 Institute Curie. All rights reserved.

NETI download page: <http://bioinfo.curie.fr/projects/reverse-engineering/>

Author(s): Eugene Novikov (Institut Curie)

E-mail: Eugene.Novikov@curie.fr

System Requirements

NETI is written in Java and C++. It runs on Windows platforms 95/98/Me/NT/2000/XP and needs the Java Runtime Environment (JRE) to be installed:

<http://www.java.com/en/download/>

Installation

NETI can be downloaded from the NETI download page <http://bioinfo.curie.fr/projects/reverse-engineering/>

Click NETI Setup 1.2.exe to start the NETI 1.2 installer and follow the instructions*.

NETI 1.2 installation creates a “Curie/NETI 1.2” folder in the list of Programs of the Windows Start menu. This new folder contains the following entries:

- NETI 1.2 starts Network inference software;
- User Manual is a user manual pdf file;
- Uninstall NETI will remove NETI from your computer.

Installation procedure may also create a “NETI 1.2” icon on your Desktop.

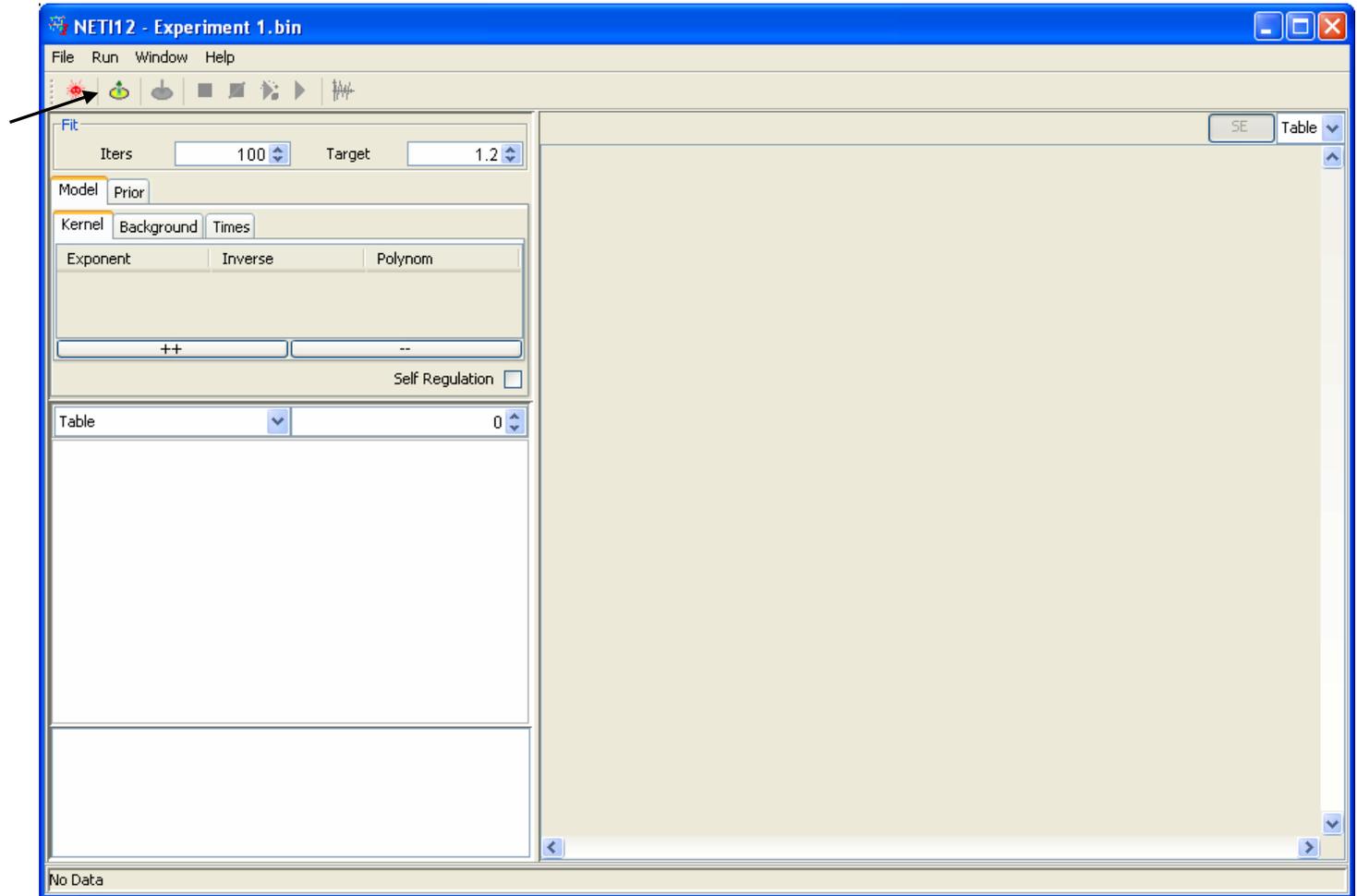
*) Installation procedure asks about the default size of the JVM (Java Virtual Machine) memory allocation pool. It is recommended to set it as large as possible, but not larger than the amount of available RAM.

Network Inference

Processing Window

Data can be downloaded using the “Load Data ...” button from the Toolbar or the Menu Item “File|Load|Data ...” (Ctrl+O).

See next page for examples of the input data formats.



Input Data Format

Without standard errors:

Header	Time	x0	x1	x2	x3	x4	x5	...
Sample	:	0	1.0101	2.0202	3.0303	4.0404	5.05051	...
E1	0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	...
E2	1	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	...
KKK	2	0.003	0.00296	0.00291	0.00287	0.00284	0.0028	...
P_KKK	3	0	0.00001	0.00002	0.00003	0.00003	0.00004	...
KK	4	1.2	1.19743	1.1904	1.17974	1.16613	1.15009	...
P_KK	5	0	0.00253	0.00948	0.01994	0.03313	0.0484	...
PP_KK	6	0	0	0.00001	0.00004	0.00012	0.00026	...
K	7	1.2	1.19999	1.19985	1.19931	1.19795	1.19533	...
P_K	8	0	0.00001	0.00008	0.00038	0.00114	0.00259	...
PP_K	9	0	0	0	0	0	0.00001	...
KPase	10	0.12	0.12	0.11997	0.11985	0.11955	0.11897	...
P_KKK_KK	11	0	0.00003	0.00007	0.0001	0.00013	0.00015	...
PP_KK_KK	12	0	0	0.00003	0.00016	0.00046	0.00104	...
KPase_PP_KK	13	0	0	0	0	0	0	...
...

With standard errors (SE is indicated in the table cell [1;1]):

Header	Time	x2	SE x2	x10	SE x10	x13	SE x13	...
Sample	SE;	2.0202	2.0202	10.10101	10.10101	13.13131	13.13131	...
E1	0	3.02E-05	1.50E-06	3.00E-05	1.50E-06	3.15E-05	1.50E-06	...
E2	1	2.79E-04	1.50E-05	2.87E-04	1.50E-05	3.18E-04	1.50E-05	...
KKK	2	0.00286195	1.45E-04	0.00250212	1.32E-04	0.00260262	1.28E-04	...
P_KKK	3	2.14E-05	1.00E-06	6.97E-05	3.50E-06	9.56E-05	4.50E-06	...
KK	4	1.21922989	0.05952	1.06178469	0.052231	1.00851474	0.048417	...
P_KK	5	0.00993045	4.74E-04	0.13696104	0.007016	0.18396007	0.0098425	...
PP_KK	6	9.74E-06	5.00E-07	0.00265924	1.49E-04	0.00669657	3.58E-04	...
K	7	1.18169628	0.0599925	1.14199093	0.057234	1.0437543	0.0529535	...
P_K	8	8.16E-05	4.00E-06	0.03169311	0.0015605	0.08431202	0.0040365	...
PP_K	9	0	0.01	8.49E-04	4.20E-05	0.00607996	3.06E-04	...
KPase	10	0.11832766	0.0059985	0.11307701	0.0054215	0.08518873	0.004654	...
P_KKK_KK	11	7.16E-05	3.50E-06	2.27E-04	1.25E-05	2.92E-04	1.45E-05	...
PP_KK_KK	12	2.96E-05	1.50E-06	0.0103351	5.69E-04	0.02333924	0.0012625	...
KPase_PP_KK	13	0	0.01	0	0.01	0	0.01	...
...

Table View

Intensity values with the corresponding standard errors (*SE*).

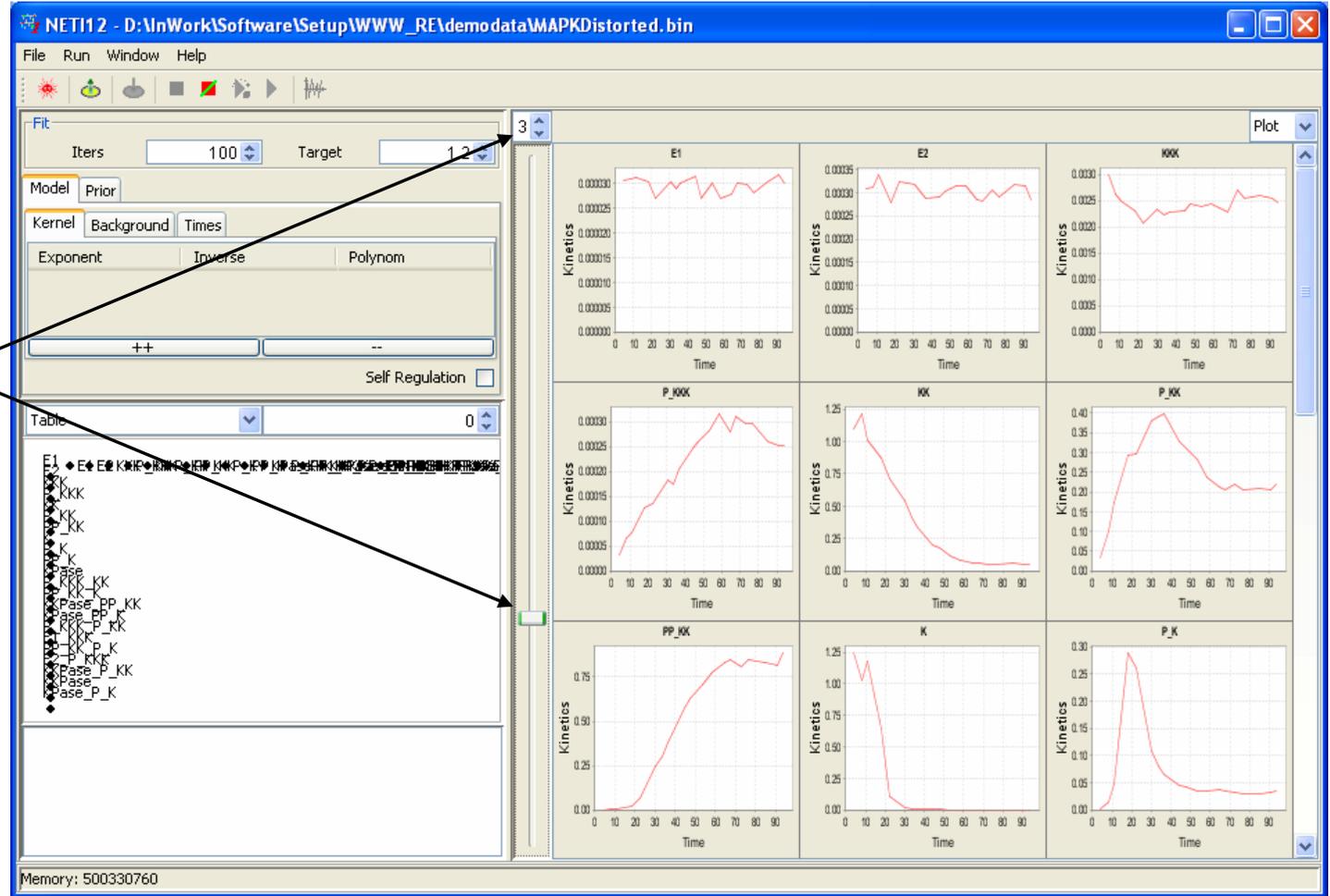
SE column can be hidden using the “*SE*” button from the local toolbar.

The screenshot shows the NETI 1.2 software window with the following table of parameters:

Header	Time	x4	SE x4	x8	SE x8	x11	SE x11
Time	SE	4.0404	4.0404	8.08081	8.08081	11.11111	11.1111
E1	0.0	3.85725260...	1.1E-6	3.09628184...	1.5E-6	3.12373743...	1.5E-6
E2	1.0	3.07532378...	1.49909999...	3.11788321...	1.49999999...	3.37839618...	1.49999
KKK	2.0	0.00298298...	1.42E-4	0.00262817...	1.35E-4	0.00248207...	1.305E-
P_KKK	3.0	3.36785888...	1.5E-6	6.44455290...	3.0E-6	7.81802572...	4.00000
KK	4.0	1.09694903...	0.05830650...	1.21132341...	0.05454449...	1.01075772...	0.05099
P_KK	5.0	0.03432677...	0.0016565	0.10067342...	0.0050945	0.17432846...	0.00797
PP_KK	6.0	1.30193776...	6.0E-6	0.00139286...	6.95E-5	0.00424773...	2.055E-
K	7.0	1.24208287...	0.05989750...	1.02269297...	0.05873800...	1.17772457...	0.05613
P_K	8.0	0.00106066...	5.7E-5	0.01391098...	7.06E-4	0.04633620...	0.00219
PP_K	9.0	0.0	1.0	1.63421023...	8.50000000...	0.00152726...	8.5E-5
KPase	10.0	0.12199318...	0.0059775	0.11166633...	0.0057275	0.11357458...	0.00520
P_KKK_KK	11.0	1.23317372...	6.5E-6	2.21276279...	1.10000000...	2.25559811...	1.35000
PP_KK_K	12.0	4.78586511...	2.30000000...	0.00514494...	2.72000000...	0.01561012...	7.68500
KKPase_PP...	13.0	0.0	1.0	0.0	1.0	0.0	1.0
KPase_PP_K	14.0	0.0	1.0	5.32821062...	3.0E-6	6.40613020...	2.95000
P_KKK_P_KK	15.0	0.0	1.0	2.11799427...	1.00000000...	3.71790448...	2.00000
E1_KKK	16.0	0.0	1.0	0.0	1.0	0.0	1.0
PP_KK_P_K	17.0	0.0	1.0	6.81570453...	3.5E-6	6.12453913...	2.99999
E2_P_KKK	18.0	0.0	1.0	0.0	1.0	0.0	1.0
KKPase_P_...	19.0	2.93667162...	1.5E-6	7.56427929...	4.00000000...	8.97975389...	5.0E-6
KKPase	20.0	2.85189337...	1.35000000...	2.18929163...	1.10000000...	1.84055799...	9.5E-6
KPase_P_K	21.0	4.30187251...	2.25E-5	0.00577453...	2.69000000...	0.01566717...	7.615E-

Plot View

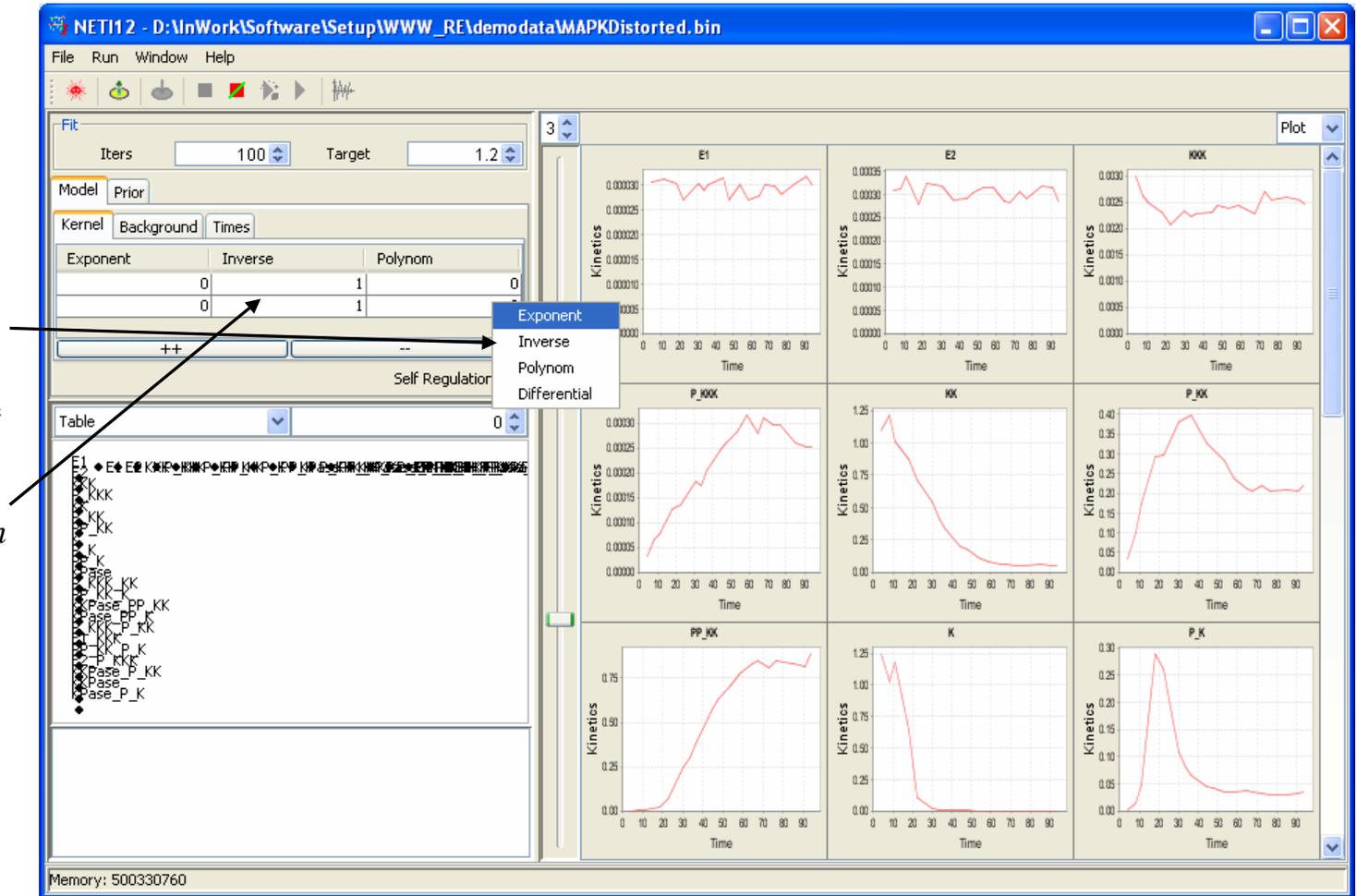
The number of the graph columns and the height of the graph rows can be adjusted using the corresponding controls.



