

## Supplement A

# MCMC\_CLIB – An advanced MCMC sampling package for ode models

Kramer, Andrei and Stathopoulos, Vasileios

May 27, 2014

This Supplement provides some illustration for the chosen model. It shows, how a parameter sample is related to a model fit. This model has been chosen as an example for a hard problem: the model is large and cannot fit all of the observations simultaneously; the measurements are sparse, uncertain, and are provided in arbitrary units. Note: The reason for the mismatch will not be discussed. The data might have been obtained under erroneous assumptions which are consequently not reflected in the modeling setup or alternatively, the model is incorrect.

### 1. Notation of uncertainty

Throughout all documentation we use this notation for errors:

$$v = 1.71(12) := 1.71 \pm 0.12 \quad a = 42(3) \times 10^6 := (42 \pm 3) \times 10^6, \quad (1)$$

where we write only the significant digits of the error estimate. The magnitude of the error is communicated by rounding the reported number to the appropriate number of digits. For example, given  $a = \alpha \pm \delta$

$$\alpha = 3.14159265 \text{ kg}, \quad \delta = 0.00014672 \text{ kg}, \quad (2)$$

will be reported as  $a = 3.14159(15) \text{ kg}$ , (3)

as opposed to  $a = (3.14159 \pm 0.00015) \text{ kg}$ . (4)

This is shorter, better suited for tables and has advantages when reporting numbers with units of measurement.

tuning duration	1000 points
target acceptance $a_0$	0.500
observed acceptance $a$	0.498
auto-correlation length $\tau_{\text{int},l}$	340(30)
failed Likelihood evaluations [CVODES]	16

**Table 1:** Sample Properties.

## 2. Calculation of uncertainty

All analysis of uncertainty was performed in accordance with [1] using the matlab script provided by the author. The used method calculates a naïve variance of an estimate  $a$  and increases it appropriately using the auto-correlation length  $\tau_{\text{int},a} \geq 0.5$ . This quantity is calculated by the script in a numerically stable way using the correlation function  $\Gamma$  ([1], Eq. 2). A notion of effective sample size and effective sampling speed can also be derived from the concepts described in this publication. The resulting estimate of uncertainty can be compared to a simple standard deviation of an uncorrelated sample with the same effective size

$$N_{\text{eff.}} = \frac{N}{2\tau_{\text{int},a}}, \quad (5)$$

where  $N$  is the actual sample size. For uncorrelated samples, these two numbers match.

## 3. Two different ways to sample the provided model

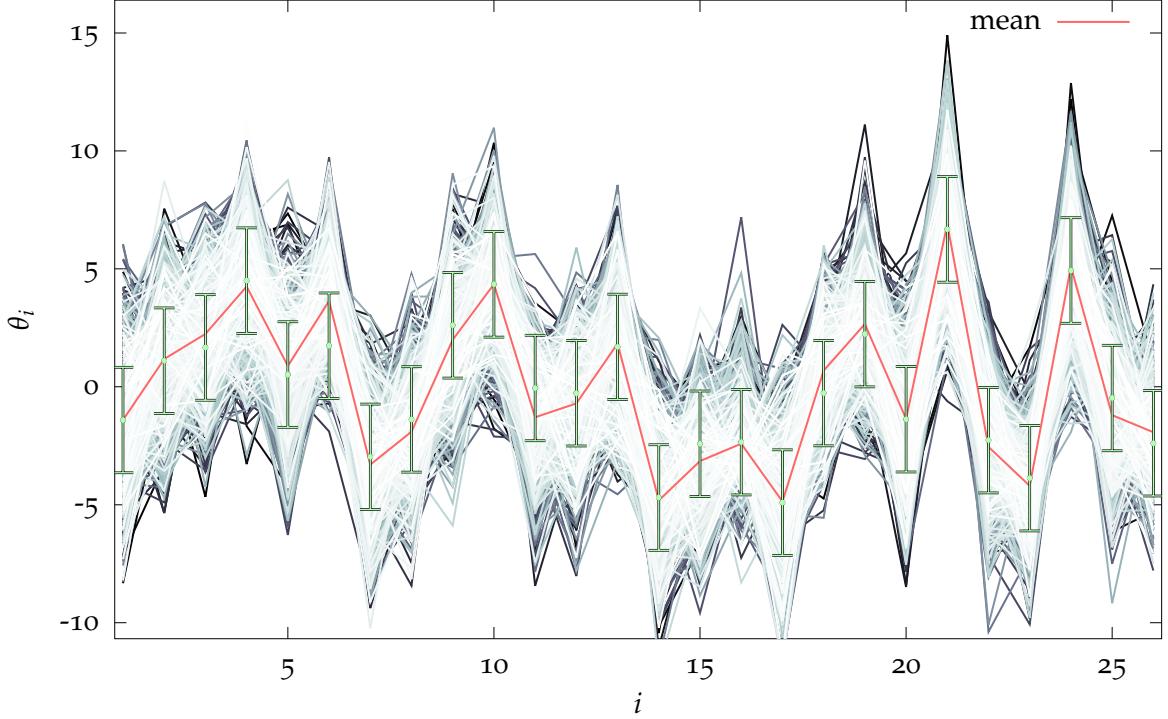
The sample was obtained in  $t_s = 267385$  s (about 74 h). Considering the auto-correlation  $\tau_{\text{int},l} = 340(30)$ , the effective sampling speed is:

$$v = \frac{N}{2\tau_{\text{int},l} t_s} = 5.50(49) \times 10^{-3} \text{ s}^{-1}. \quad (6)$$

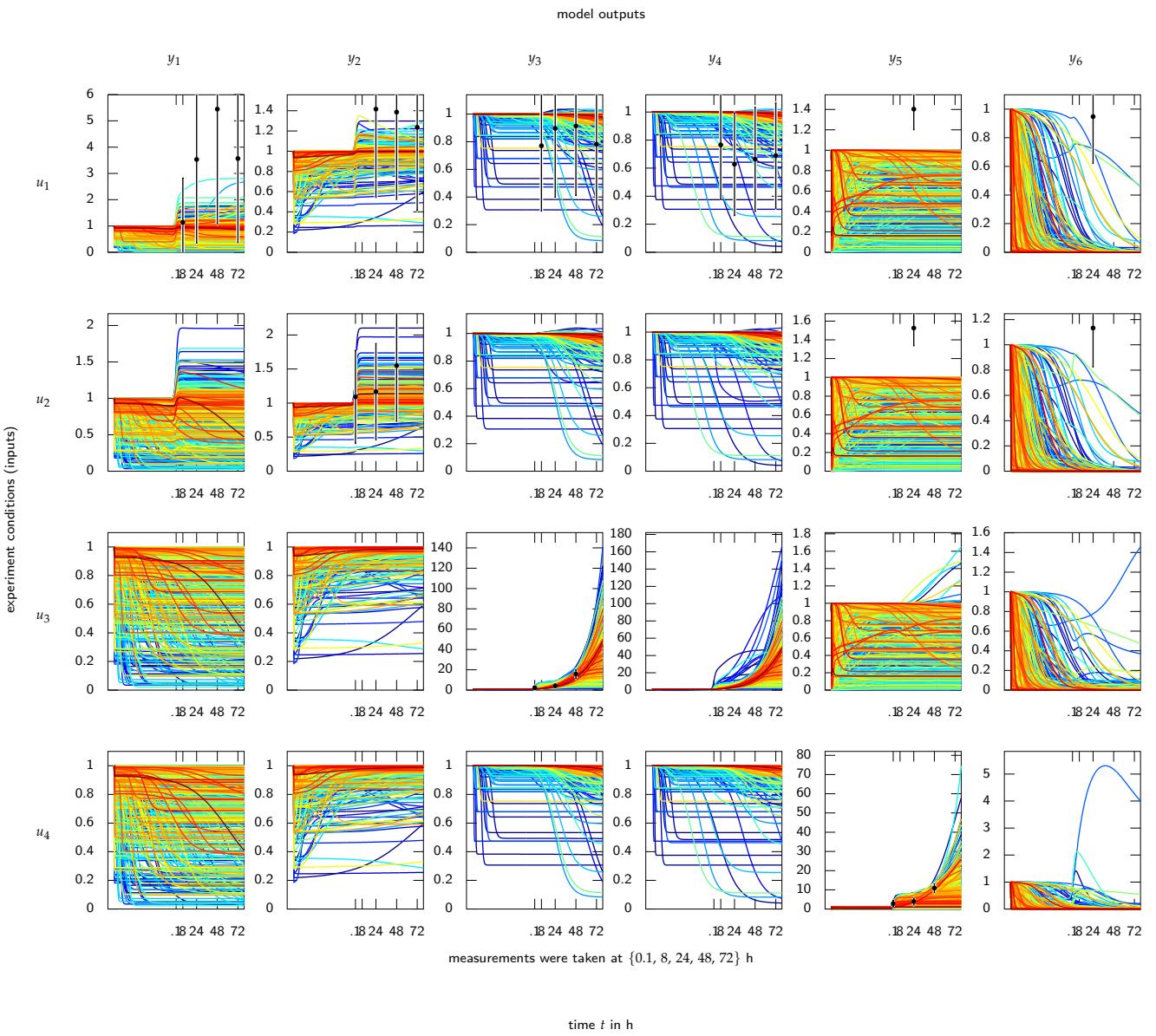
The properties recorded during sampling are summarised in Table 1. The sample is plotted in Figure 1 using parallel coordinates (each sampled parameter vector is shown component wise, as a line). Figure 2 shows the output trajectories obtained from forward simulations using the sampled parameters. There is partial agreement between model and data (black errorbars).