Inference Model Definition: Inverse Model

Using the context menu of the Model table the default representation for the kernel function can be selected:

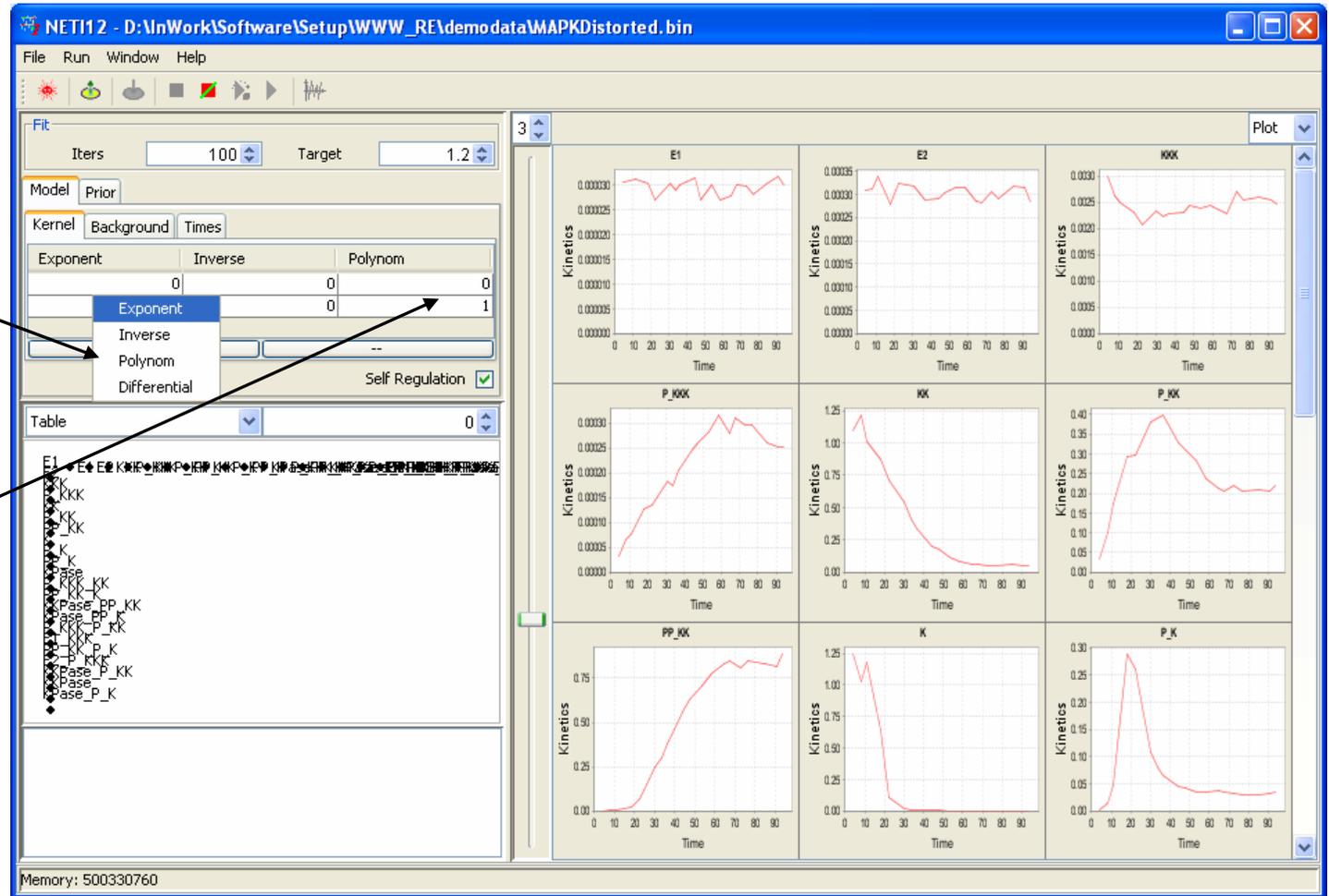
Inverse kernel function



Inference Model Definition: Polynomial Model

Using the popup menu of the Model table the default representation for the kernel function can be selected:

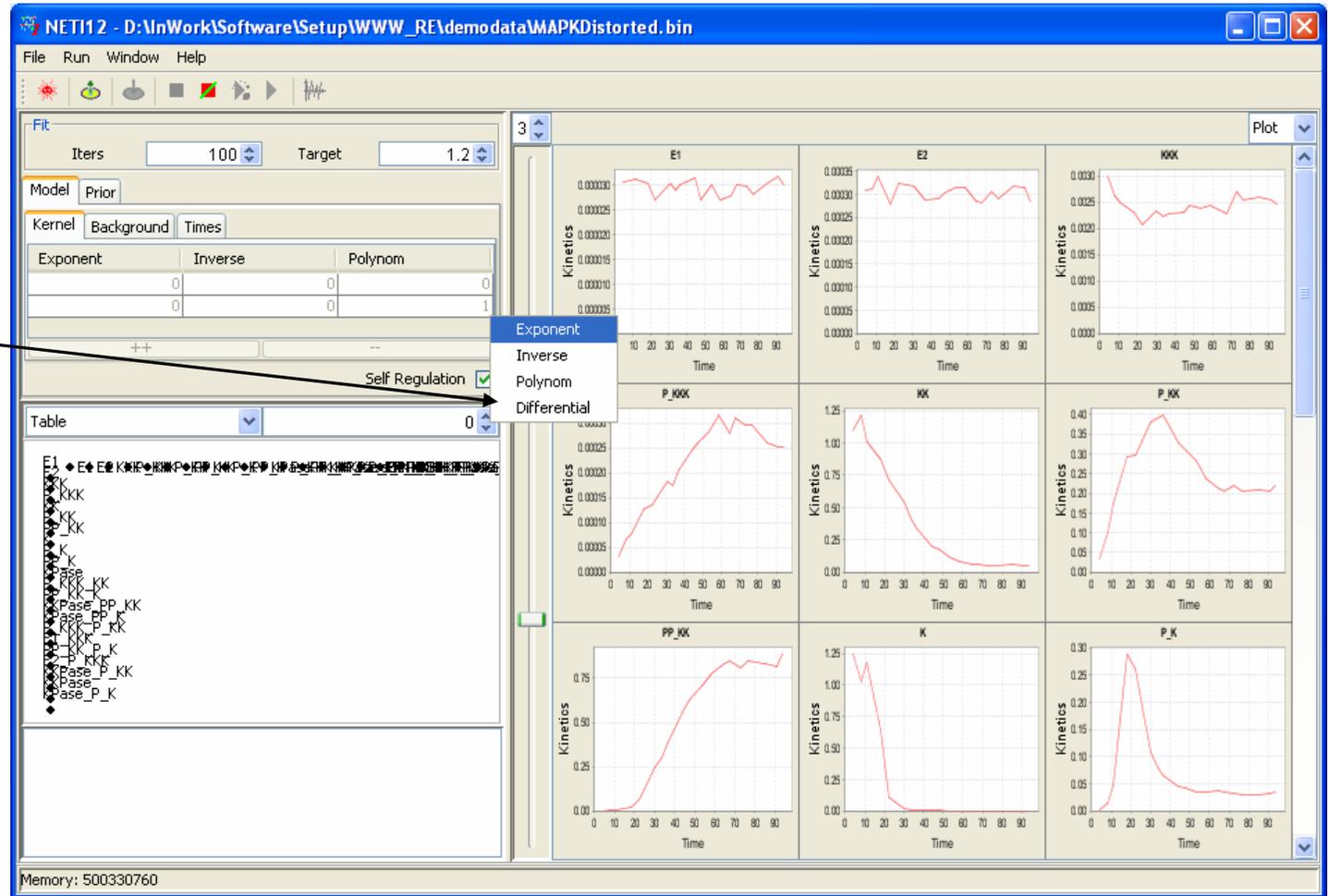
Polynomial kernel function



Inference Model Definition: Differential Model

Using the popup menu of the Model table the default representation for the kernel function can be selected:

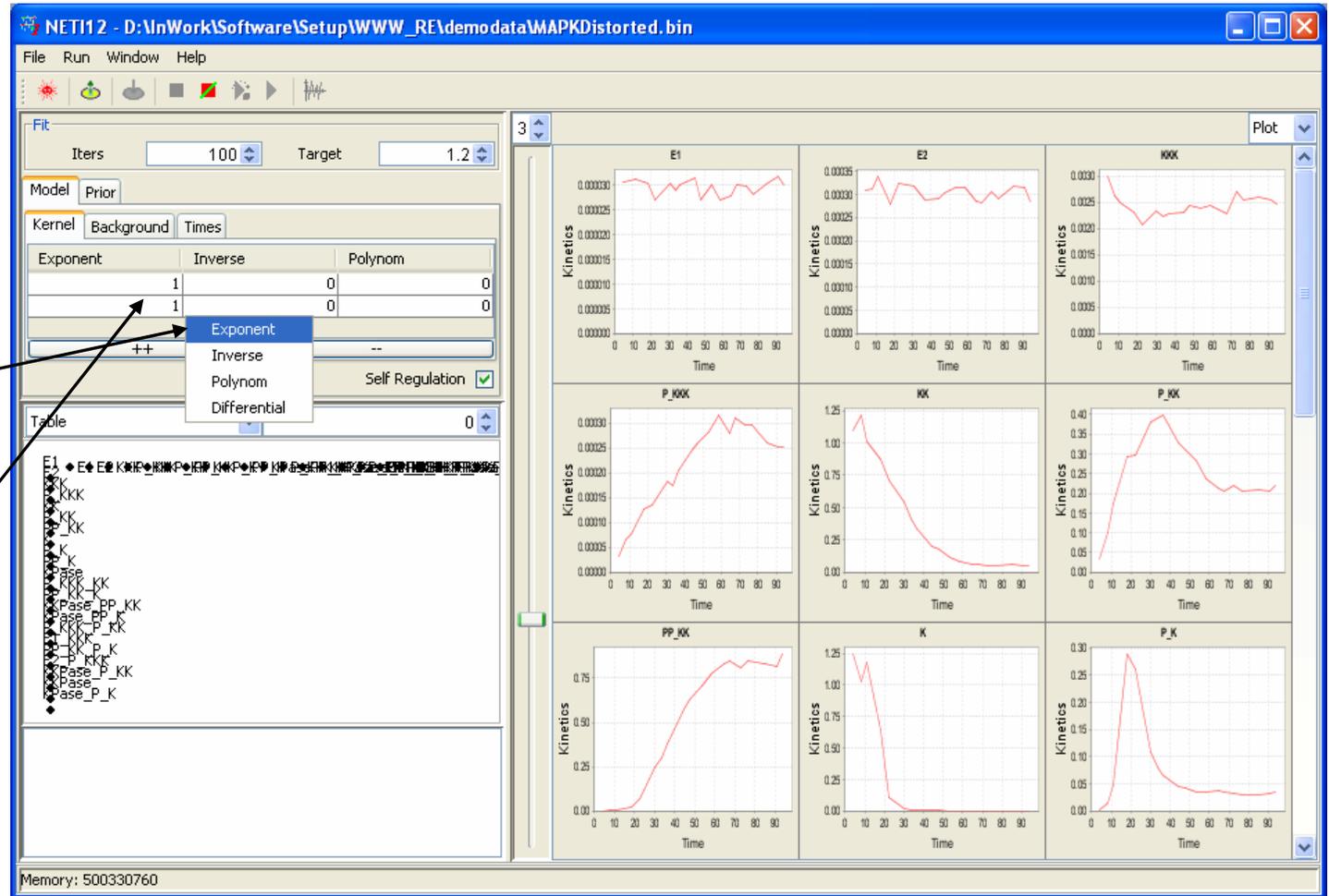
First-order, linear ordinary differential equations



Inference Model Definition: Exponential Model

Using the popup menu of the Model table the default representation for the kernel function can be selected:

Exponential kernel function



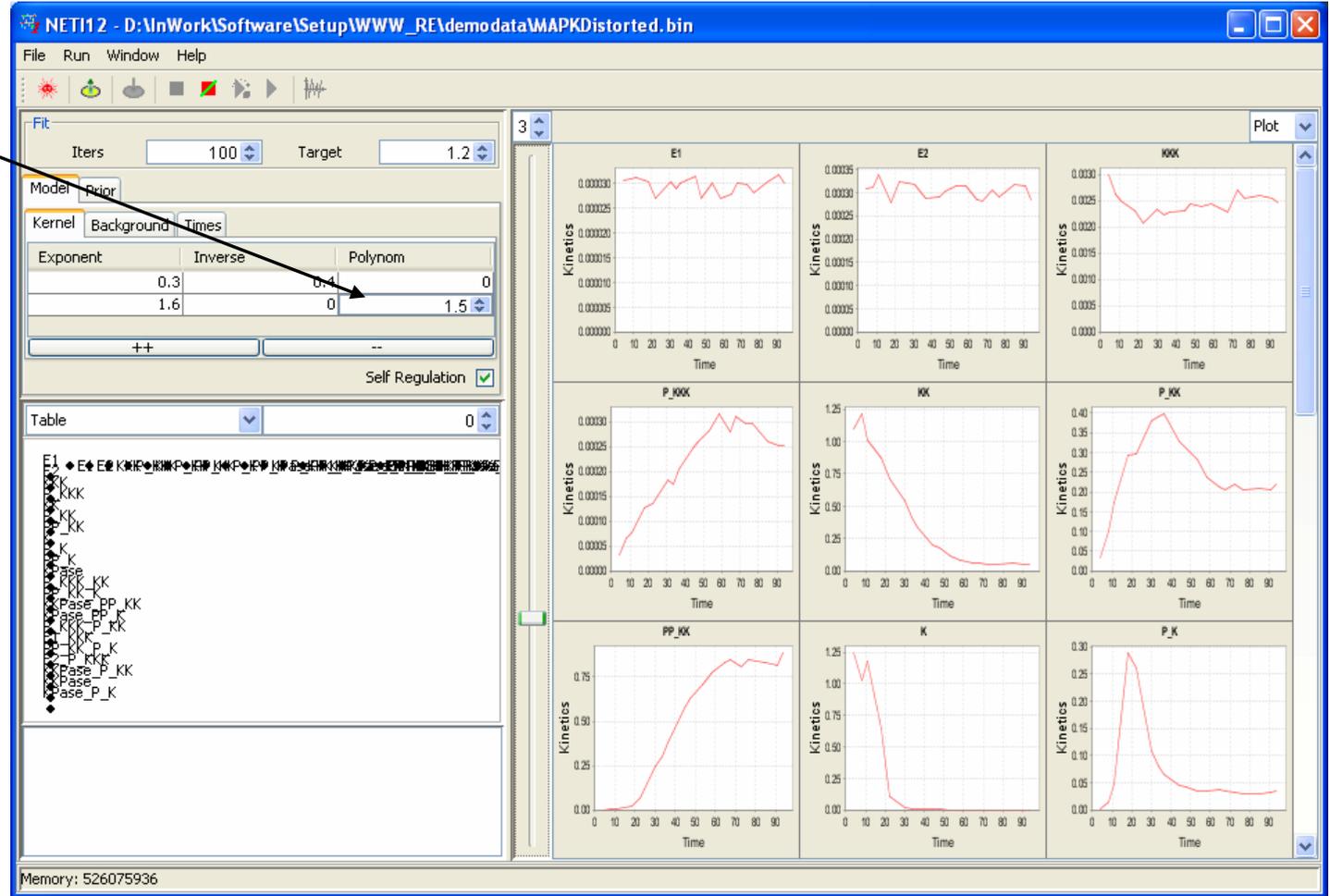
Inference Model Definition: Power-Factors

Power-factors for any term of any kernel function can be adjusted using the corresponding spinner boxes.

If the power-factors differ from zero for two or three functions, these functions are combined in a product. For example, the kernel function defined by the given combination of the power factors takes the form:

$$w_{ij}(t) = u_{1,ij} \exp\left\{-\left(t/\tau_{e1}\right)^{0.3}\right\} \times \left\{1 + \left(t/\tau_{i2}\right)^{0.4}\right\}^{-1} + u_{2,ij} \exp\left\{-\left(t/\tau_{e2}\right)^{1.6}\right\} \times t^{1.5}$$

Novikov E, Barillot E: *Model selection in the reconstruction of regulatory networks from time-series data*, submitted to *Algorithms for Molecular Biology*.

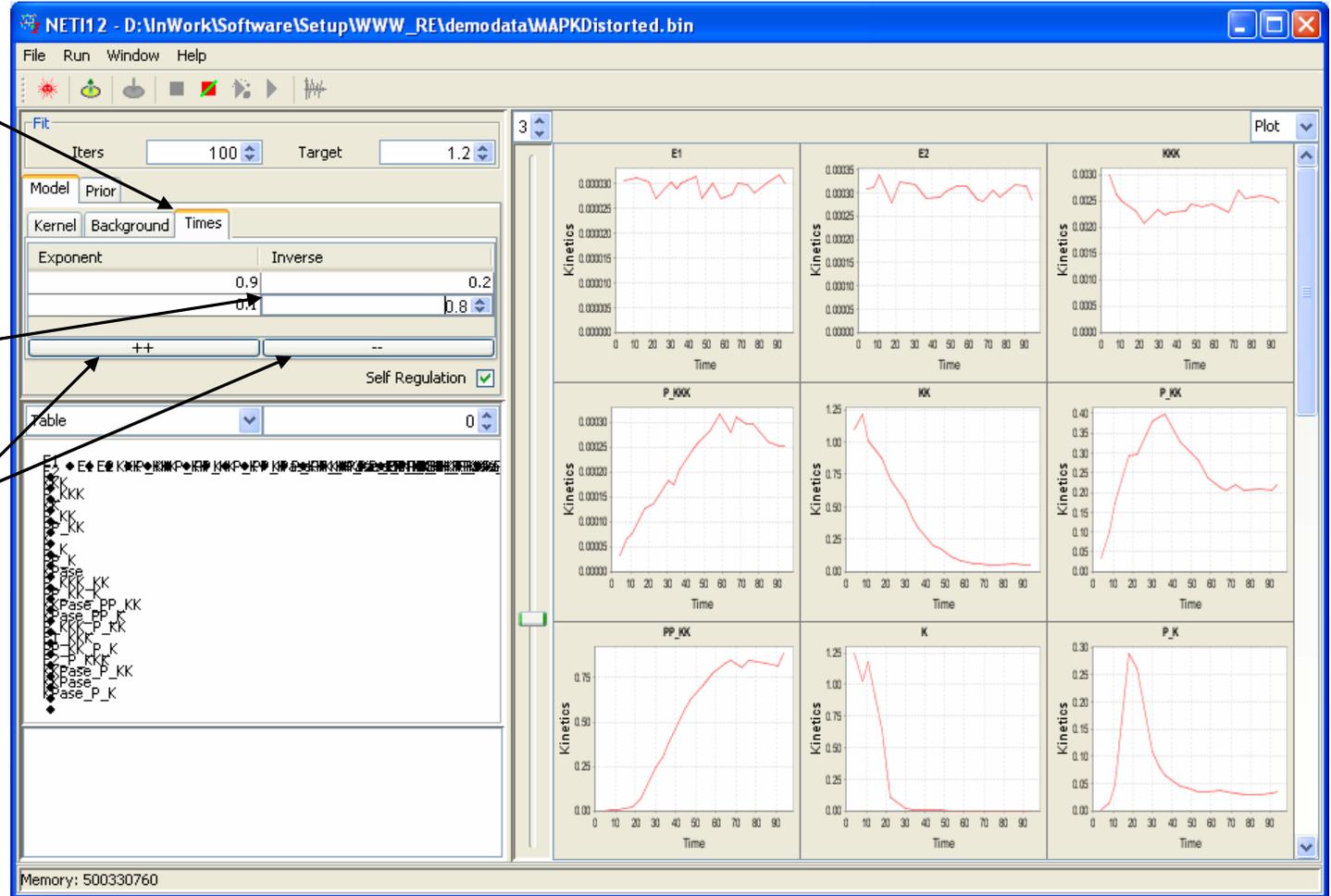


Inference Model Definition: Characteristic Times

Characteristic times for the Exponential and Inverse kernel functions can be set using the corresponding spinner boxes.

The characteristic times are defined as $\alpha * T$, where α is taken from the table, and T is the measurement time range.

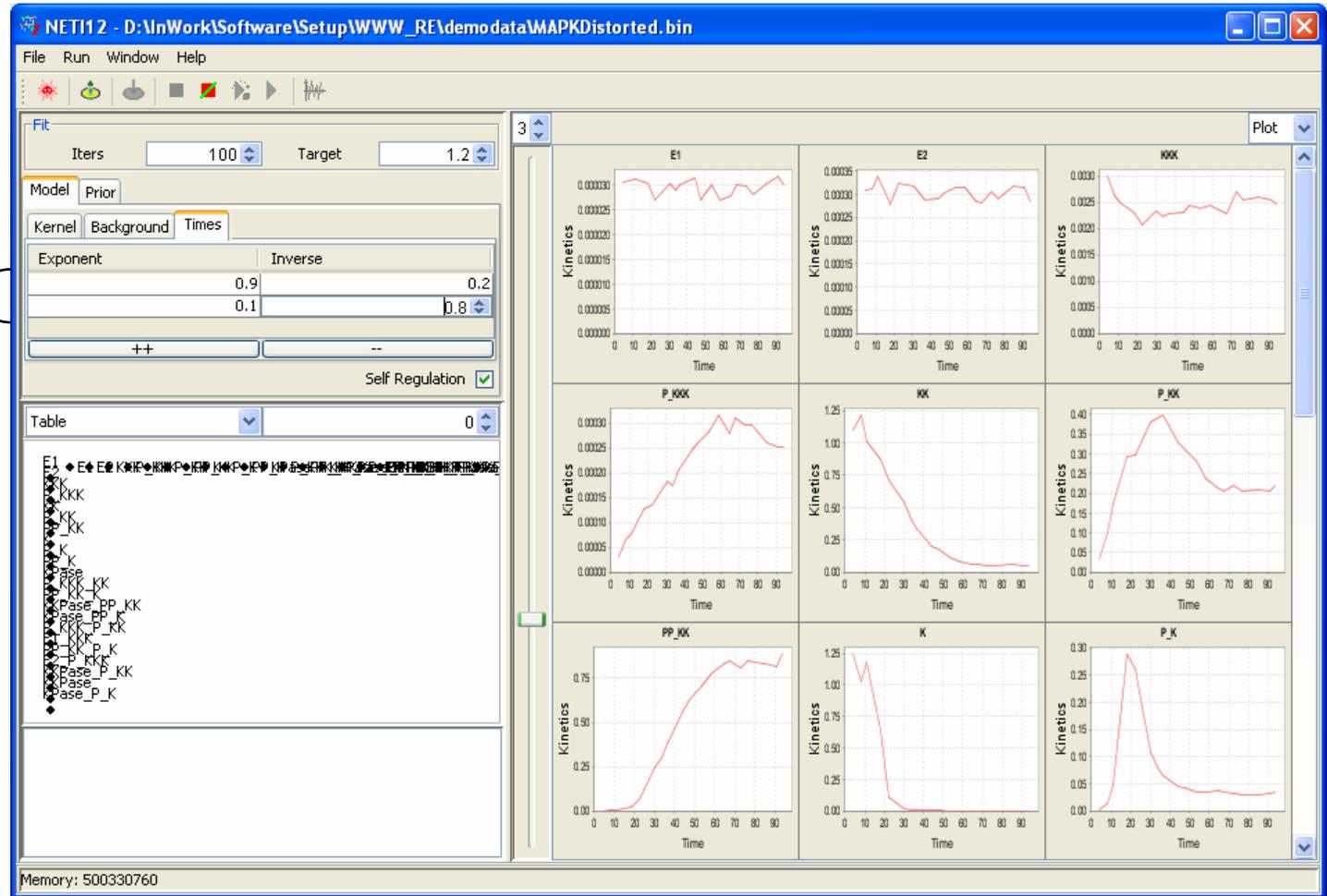
The number of times can be increased using the “++” button or can be decreased using the “—” button.



Inference Model Definition: Characteristic Times Selection

The number of characteristic times should not be smaller than the number of terms in the kernel functions.

If the number of characteristic times is larger than the number of kernel terms, all possible combinations of characteristic times are consequently substituted in the kernel function and fitting is performed for each tested link. The combination ensuring the best fitness value (χ^2) is preserved for this link.

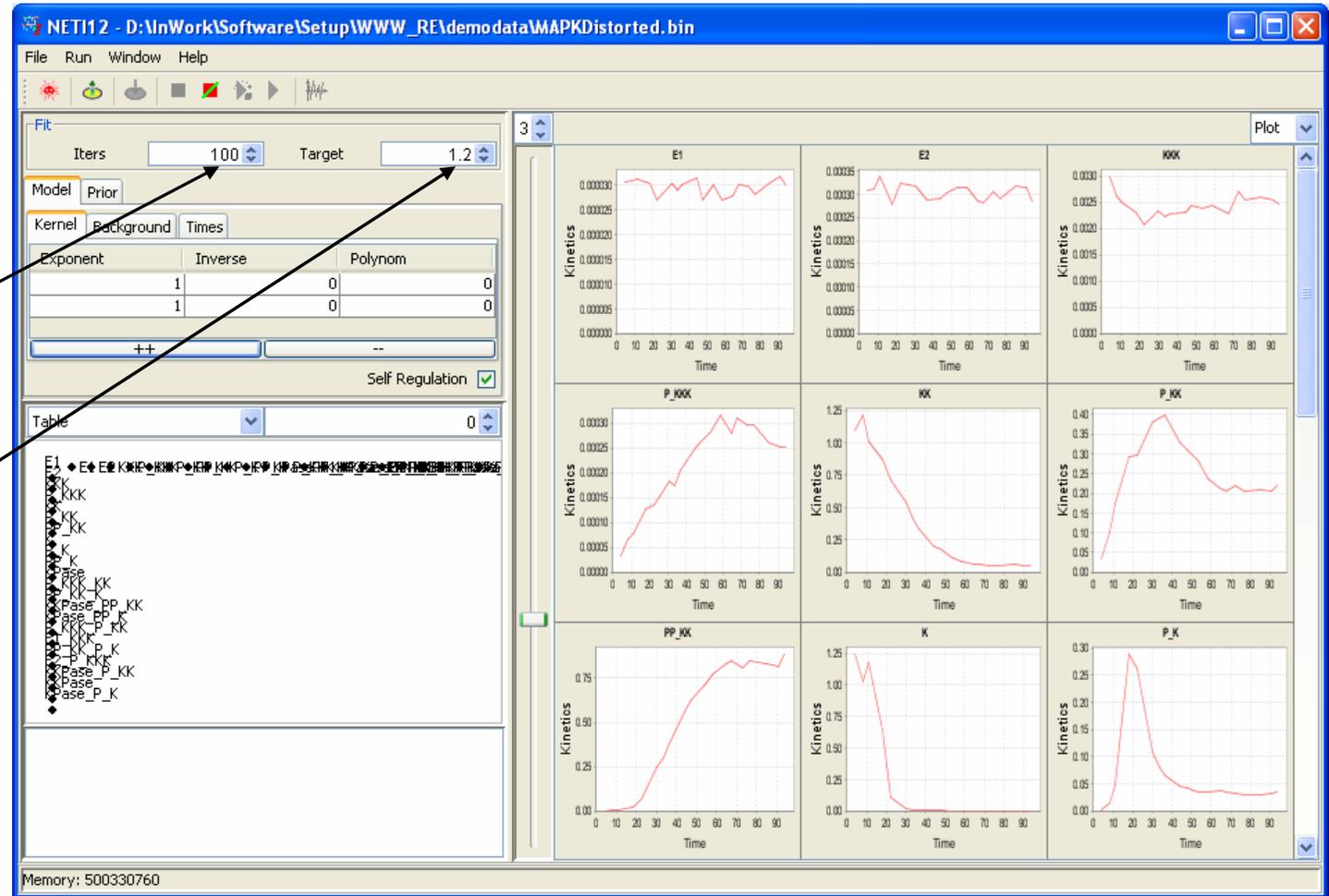


Forward Selection Stopping Criteria

Iters is the stopping criterion: maximal number of links that can be created.

Target is the stopping criterion: minimal limit for the χ^2 overall fitness criterion.

Novikov E, Barillot E: *Regulatory network reconstruction using an integral additive model with flexible kernel functions*, submitted to *BMC Systems Biology*.

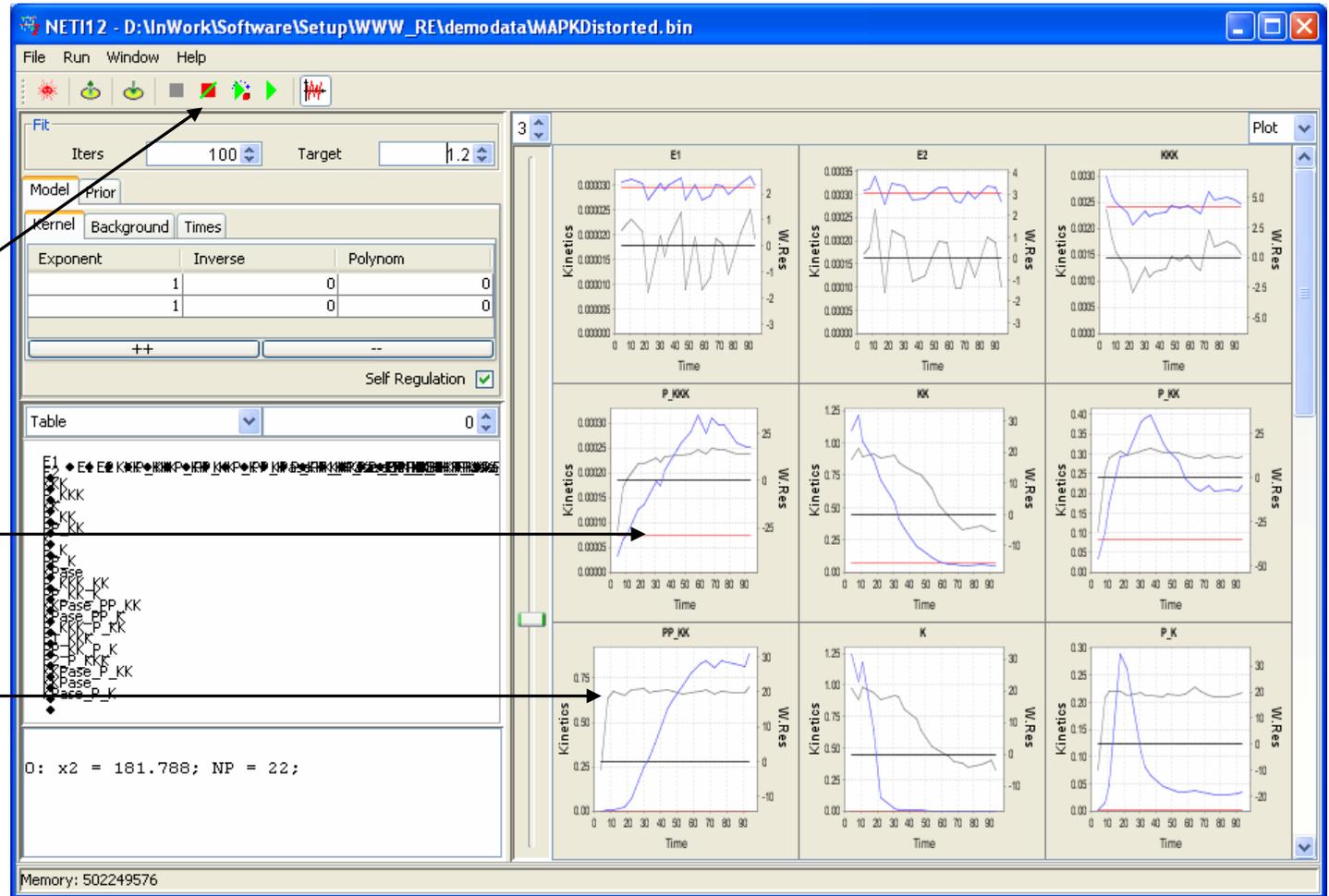


Initialization

Press the “Init” button from the Toolbar or select the Menu Item “Run|Init” (F3) to initialize the inference.

The default (background) behavior will be assigned to each node (red line).

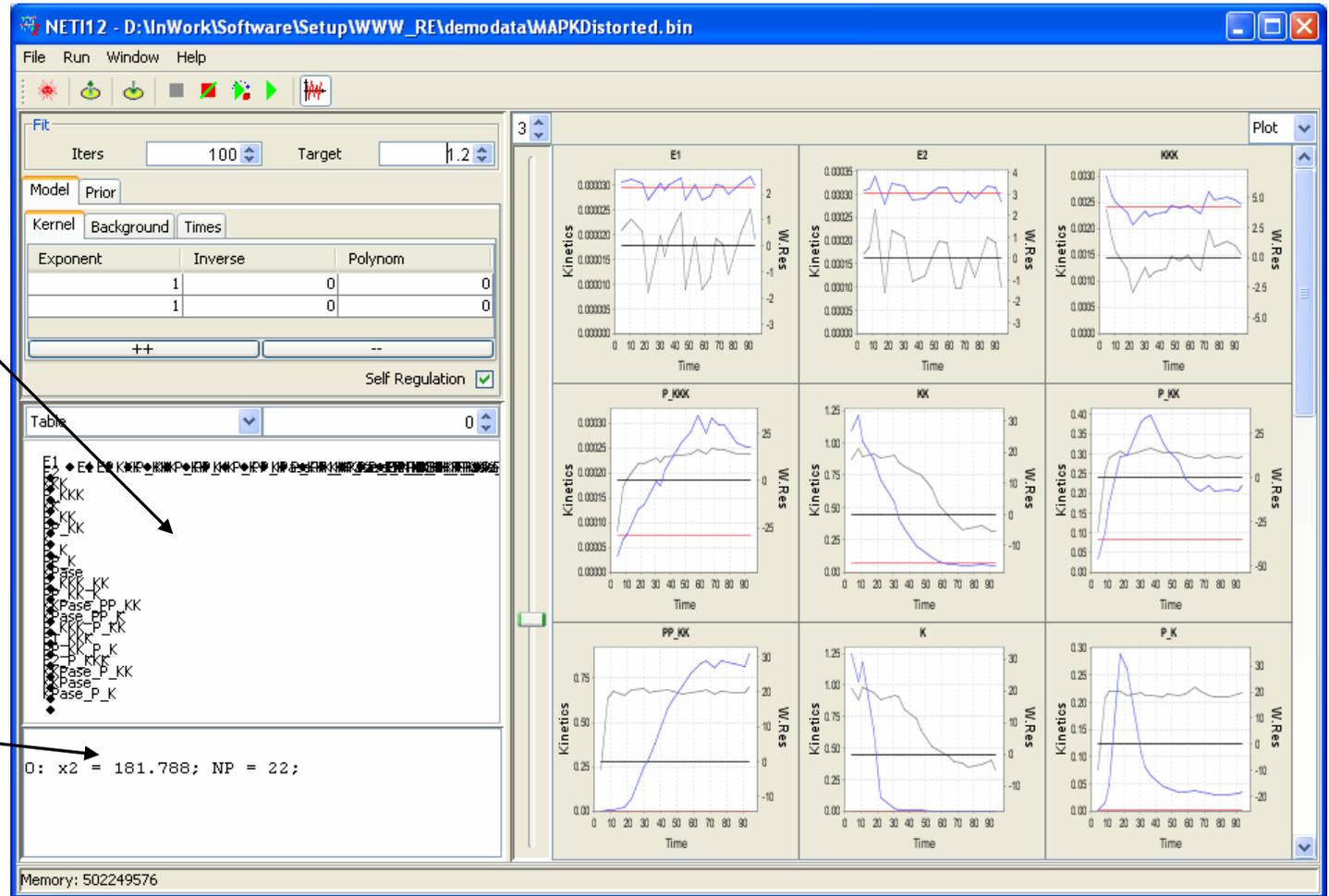
Gray line is the residuals: the differences between the experimental and reconstructed time series.



Initialization: Interaction Table

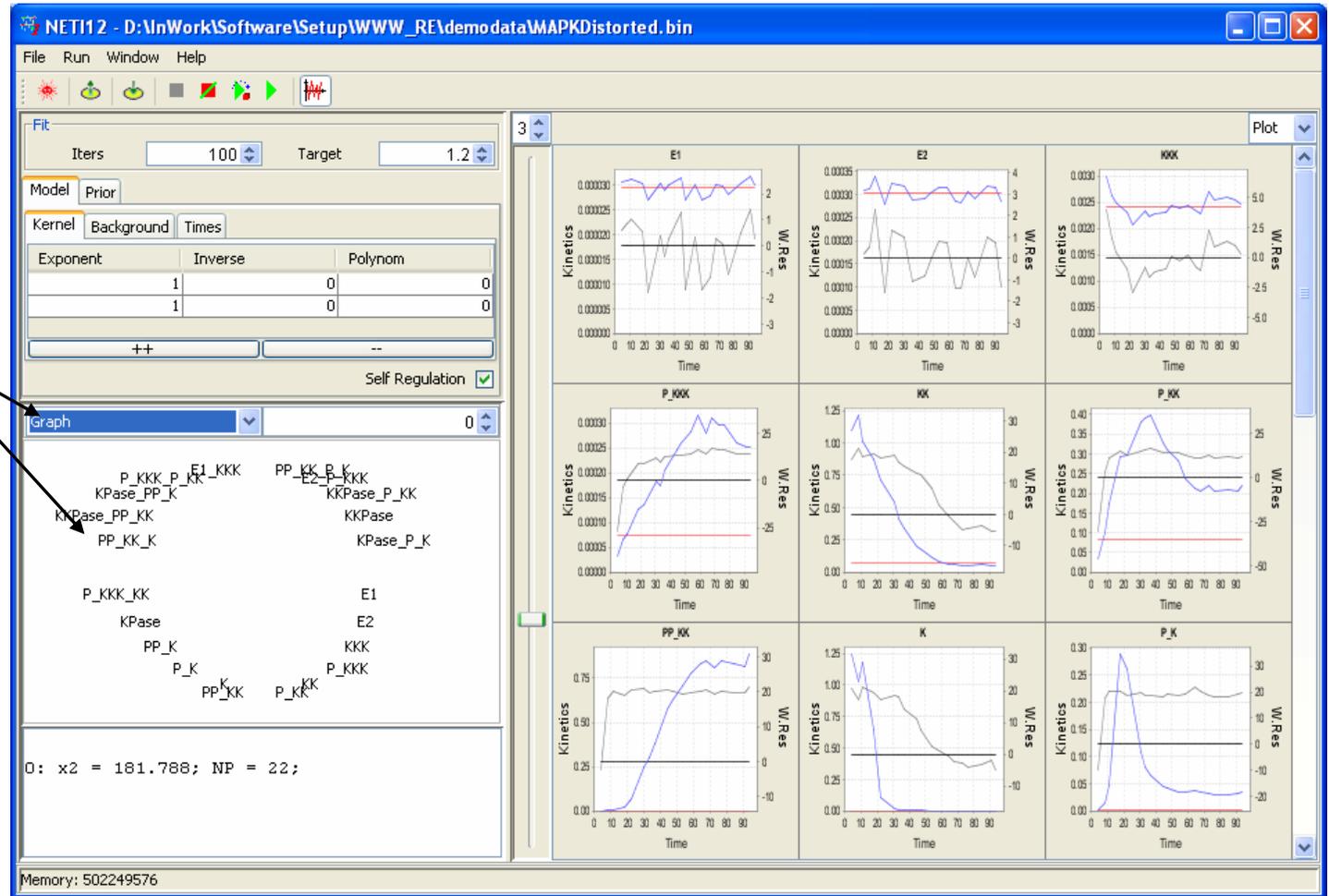
The interaction table is empty as there are no interacting nodes so far.

The number of iterations (links) and the fitness criterion are shown in the log window.



Initialization: Network Graph

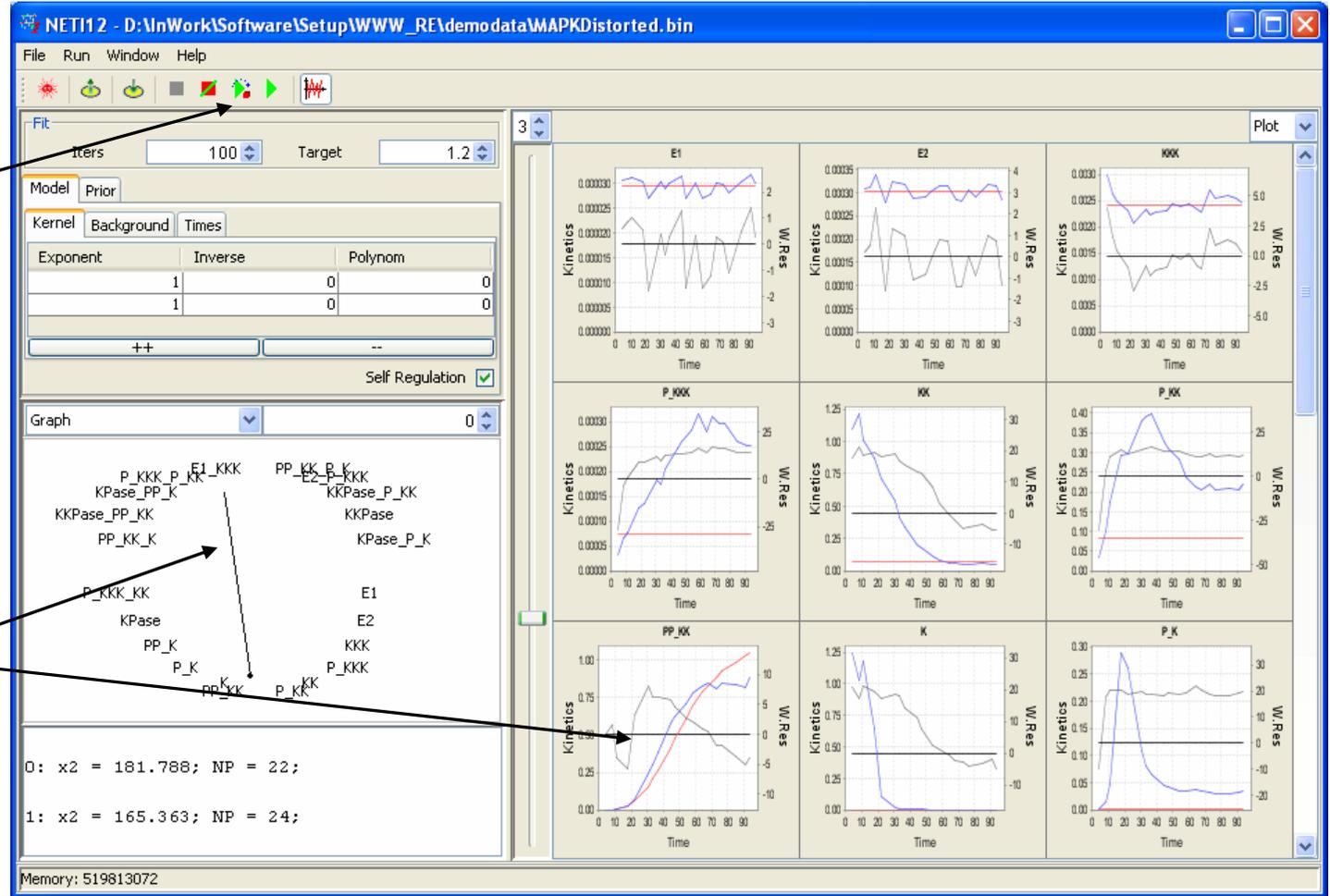
The interaction table can also be presented as a network graph.



Step-by-Step Inference

Network inference can be performed in the “step-by-step” mode, using the “Step” button from the Toolbar or the Menu Item “Run|Step” (F4).