### 3.1. Alternative sampling approach

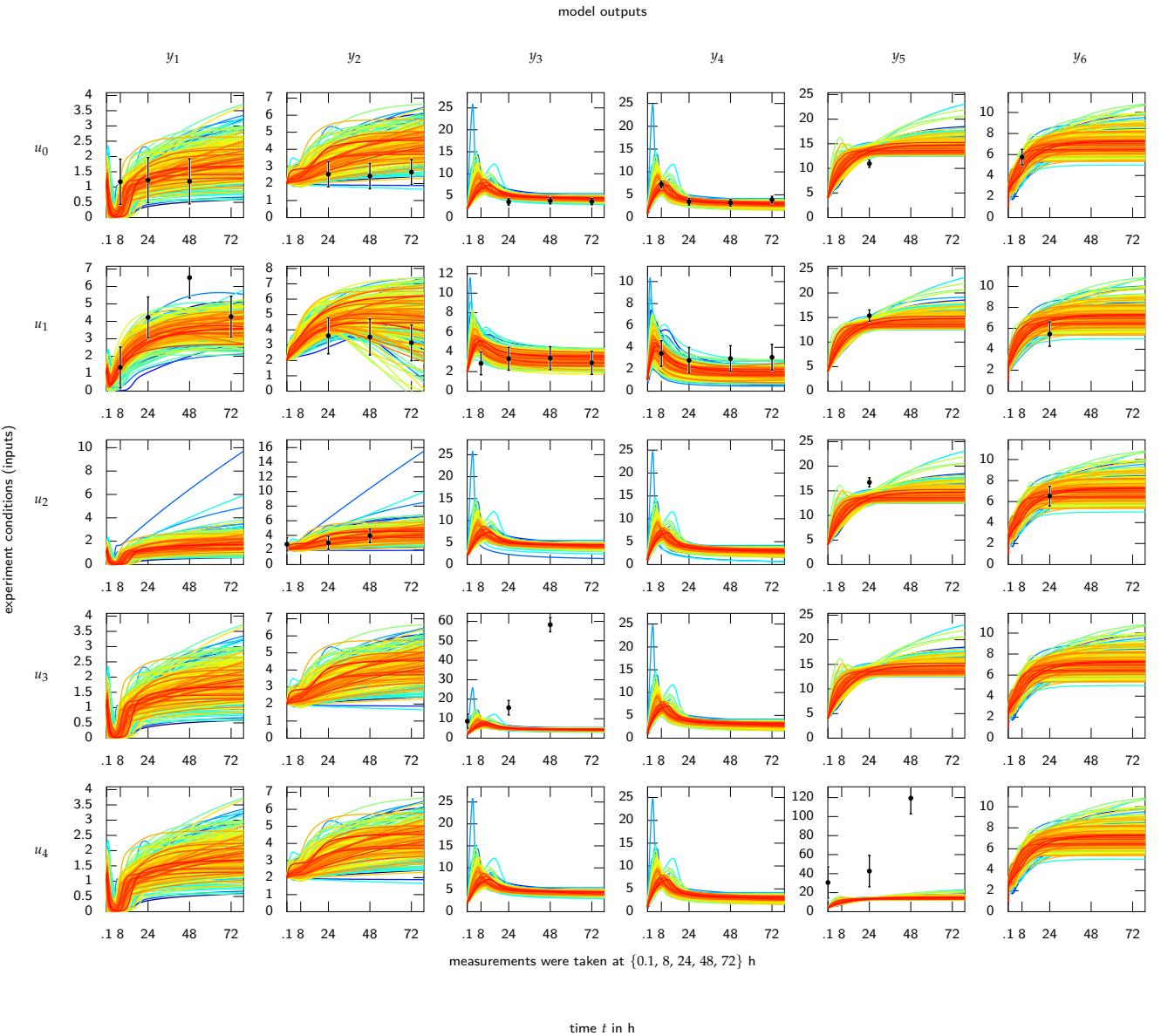
As in the given example, it is often the case that observations  $y_i(t_j, \theta, u_k)$  are reported in arbitrary units. However, we do not fit the unknown data scaling constant. We expect this scaling to be different when the experiment is repeated. In biological experiments