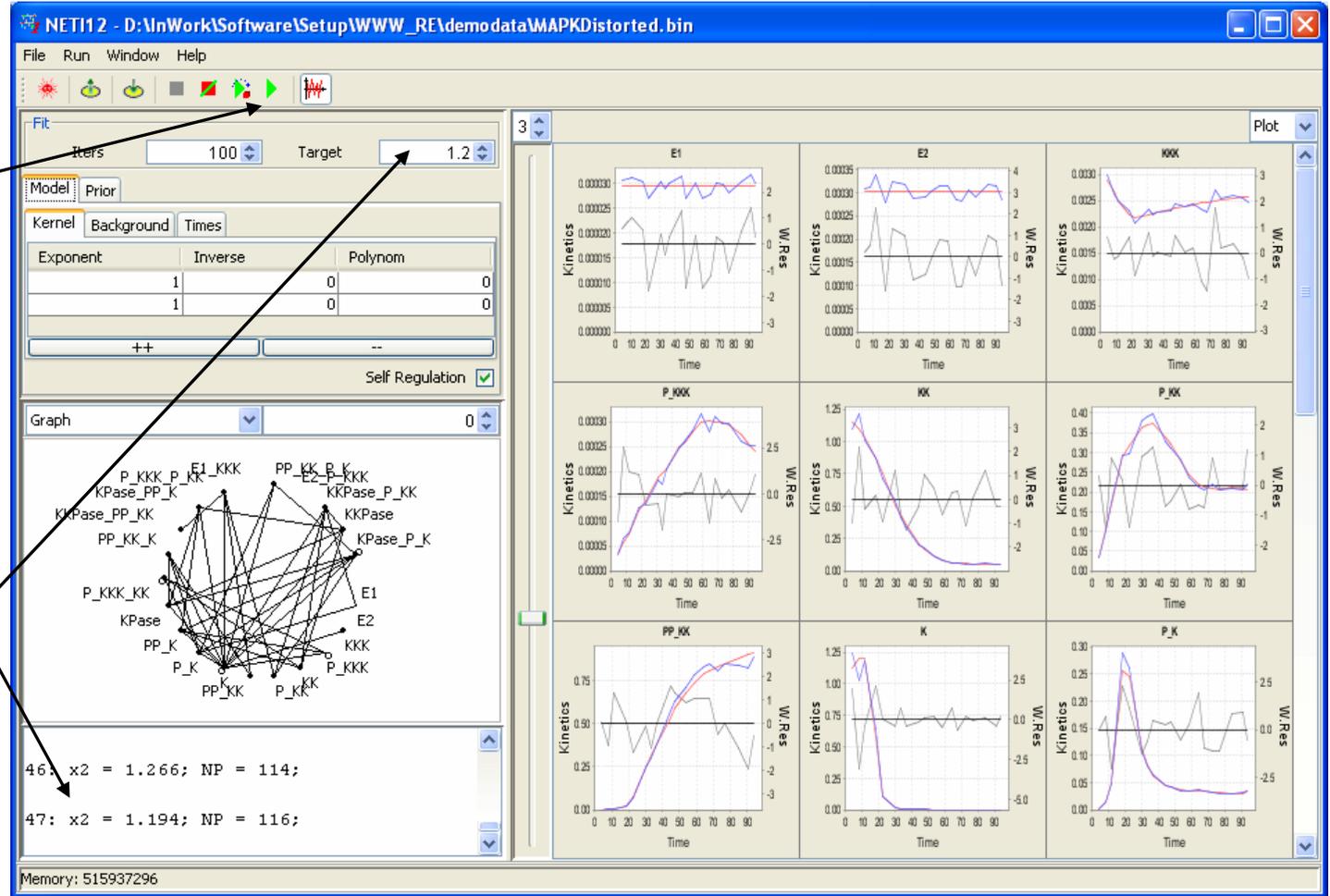
At each step, the procedure selects the node with lowest fitness and finds another node, which can explain the behavior of the given node in the best way (ensuring the lowest χ^2 criterion value for the given node).



Automatic Inference

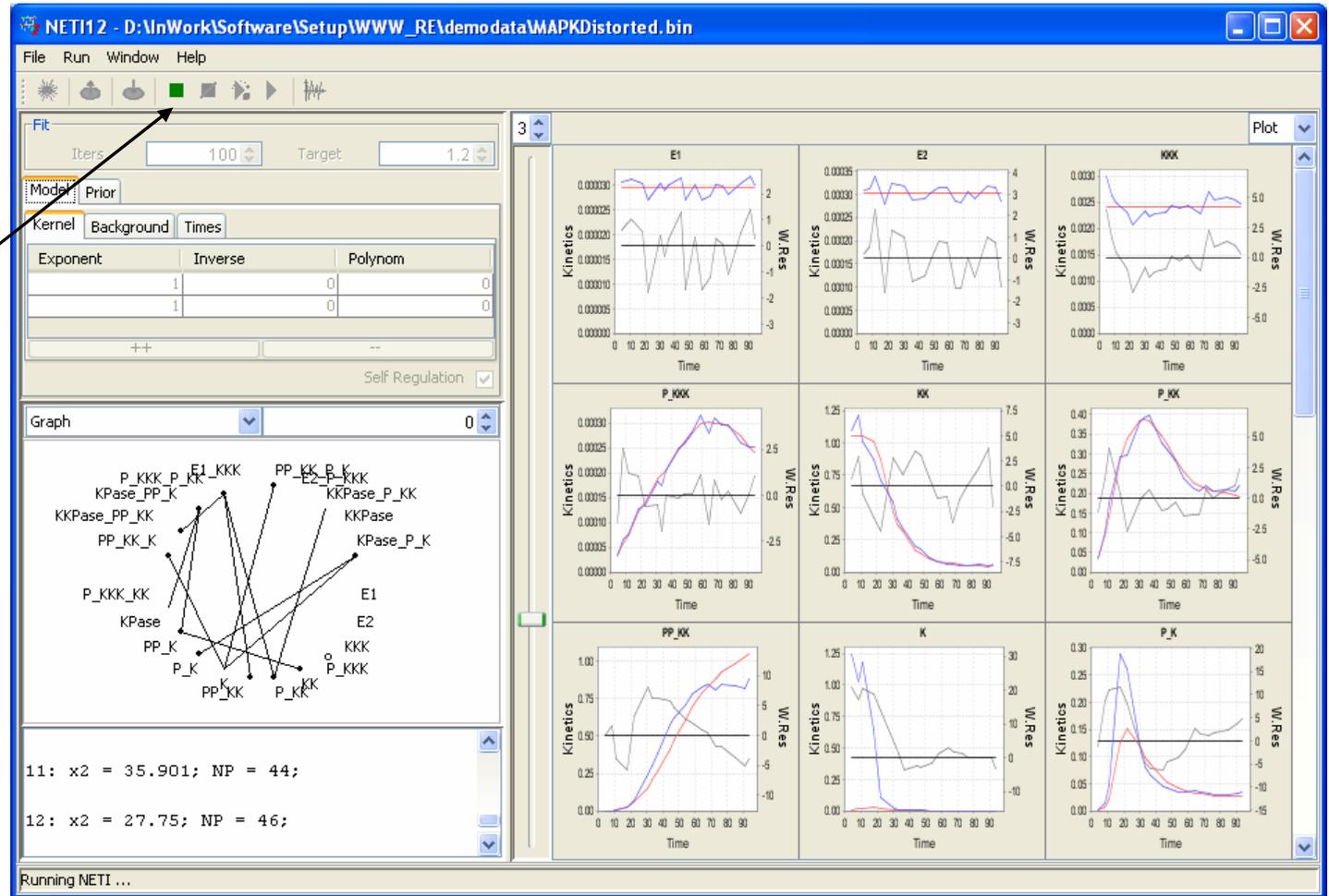
Network inference can be performed in the automatic mode, using the “Run” button from the Toolbar or the Menu Item “Run|Run” (F5).

The procedure performs until the number of created links is higher than the *Iters* value, or until the χ^2 value is lower that the *Target* value.



Terminate Processing

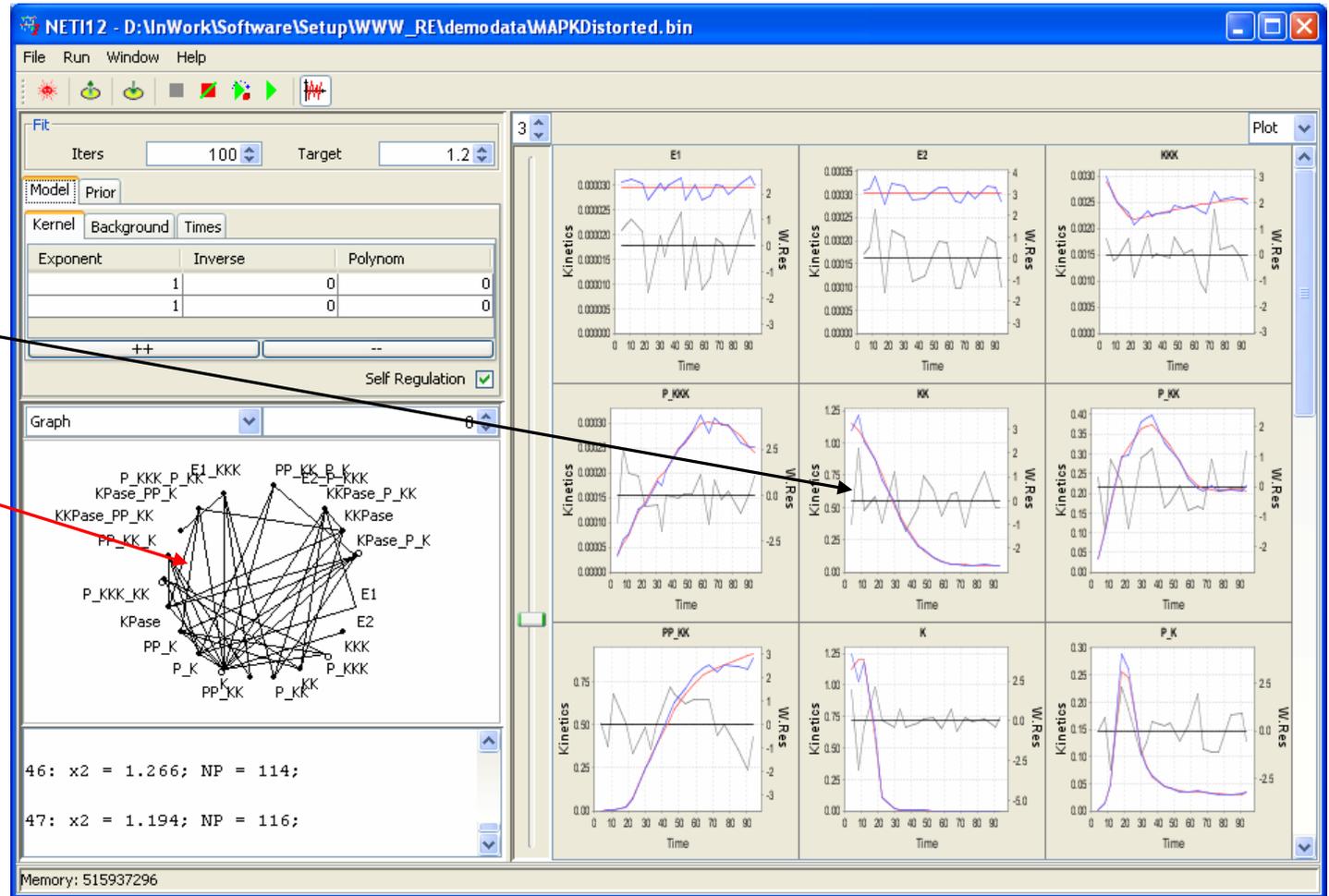
Processing can be stopped by pressing the “Stop” button on the Toolbar or selecting the Menu Item “Run|Stop”.



Results

The quality of fit can be visually appreciated.

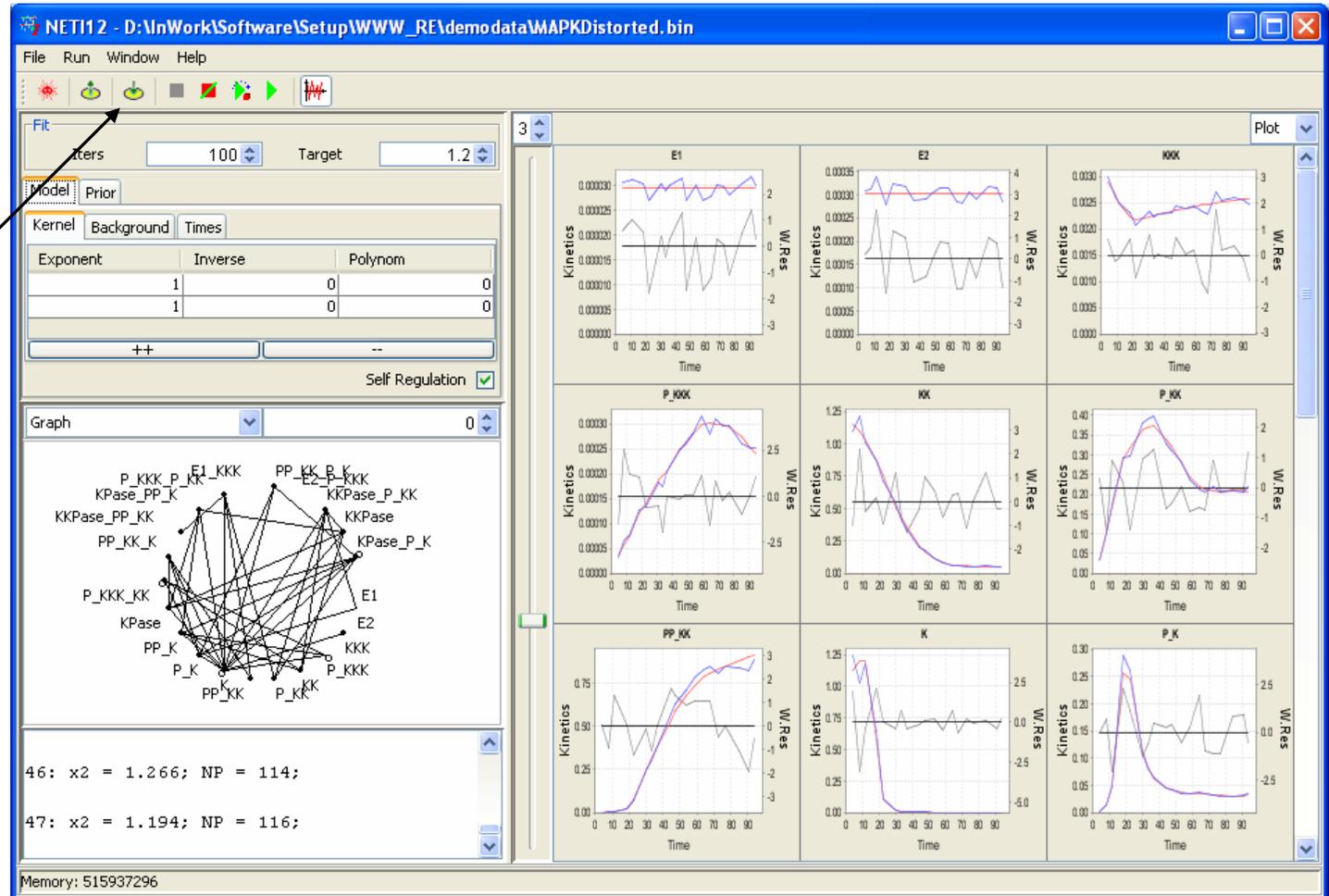
The resulting network graph.



Save Results

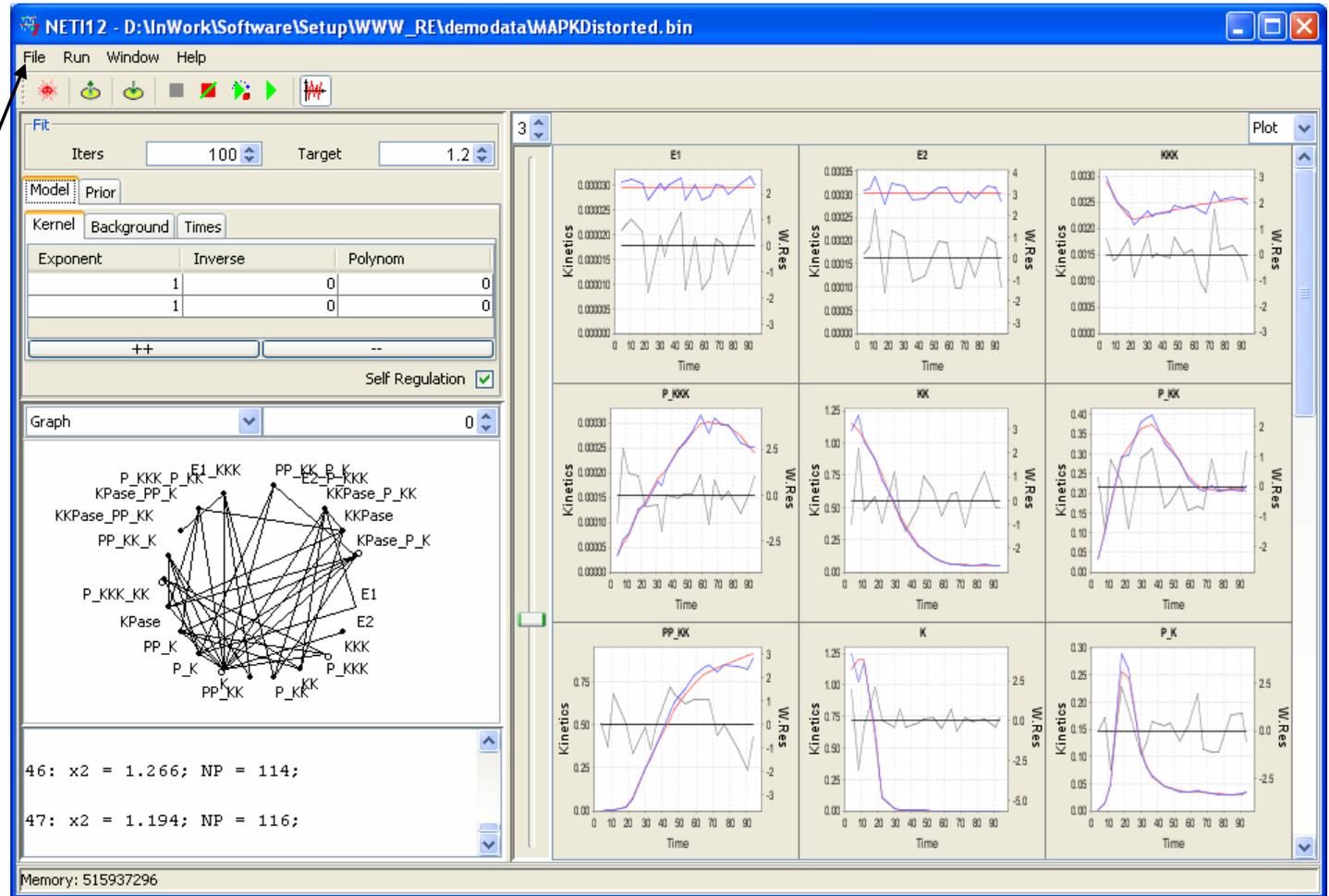
To save the results use the “Save Analysis ...” button from the Toolbar or the Menu Item “File|Save|Save Analysis ...” (Ctrl+S).

The results are saved as a list of links in the text file (importable into Microsoft Excel).



Set Batch Options

Using the Menu Item “File/Set Batch Options”, all settings can be saved to be applied to other data sets.



Adaptive Model Selection

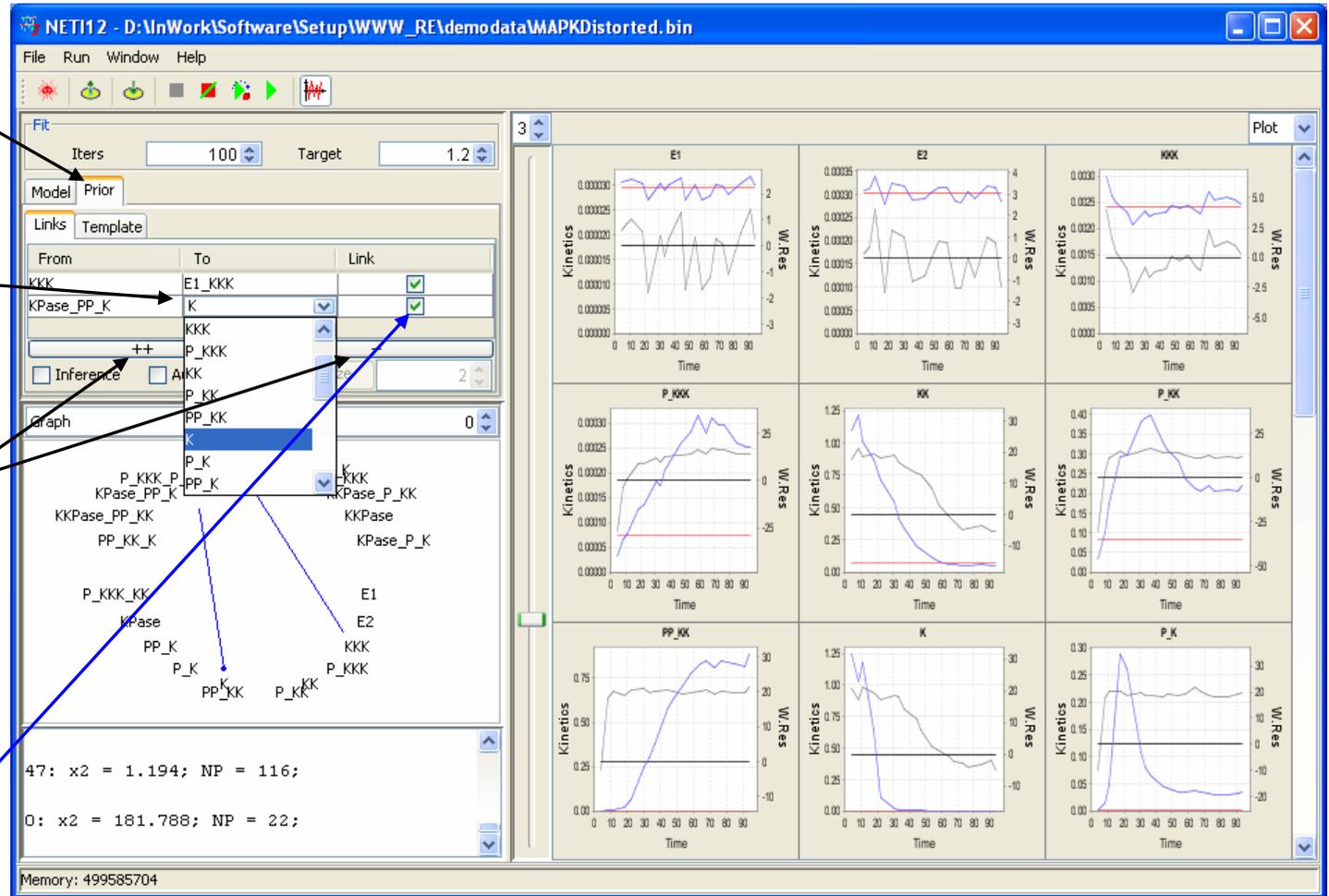
Prior Links Table

Prior links are defined in the “Links” table from the “Prior” tab.

The nodes to be linked are selected from the lists of measured nodes.

The number of prior links can be increased using the “++” button or can be decreased using the “—” button.

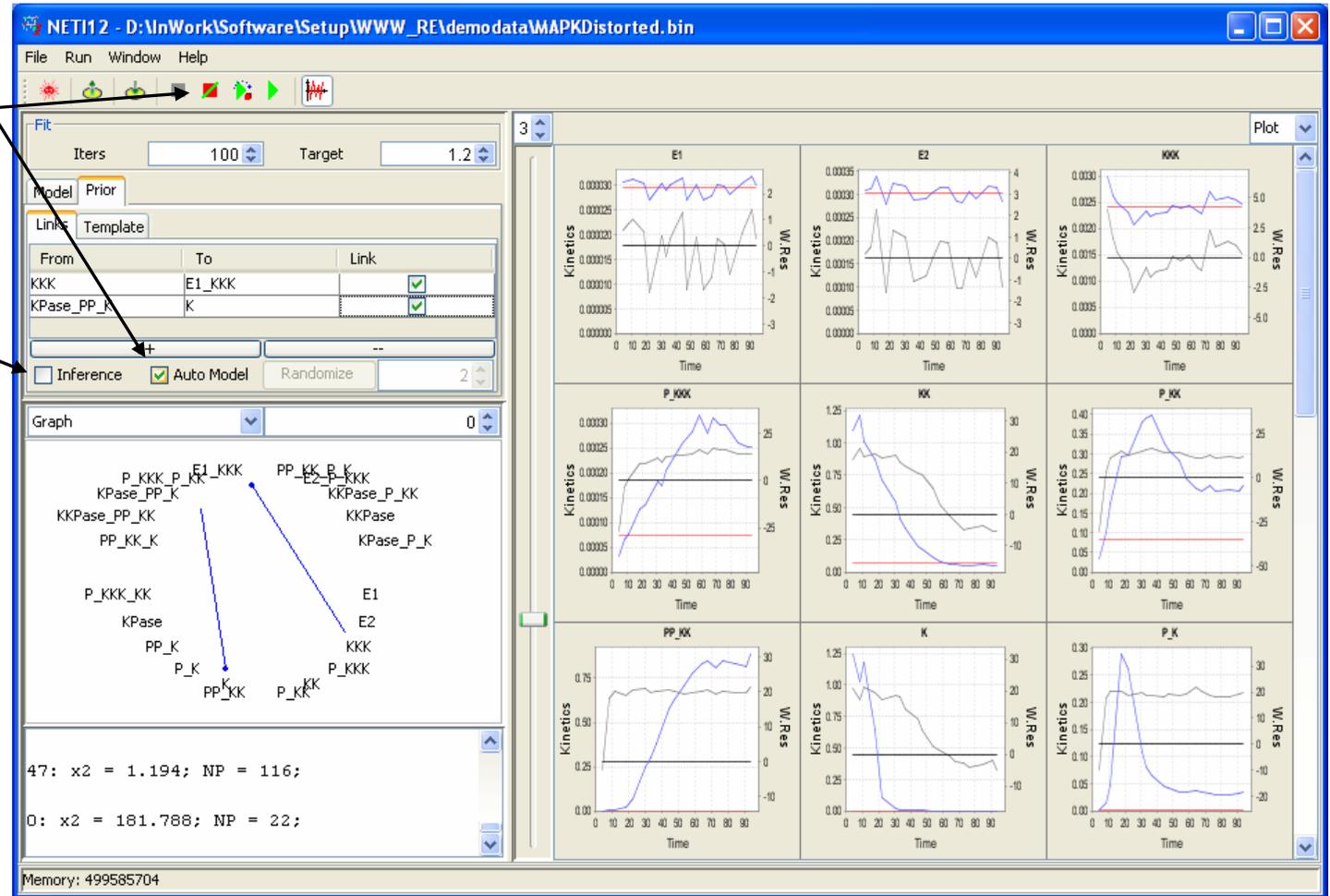
Only the selected links will be used for the adaptive model selection.



Prior Links Use

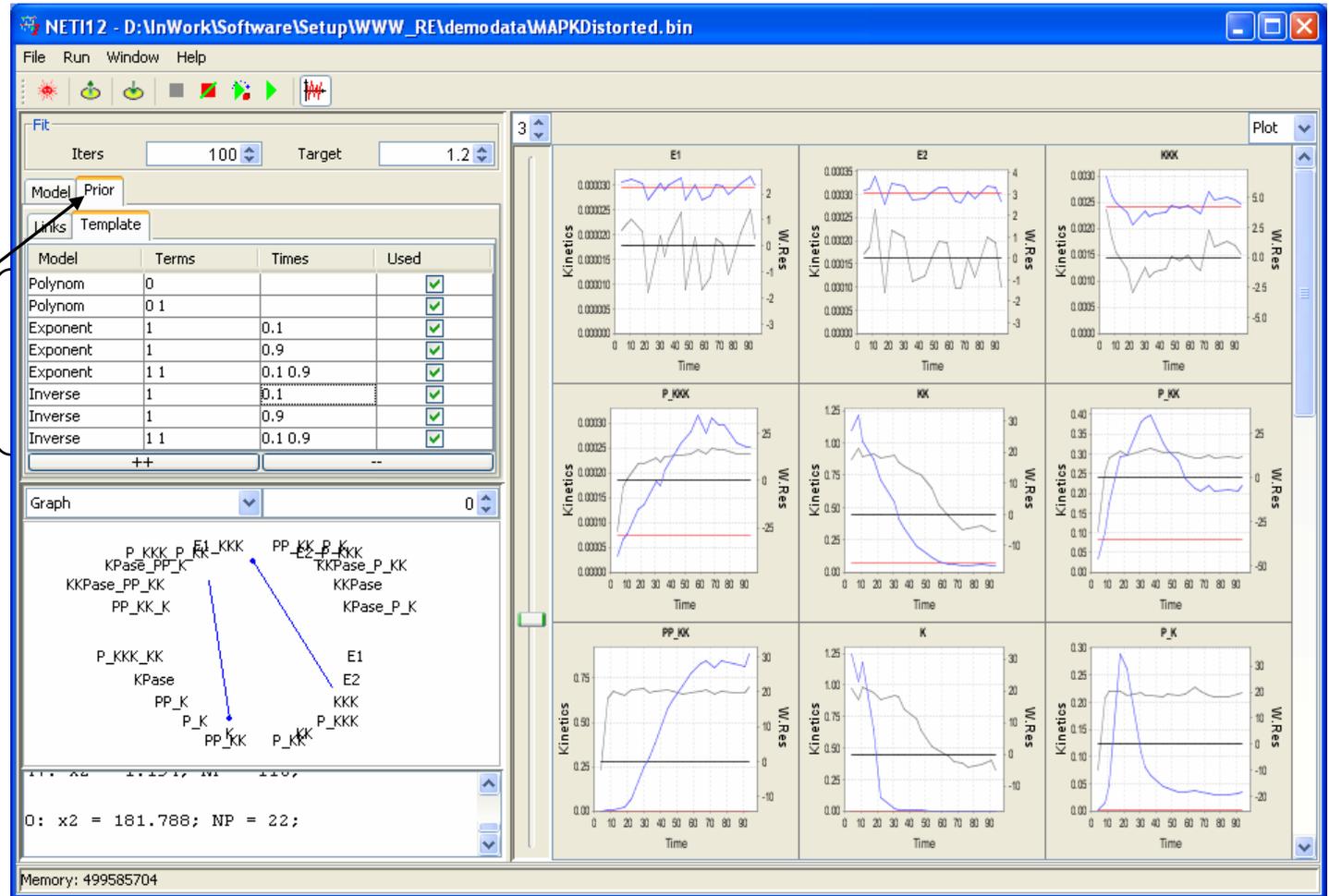
If the “Auto Model” checkbox is selected the model selection procedure will be performed at the initialization.

If the “Inference” checkbox is selected, the prior links will be used not only for the model selection but also will be imbedded in the selected model for final reconstruction.



Models Library

The models to be tested are defined in the “Template” table from the “Prior” tab.



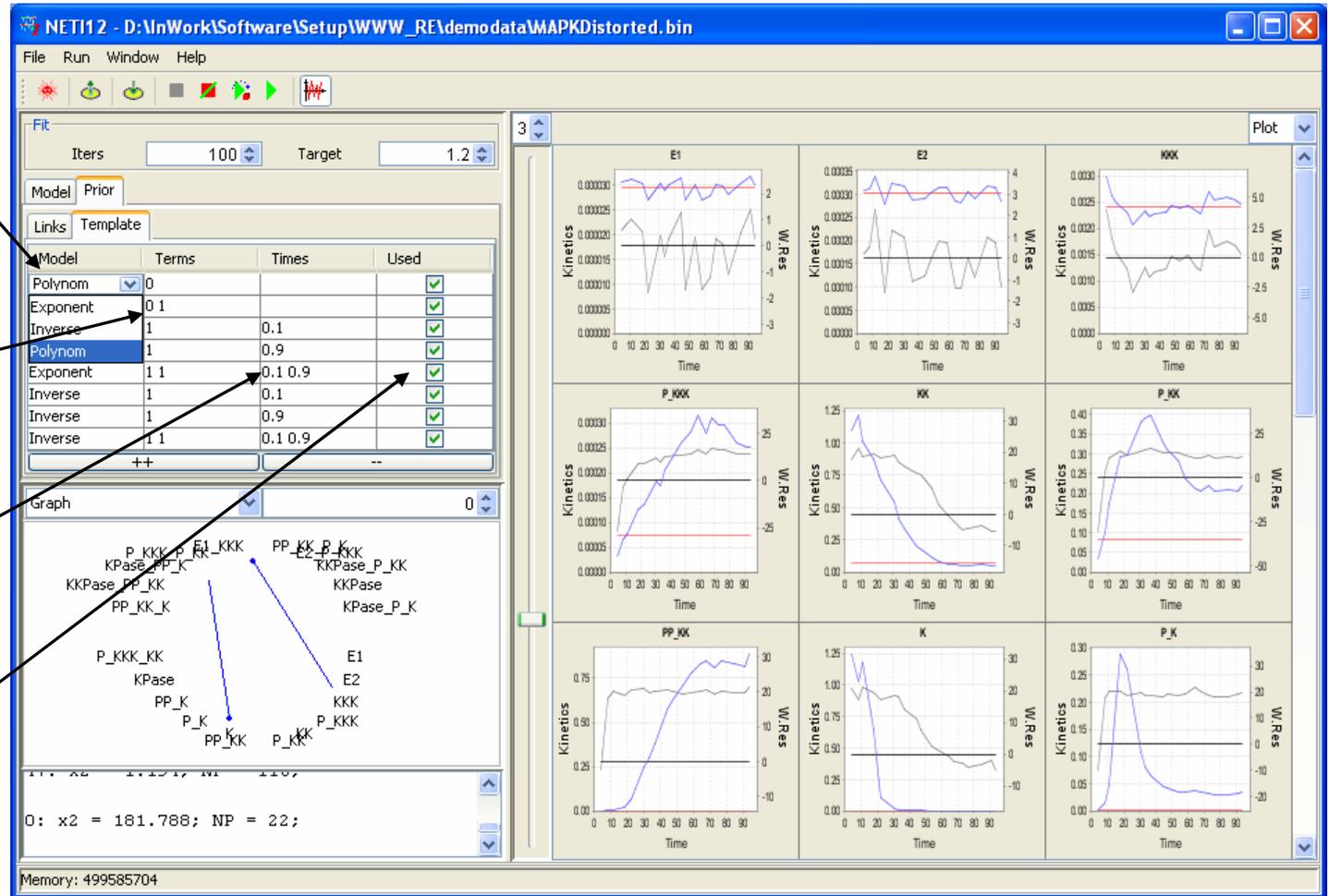
Models Template

The model is selected from the three implemented models: *Exponential, Inverse, Polynomial*.

The number of terms for each model is defined by the number of Power-Factors separated by spaces.

The corresponding characteristic times (α) should also be separated by spaces.

Only the selected models will be used in testing.



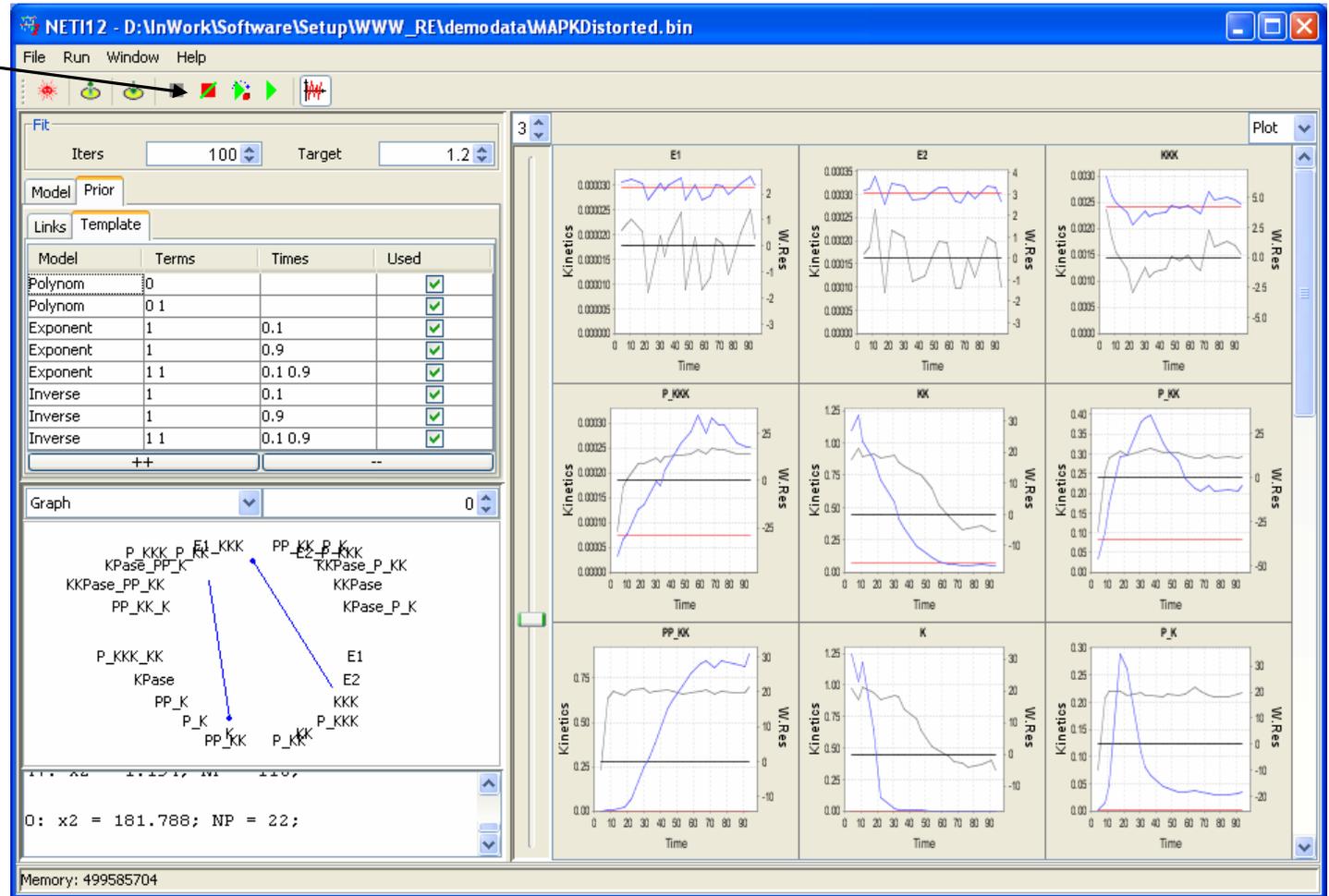
Adaptive Model Selection

At initialization, the FS procedure uses consequently each of the defined models to reconstruct the network.

The number of correctly recovered prior links is counted after the number of iterations (NI) defined by the number of prior links (2, in this case).

If NI is not sufficient to unambiguously identify the best model, the NI is increased by one and the FS procedure starts again.

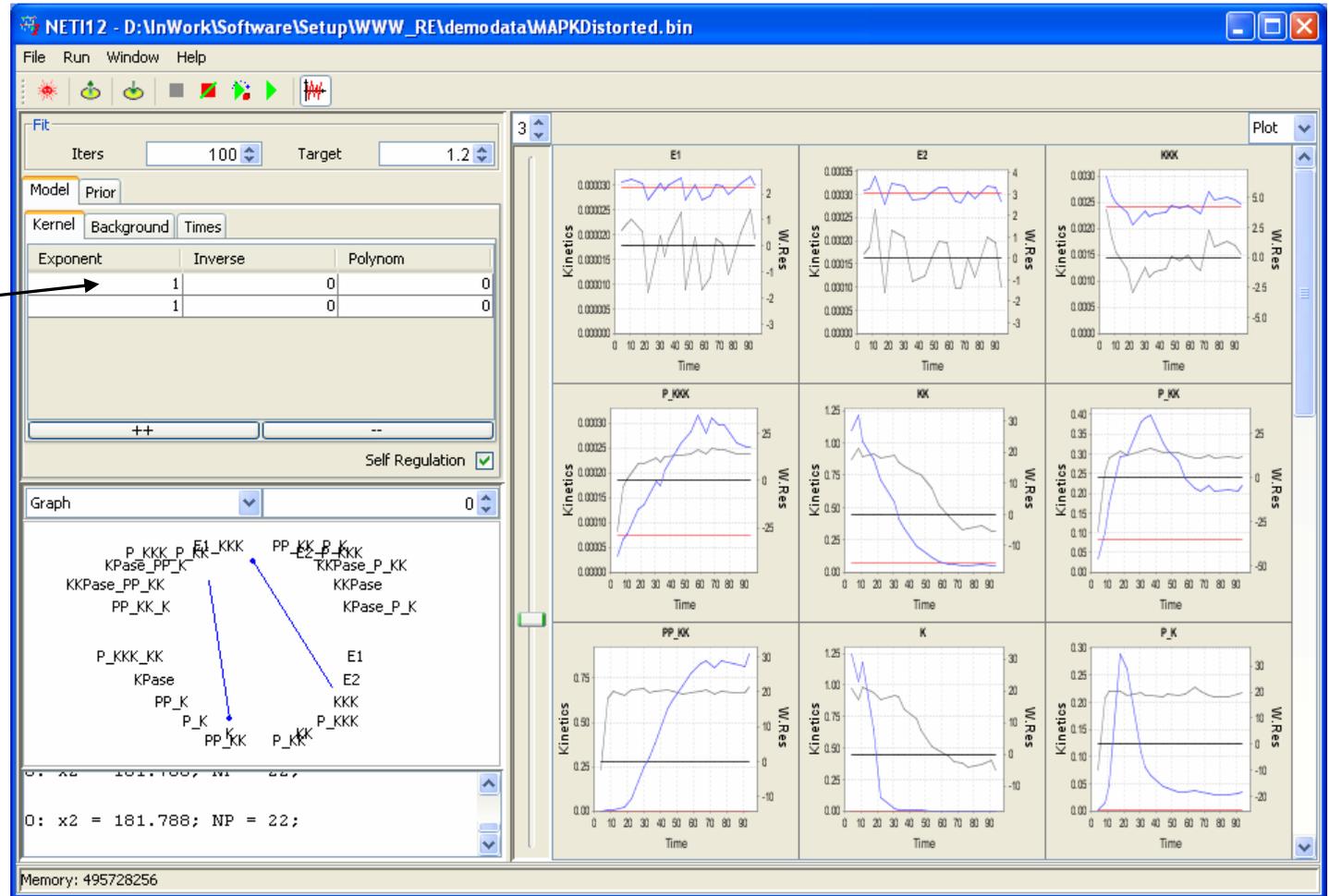
Novikov E, Barillot E: *Model selection in the reconstruction of regulatory networks from time-series data*, submitted to *Algorithms for Molecular Biology*.



Model Selection Result: Kernel

The model selection procedure, after testing the eight models defined in the “Template” table, selects the best model and copies the identification of this model into the tables “Kernel” and “Times” of the “Model” tab.

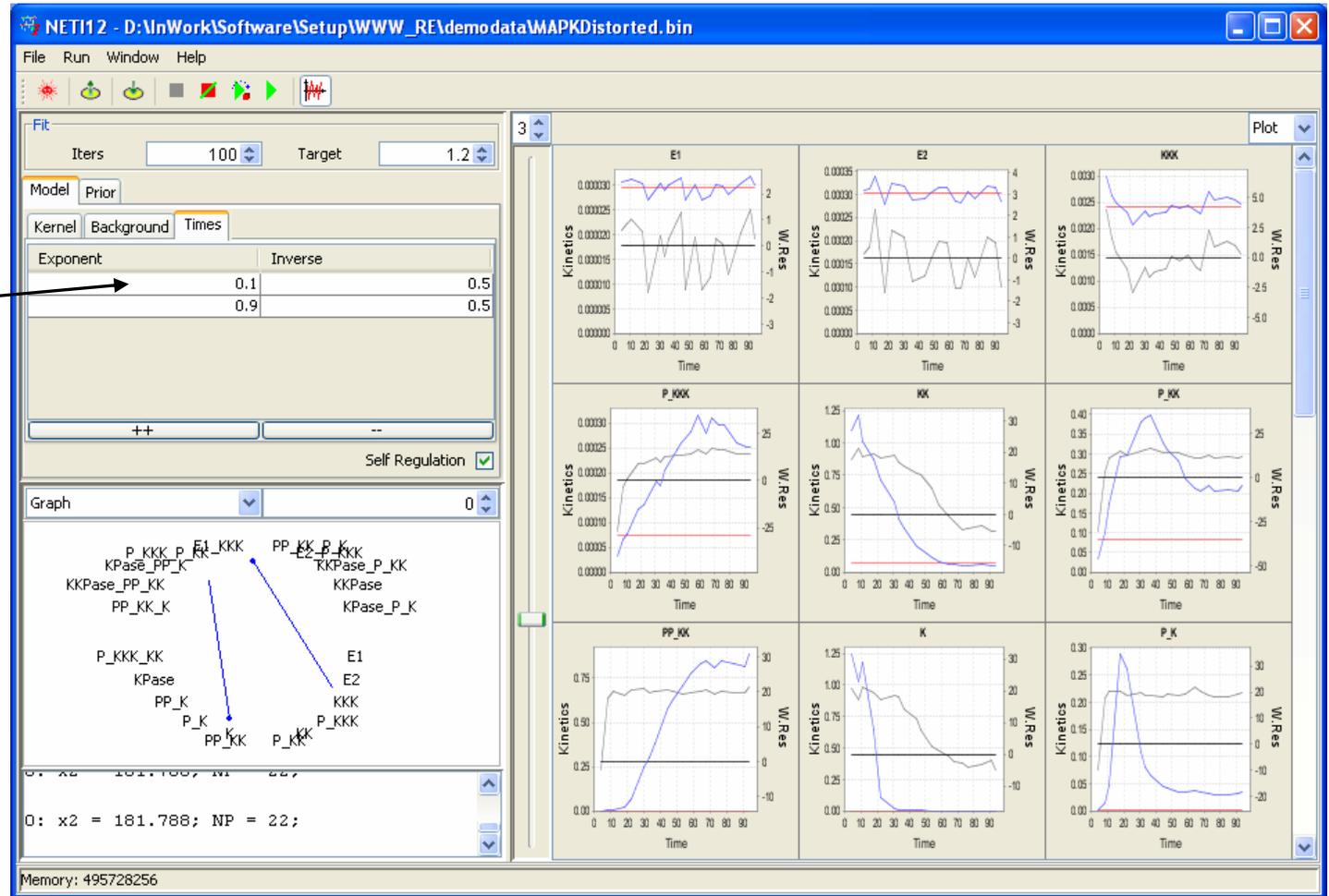
Kernel table



Model Selection Result: Times

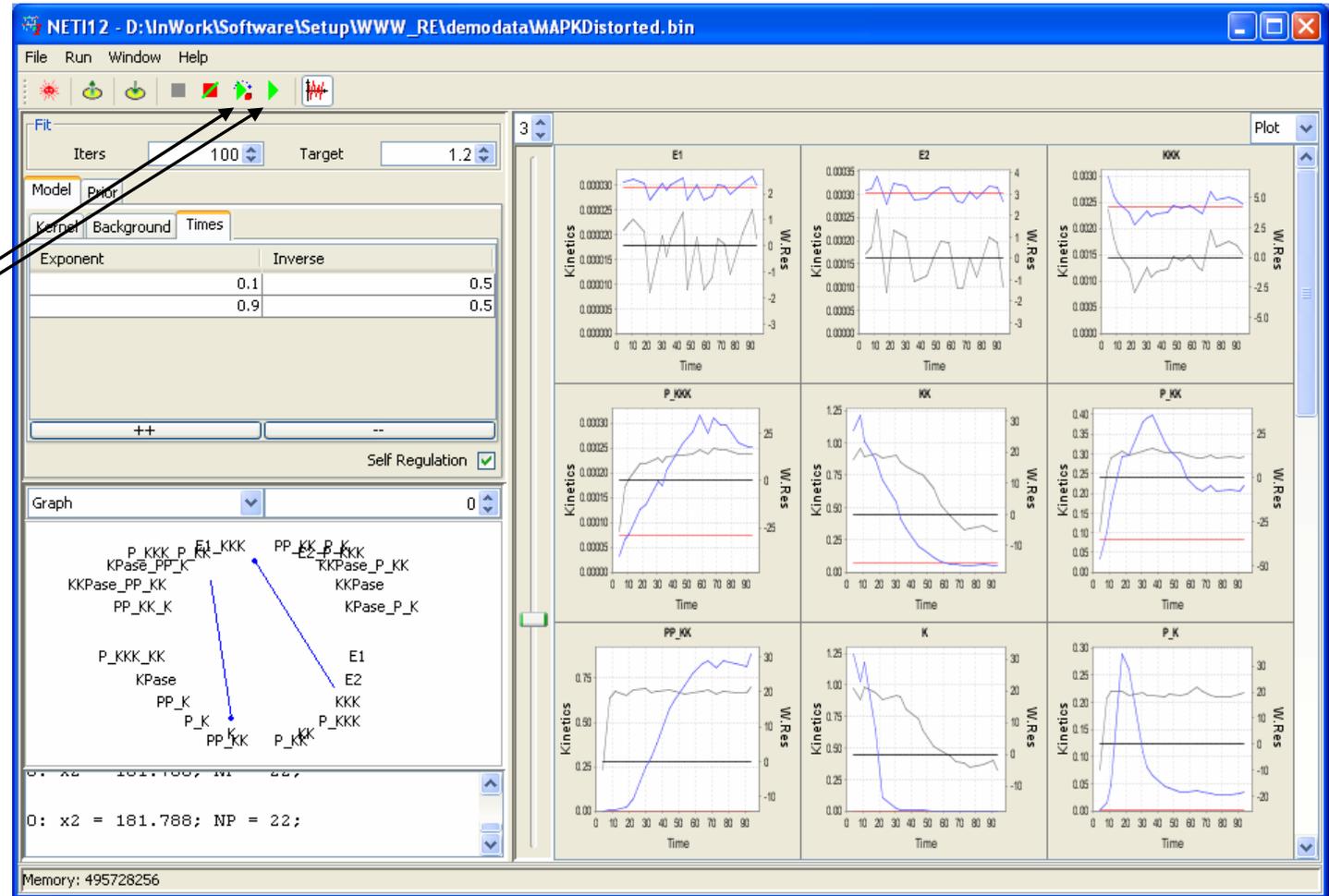
The model selection procedure, after testing the eight models defined in the “Template” table, selects the best model and copies the identification of this model into the tables “Kernel” and “Times” of the “Model” tab.

Times table



Network Reconstruction with the Selected Model

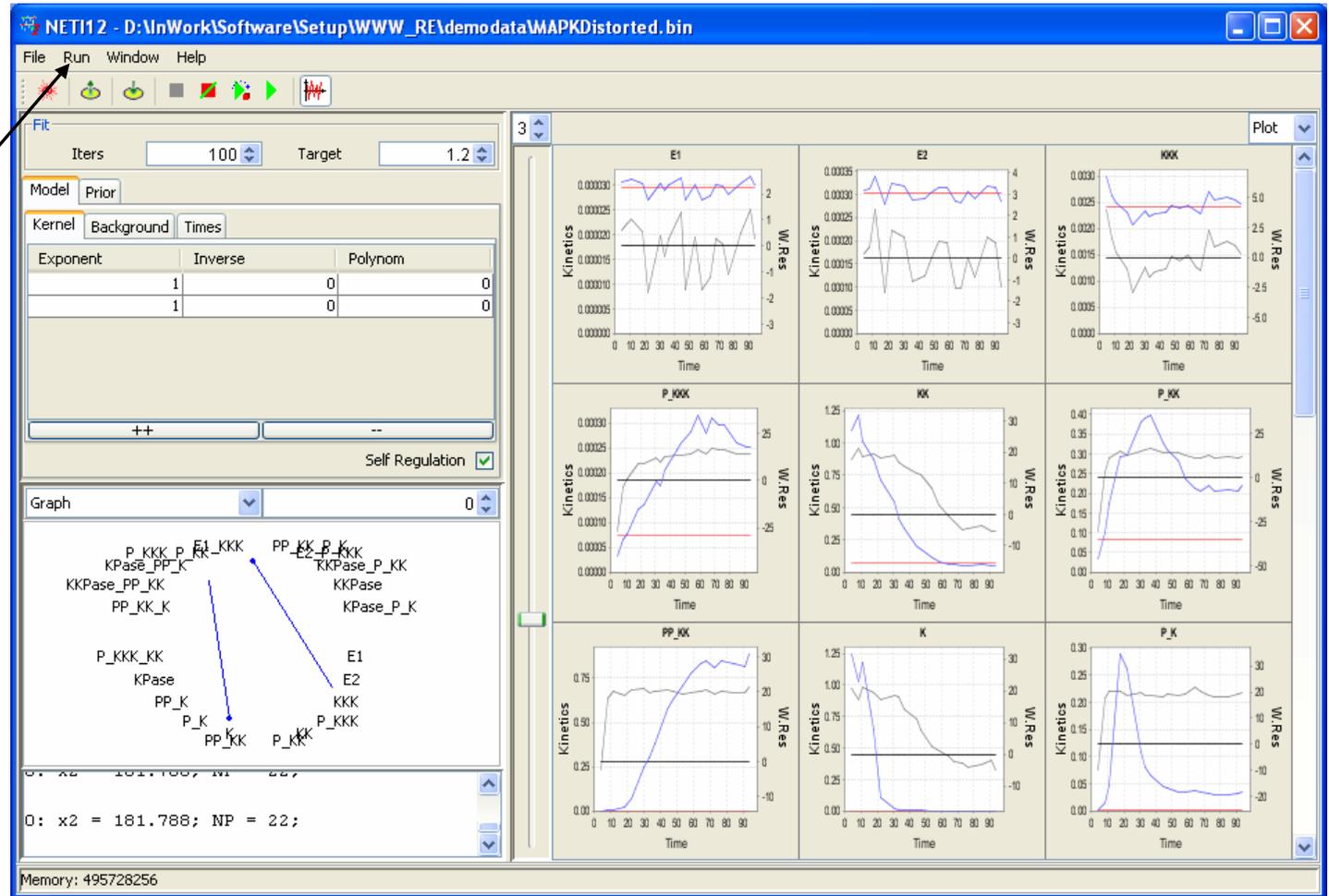
Further processing can be continued in the either “step-by-step” or automatic inference mode.



Network Simulator

Run Network Simulator

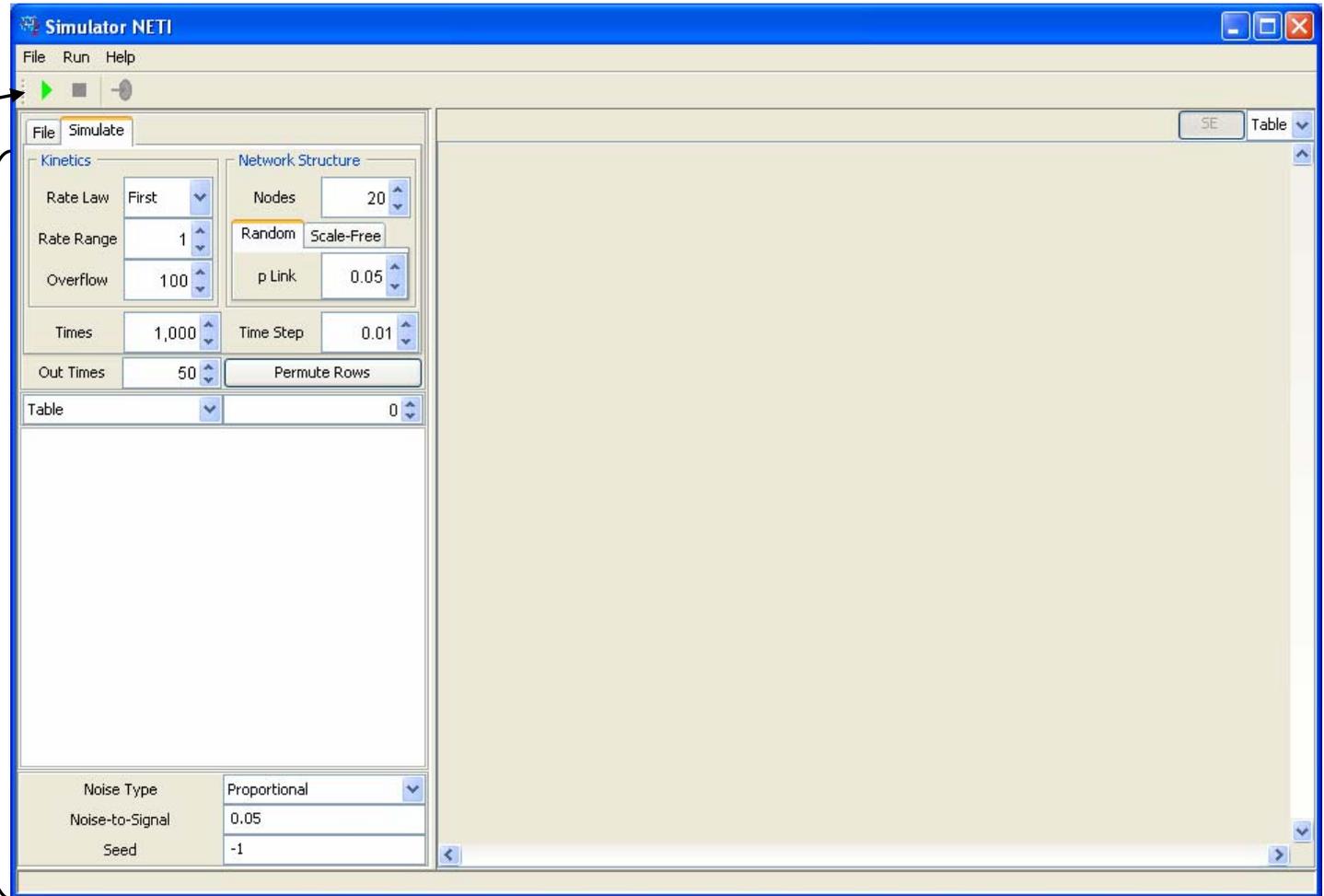
To open Network Simulator select the Menu Item “Run|Simulator”.



Network Simulator Window

To start simulations press the “Run Simulations” button from the Toolbar or select the Menu Item “Run|Run Simulations” (F5).

To simulate data the following parameters should be defined.

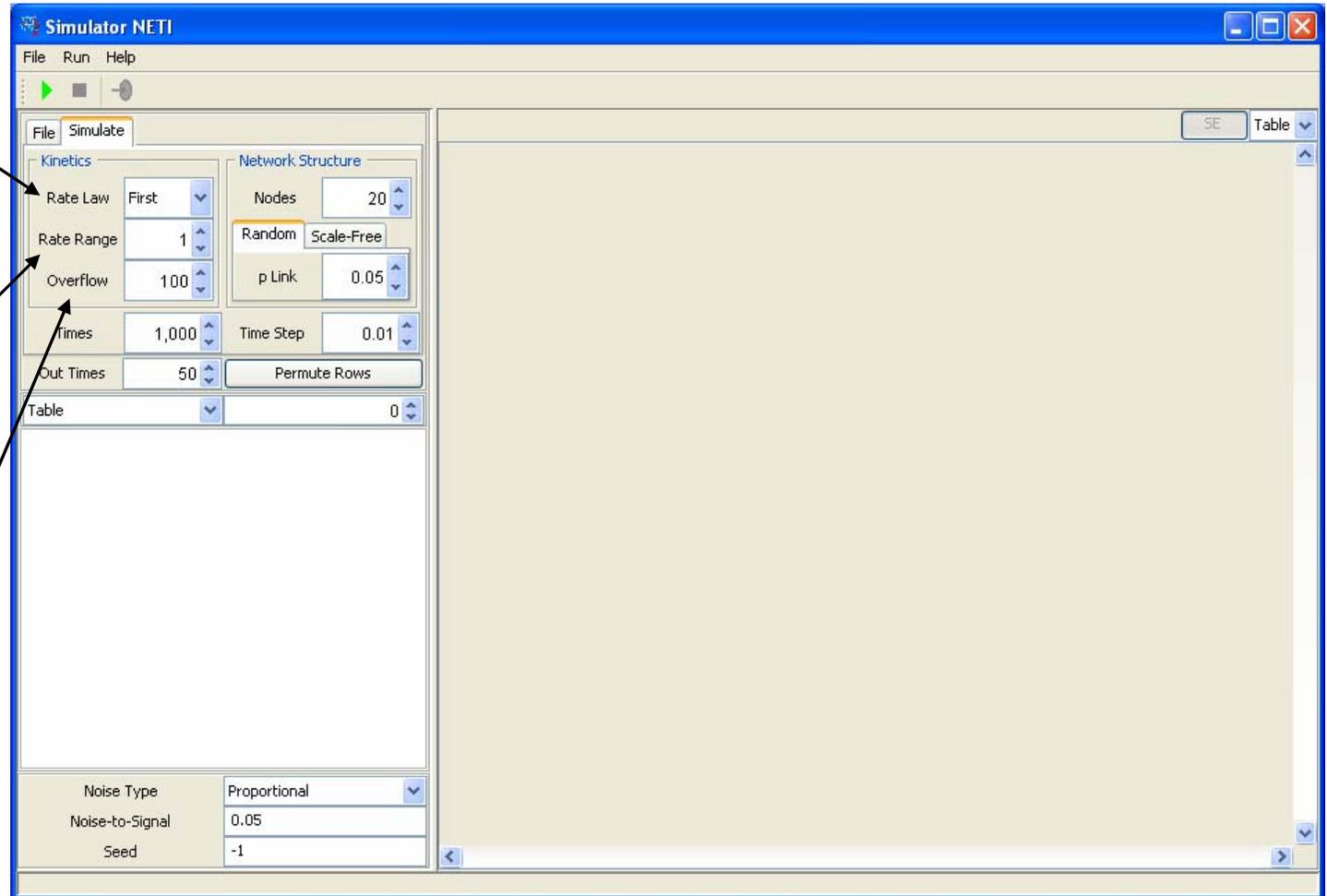


Artificial Networks (I)

Simulation model is defined by a set of ordinary differential equations with either *first-* or *second-order* kinetic *rate laws*.

The rates of kinetic equations are randomly selected from the interval: $[-Rate\ Range; +Rate\ Range]$.

Generated time series is rejected, if it exceeds *Overflow*; procedure tries to find the structure and kinetic parameters without the overflow.



Artificial Networks (II)

Number of Nodes

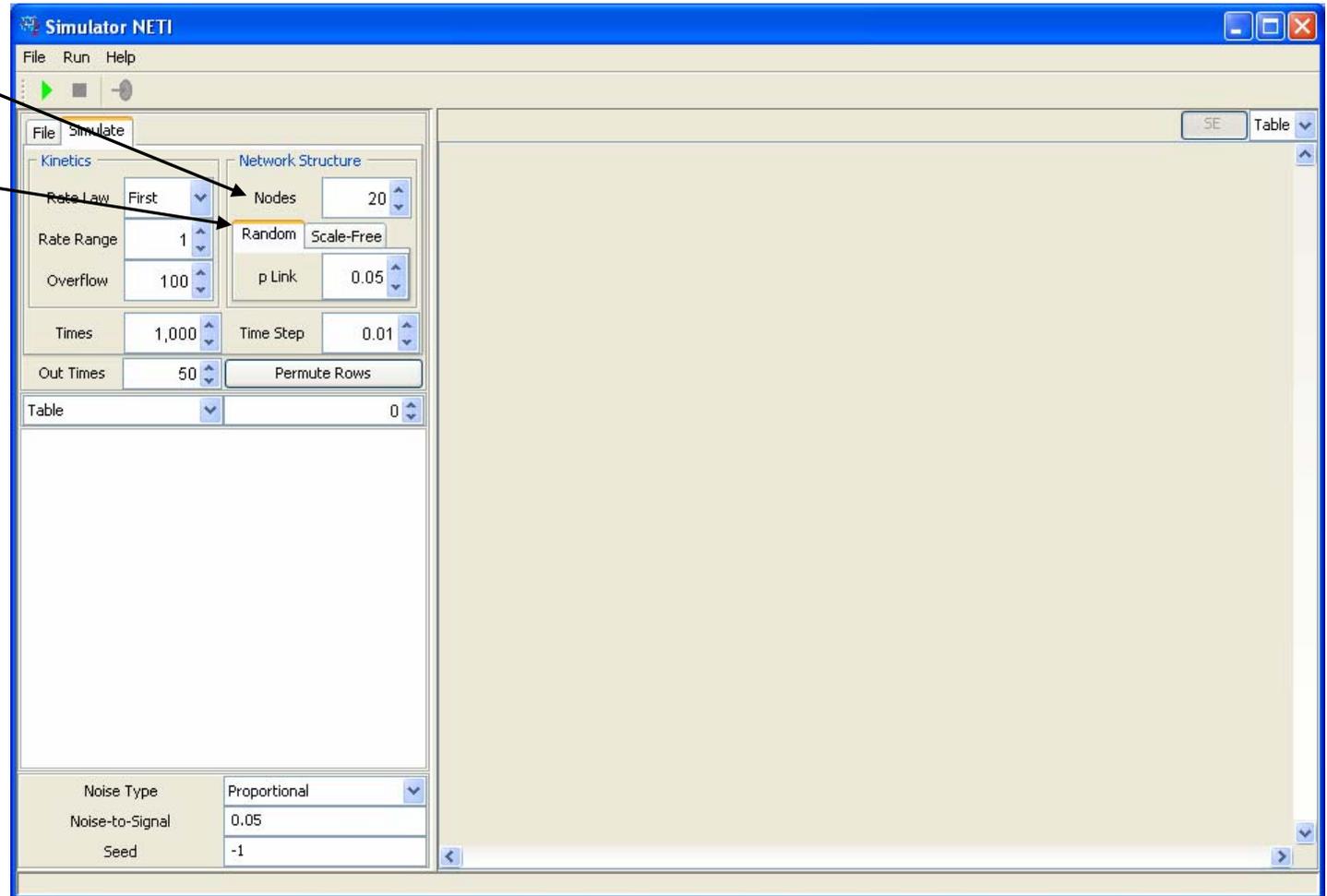
Network Topology:

Random
Scale-Free

Random topology: any two nodes are connected with the probability p independently of the other connections.

Scale-free topology: the number of links at each node is approximated by a power-law distribution $p(k) \sim k^{-\gamma}$.

Novikov E, Barillot E: *Regulatory network reconstruction using an integral additive model with flexible kernel functions*, submitted to *BMC Systems Biology*.

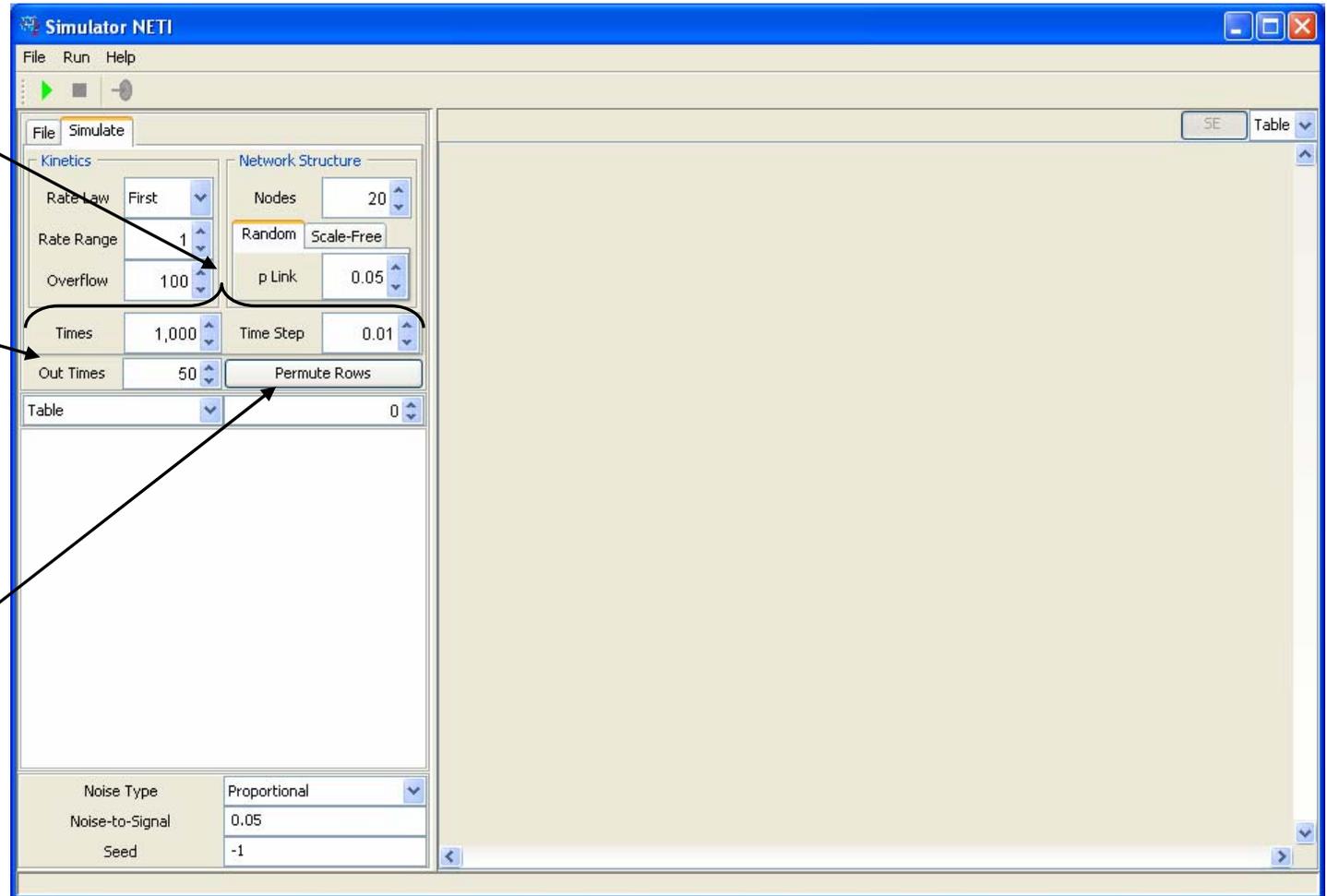


Artificial Networks (III)

Number of Time Steps (*Times*) and *Time Step* to generate idealistic time series.

Out Times defines sampling frequency. In this example, the original 1000-point time series are converted into 20 intervals of 50 points. At each interval the output time point is randomly selected.

Generation of the permuted data, i.e. when node labels are randomly assigned to generated time series.

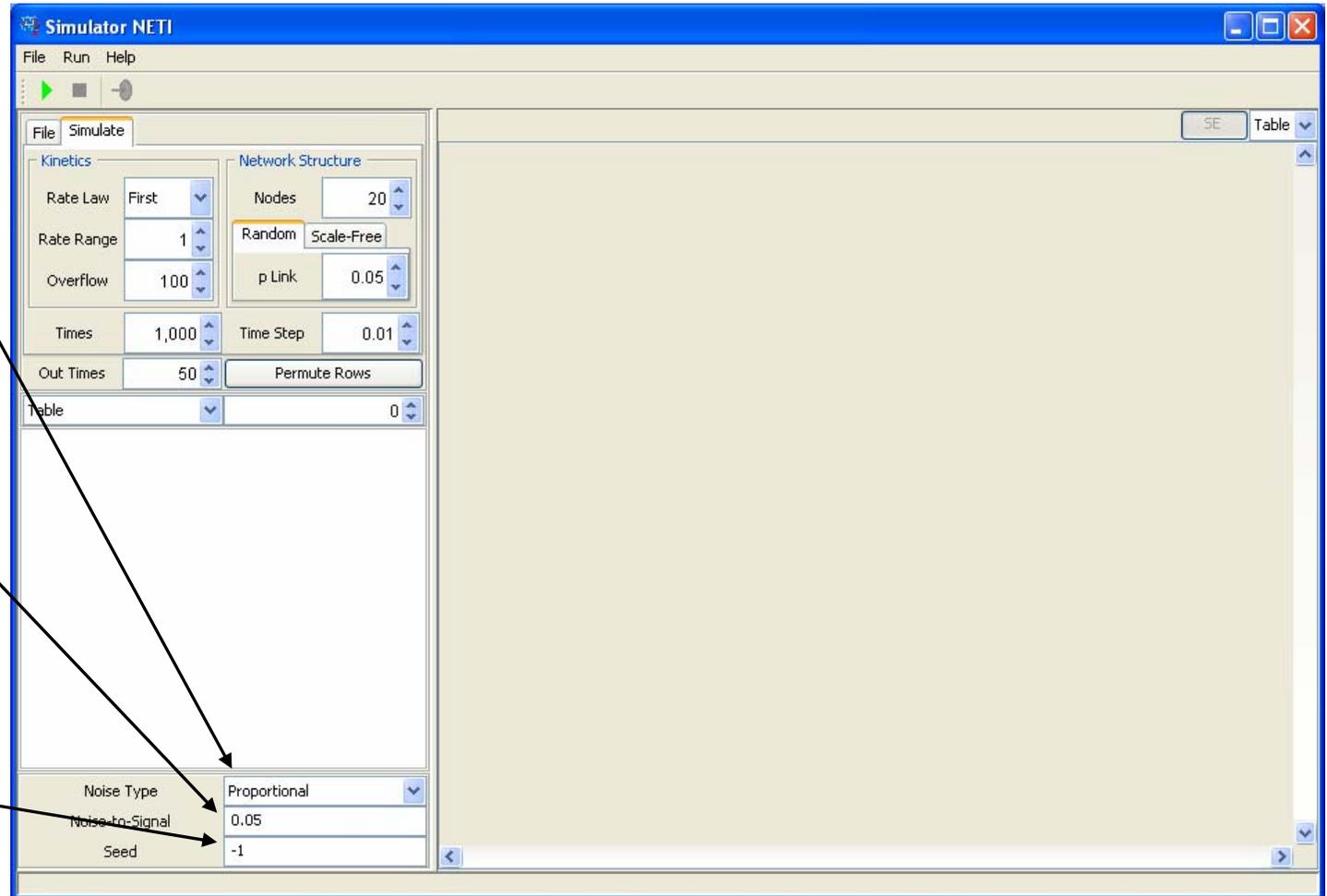


Additive Statistical Noise

Model for the standard deviation of the additive noise. It can be constant, proportional to signal, or proportional to the square root of signal.

Noise-to-signal level for the additive statistical noise. This noise is finally added to each data point.

Seed for random number generator (selection -1 as a seed will initiate the random generator with automatically (or randomly) chosen seed).

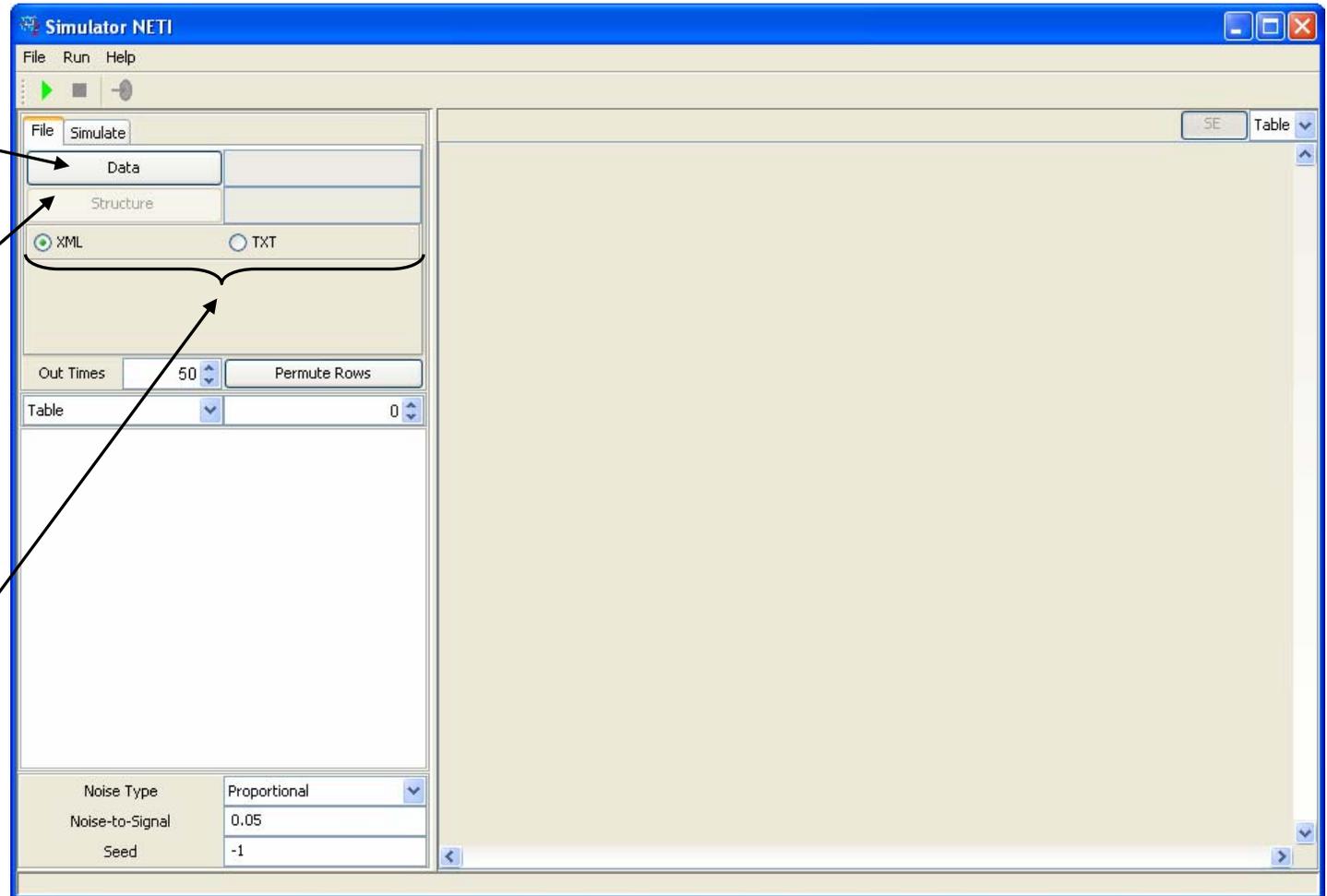


SBML Modules (I)

Simulator can import some intermediate data, typically time series generated by SBML modules.

The structure of the network can also be imported. It allows to compare the structure used in data generation with the structures obtained by the inference algorithm.

The structure can be defined by the *xml* or *txt* files.

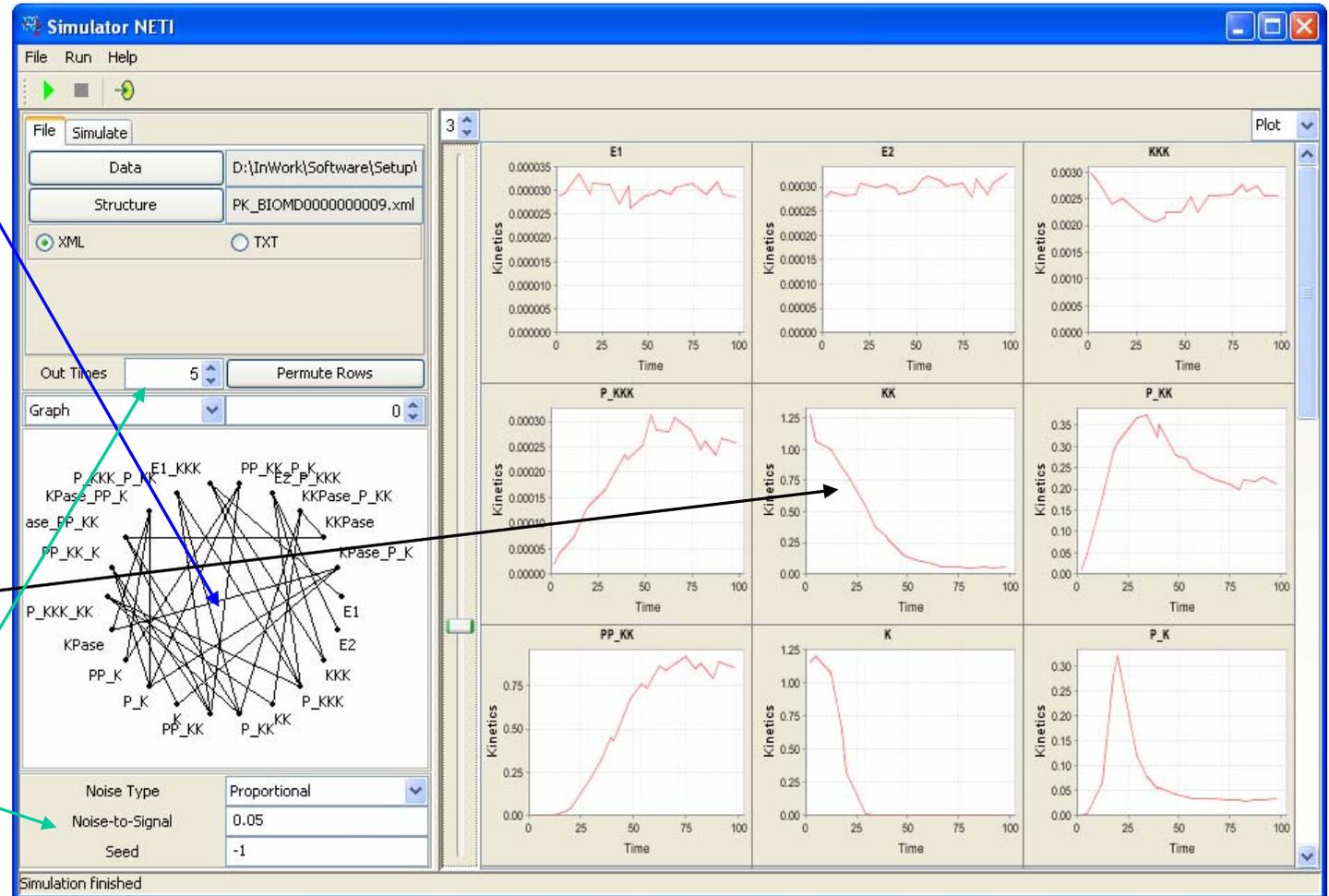


SBML Modules (II)

Imported model (MAPK).

Idealistic time series
(generated in JDesigner).

These can be further
distorted by non-
homogeneous time
sampling and adding
statistical noise.

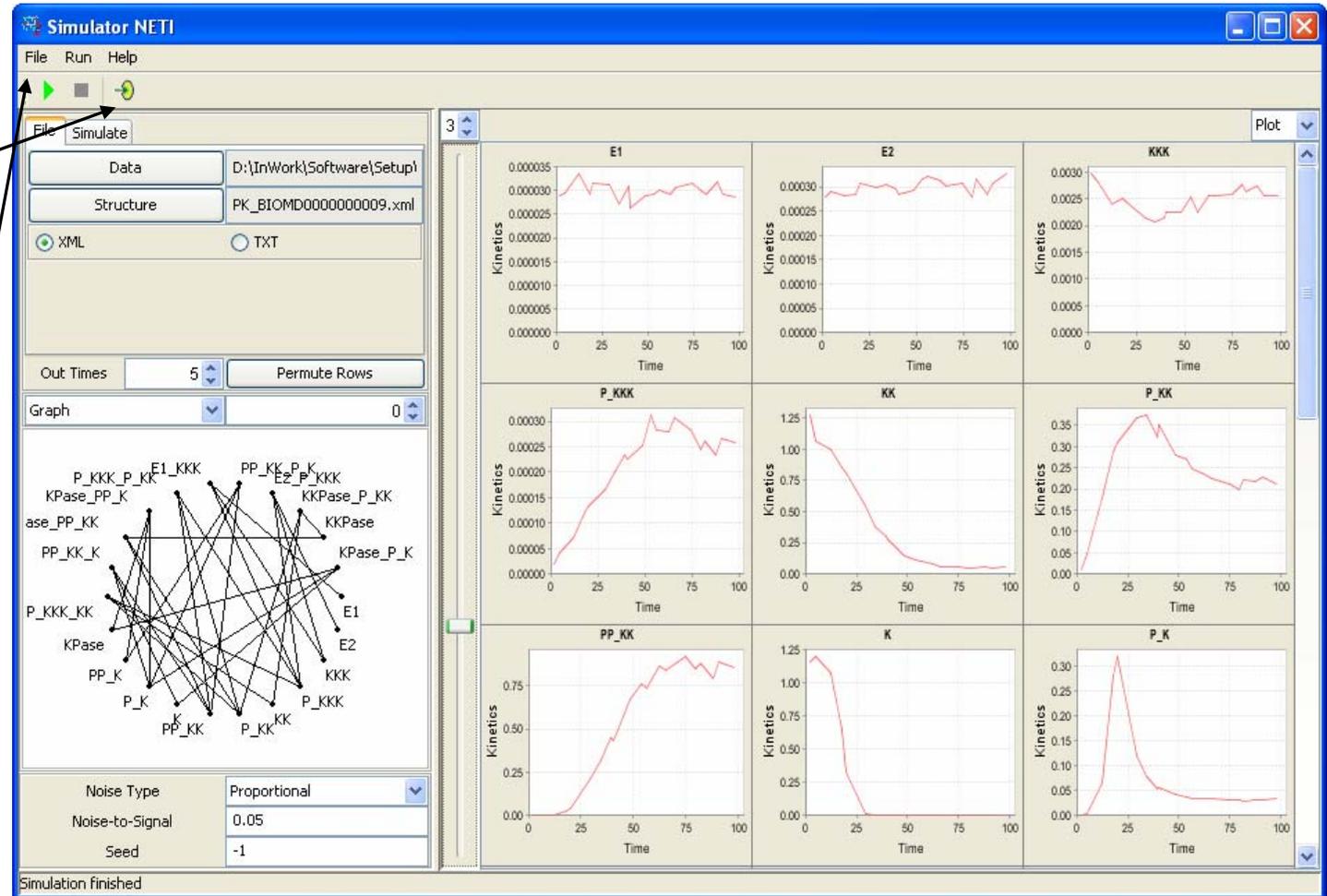


Export Data

To send the generated data in the Processing Window, use the “Send Data” button from the Toolbar or the Menu Item “File|Send Data” (Alt+→).

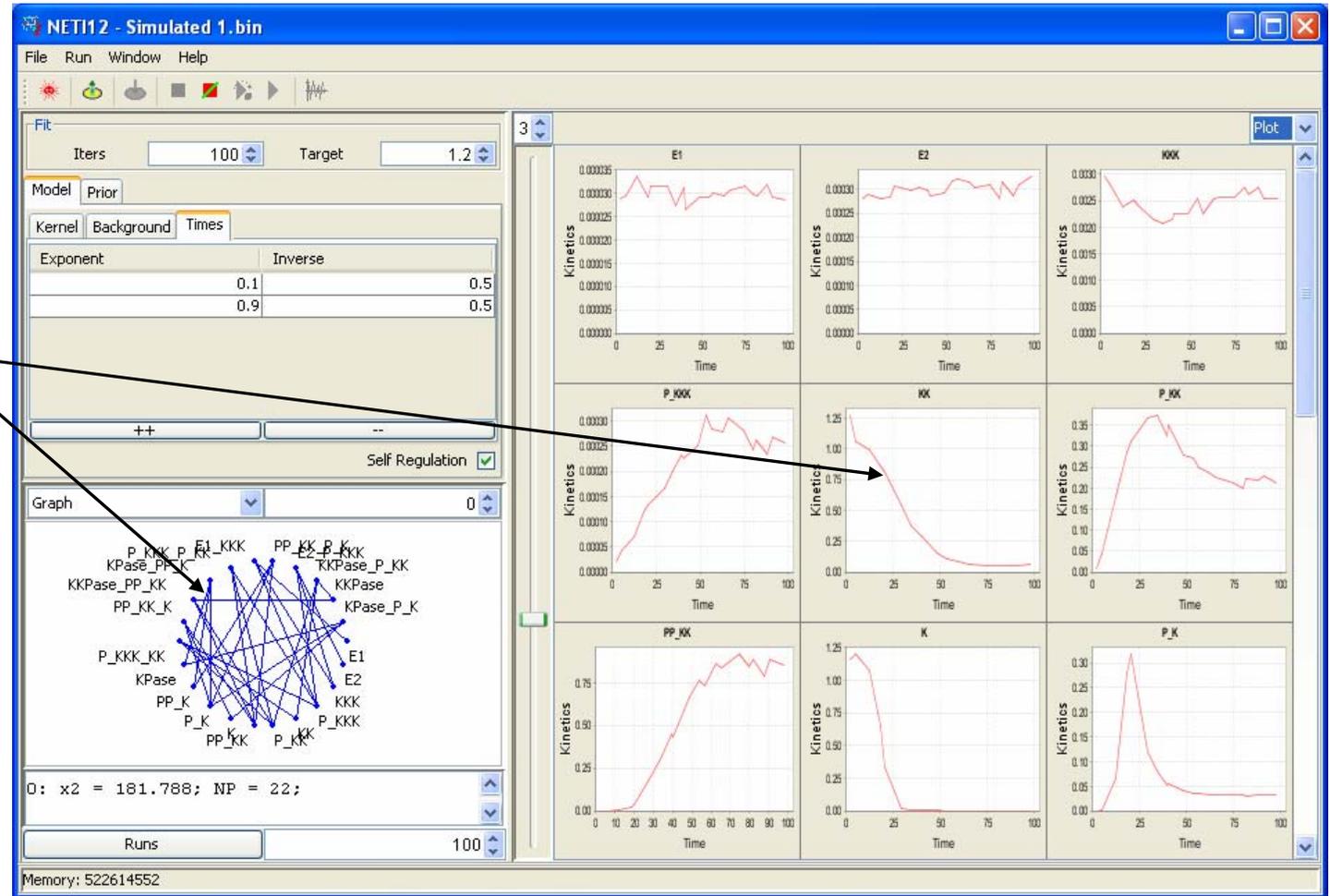
To save the generated data in the text file use the Menu Item “File|Save Data” (Ctrl+Shift+O).

To save the generated structure in the text file use the Menu Item “File|Save Data” (Ctrl+Shift+U).



Simulated Data Downloaded

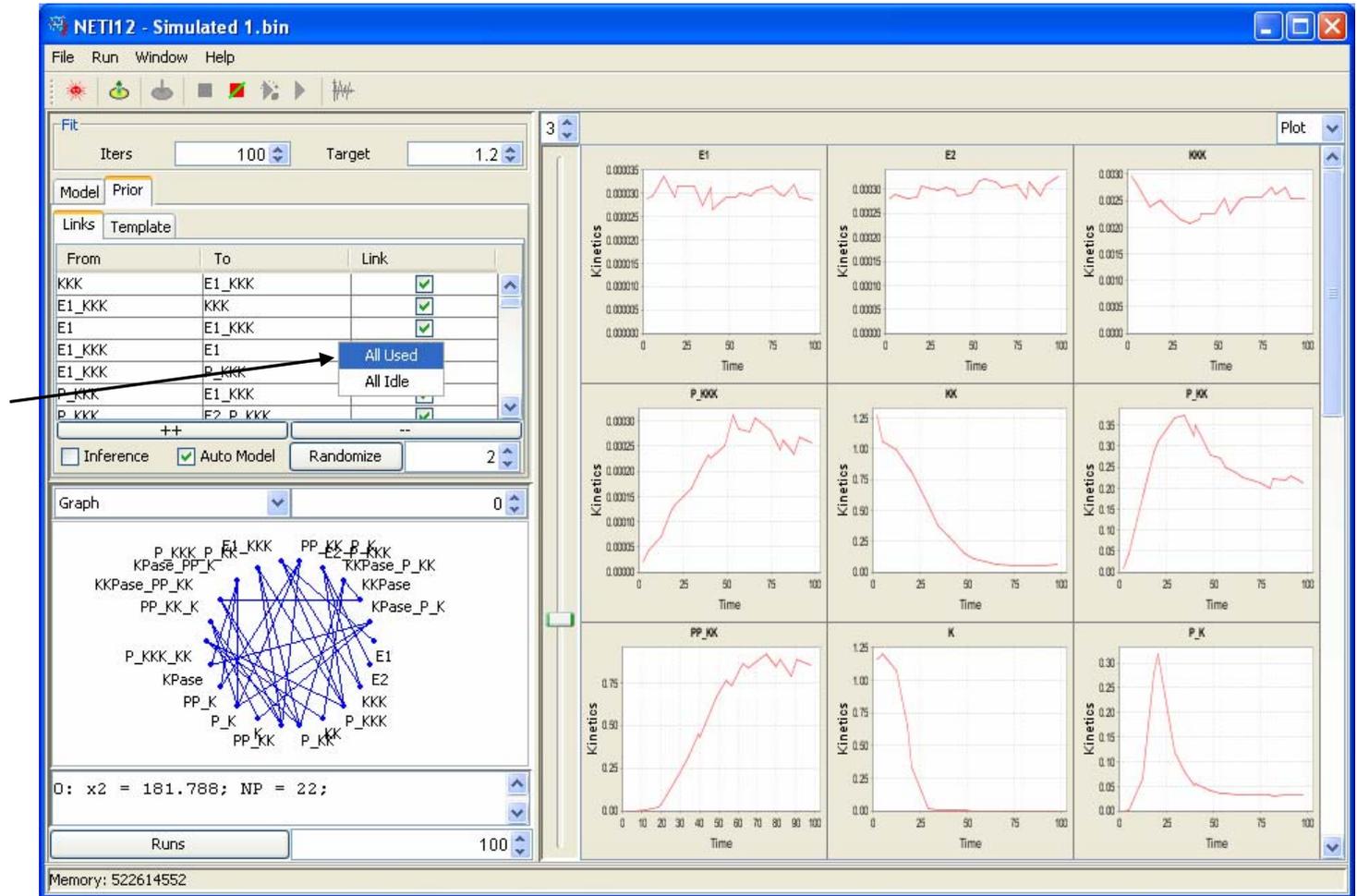
The simulated time series and the corresponding structure have been downloaded in the processing application.



Prior Links from Simulator

All links from Simulator are, by default, assumed to be unknown.

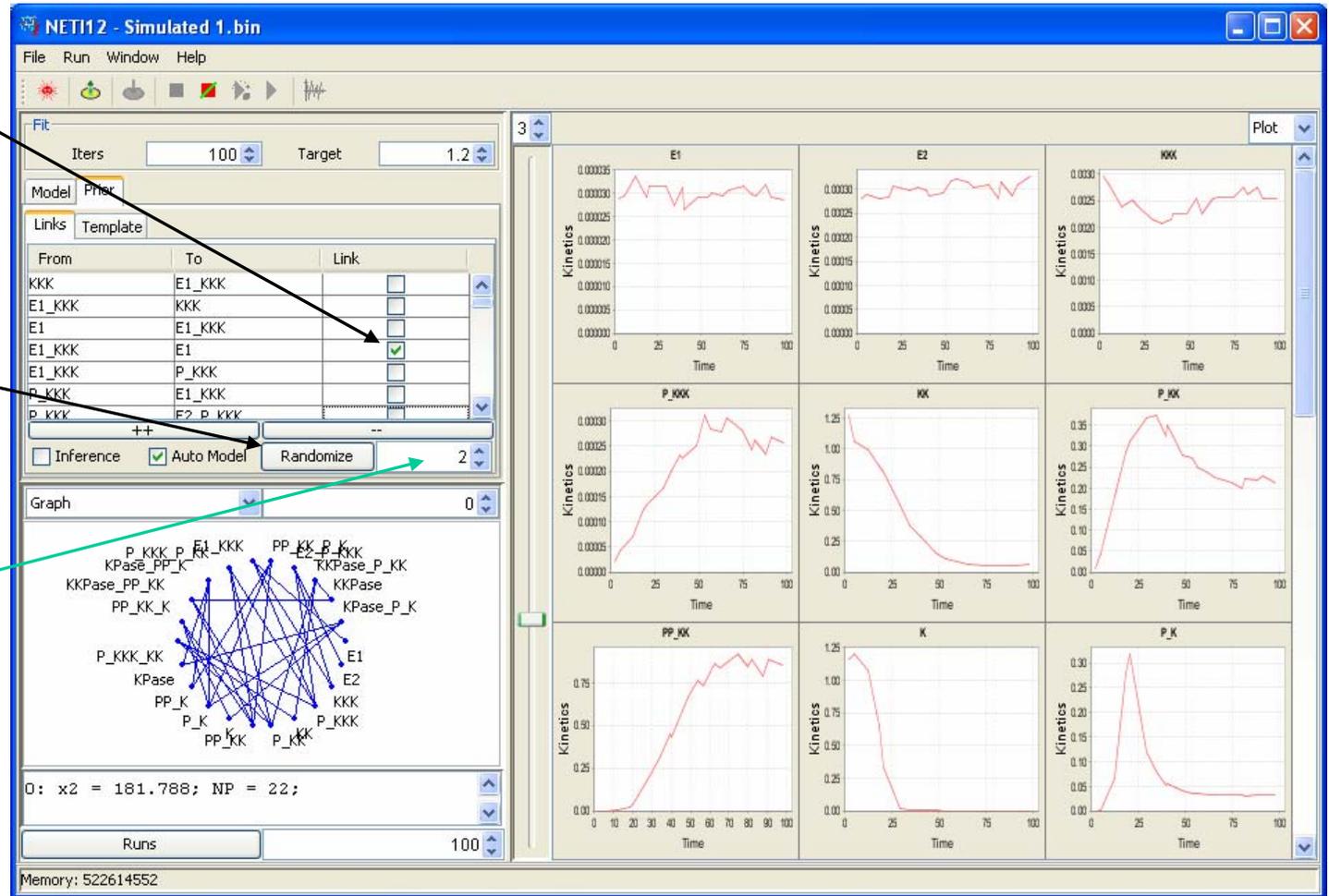
All they can be included for the adaptive model selection using the Menu Item “All Used” from the context menu.



Selected Prior Links from Simulator

Prior links can be selected manually using the check box field “Link” of the “Links” table.

Prior links can be selected randomly using the button “Randomize”; the number of selected links is defined by the corresponding spinner box.



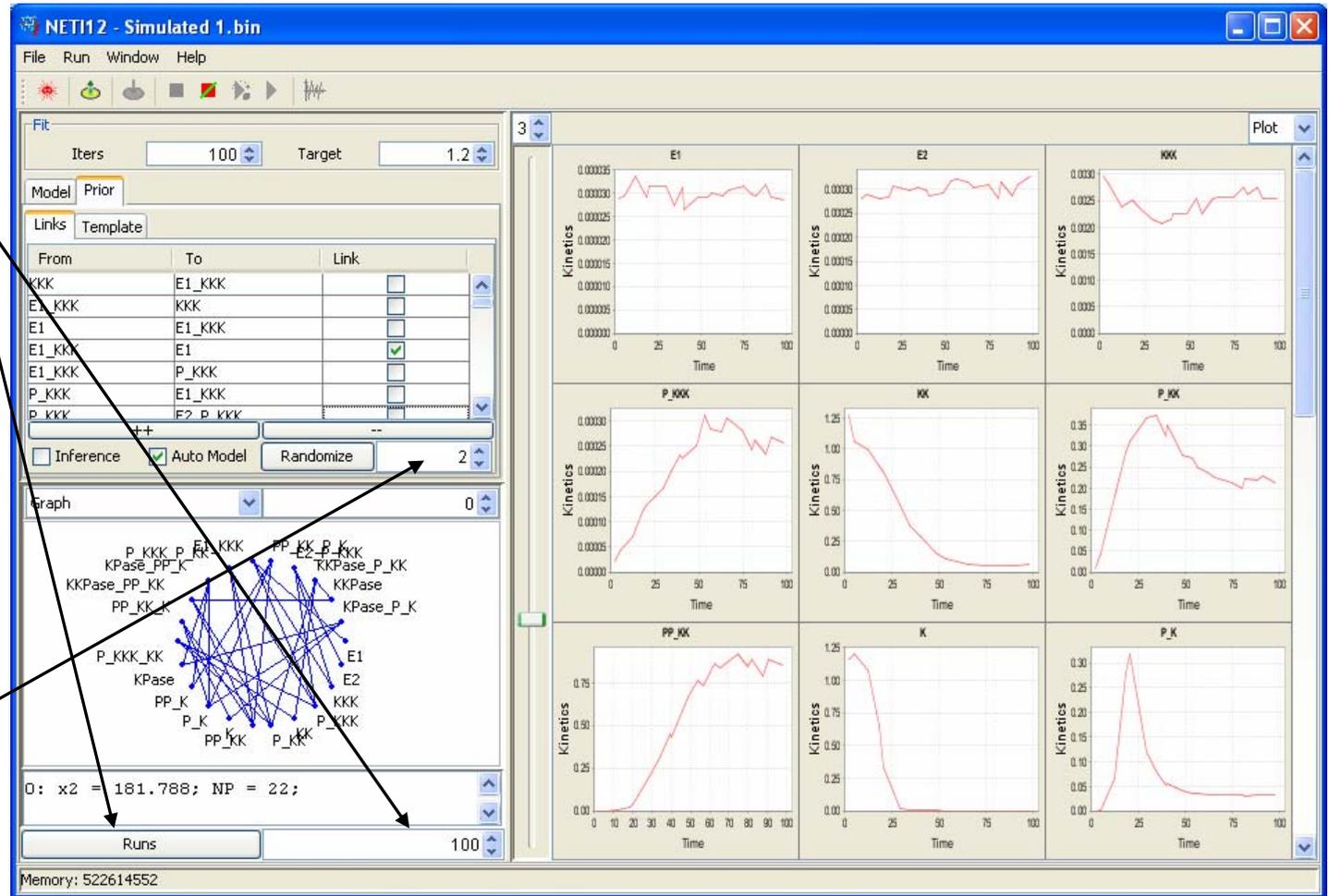
Multi-Run Simulations

Using the button “Runs” the simulation procedure with the follow-up processing is repeated 100 times to collect the statistics.

A different network structure, different link parameters, different time sampling and different noise realizations may be generated at each run.

If the properties of the network is defined by the external file (e.g. SBML) network structure, kinetic laws and kinetic parameters remained unchanged.

Different prior links are also generated at each run. The number of prior links is defined by the corresponding spinner box.



Measured Statistical Characteristics

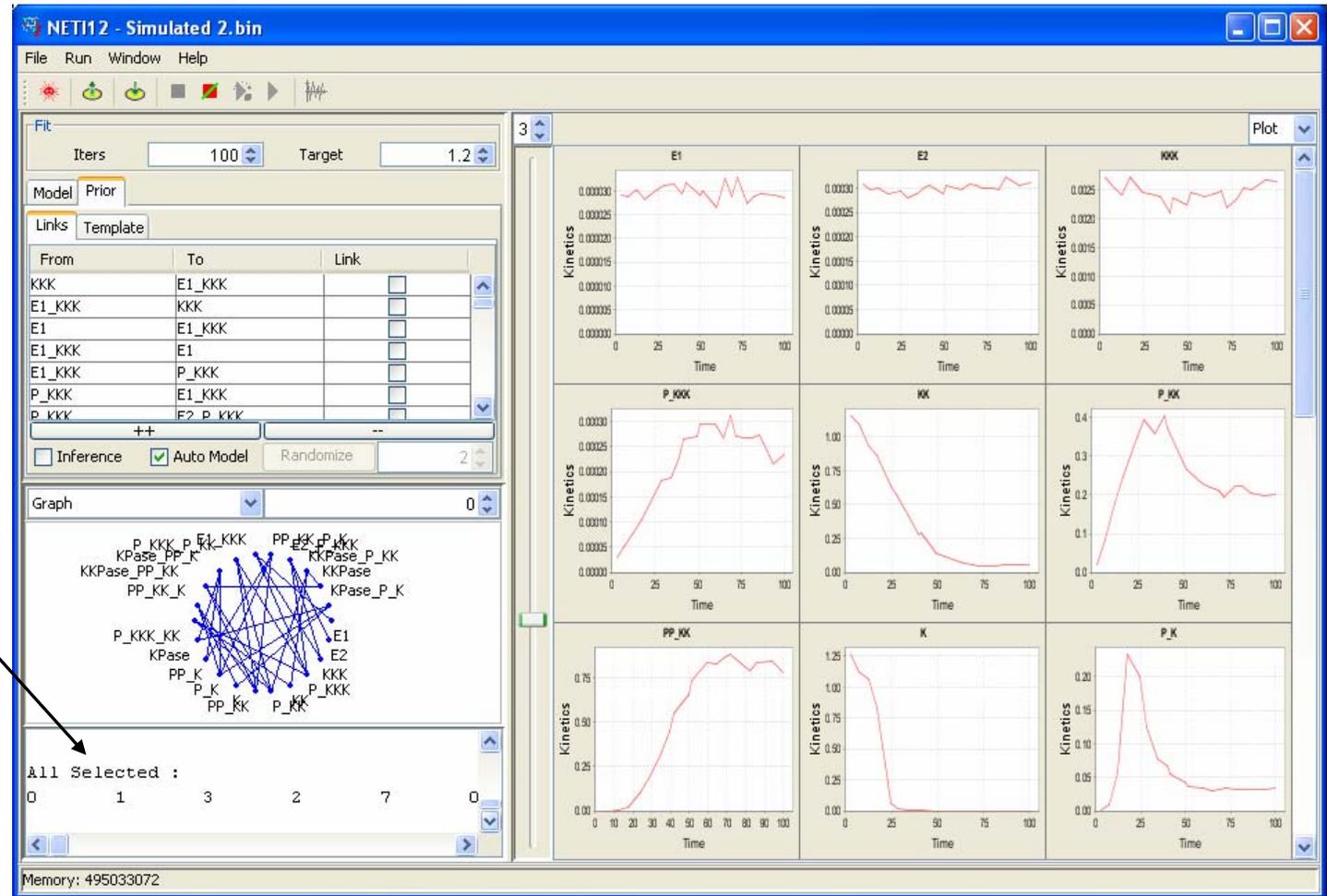
The collected statistical characteristics are the averaged dependencies on the total number of links, of:

- χ^2 criterion;
- Positive Predictive Value:
 $PPV = TP/(TP+FP)$;
- Sensitivity
 $Se = TP/(TP+FN)$;

The results in the log window:

- The selected models at each run: model numbering corresponds to the Template table;

Novikov E, Barillot E: *Model selection in the reconstruction of regulatory networks from time-series data*, submitted to *Algorithms for Molecular Biology*.



Save Statistical Characteristics

The statistical dependencies can be exported using the “Save Analysis ...” button from the Toolbar or the Menu Item “File|Save|Analysis ...” (Ctrl+S).