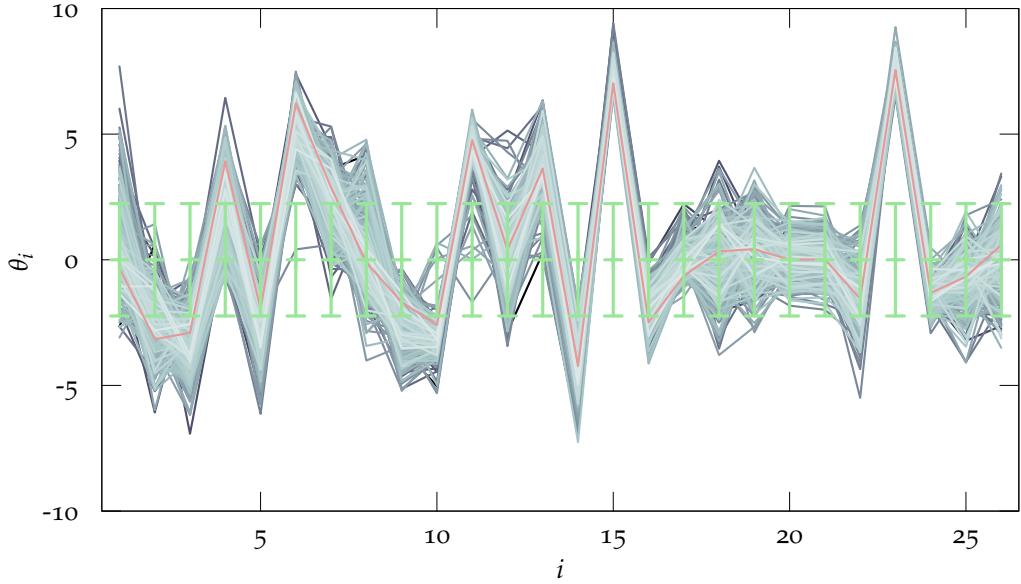
**Figure 1:** The Posterior Sample of size  $N = 1 \times 10^6$ , as obtained from the MCMC\_CLIB; thinned out using the integrated autocorrelation length  $\tau_{\text{int},l} = 340(30)$  (estimated using the log-posterior  $l$  as observable). The red line indicates the maximum posterior estimate, while the green errorbars indicate the used prior distribution (Gaussian). The colorpalette reflects the value of the log-posterior for each sampled point. Here, we have incorporated our prior knowledge about the parameter values in the prior parameter density. Yet, we have chosen to let the prior span several orders of magnitude.



**Figure 2:** output trajectories per parameter sample member; to reduce the number of lines these trajectory lines are once again plotted for every  $3534^{\text{th}}$  ( $2(\tau_{\text{int},l} + \delta_\tau)$ ) parameter in the parameter sample. The green, vertical errorbars represent the data and its measurement error estimation. The data is sparse (not all outputs were observed at all measurement instances. The model does not fit all data points.)



**Figure 3:** Treating the data as absolute, the fit changes ( $u_0$  is the setup for the reference experiment). Some data points get an improved fit, some system behaviour cannot be reproduced. Note that this treatment of the data is incorrect and is done for illustration purposes; we show the difference between fitting absolute and relative data. To obtain these Trajectories, we have used a different, uninformative, iid, Gaussian prior during sampling. The resulting parameter sample is shown in Figure 4



**Figure 4:** The parameters obtained from the assumption that provided data is absolute (molecule numbers or mols). Here, an uninformative prior was used.

this unknown scaling will reflect conditions like dilution of bacteria samples, exposure time for western blotting and similar, which does change with repetition. Therefore, we are not interested in the values of these scaling constants<sup>1</sup>. To deal with this issue, we take the ratios between the observations and some reference experiment results, which are often called «control» by biologists. This reference measurement is performed under identical observation conditions and has the same scaling; thus, the scaling cancels in the ratios. But, since the *simulation* of this reference value using the model depends on the parameters  $\rho = \exp(\theta)$ , we have a parameter dependant scaling:

$$\tilde{y}_{ijk} = \frac{y_i(t_j, \theta, u_k)}{y_i(t_j, \theta, u_0)}, \quad (7)$$

where  $u_0$  characterises the reference conditions. This results in dramatically different fits. This method effectively eliminates all observation parameters and results in a smaller MCMC problem than trying to estimate the observational scaling.

Just for demonstration purposes, we ignore the source of the data and pretend, that the measurements are absolute (gauged). The reference experiment has been included as input line  $u_0$ . Although this is wrong, the problem size is the same and the model can still fit some of the data, but note that different data entries are fitted well compared to the relative data fit. Figure 3 shows this alternative fit. Specifically the model cannot fit the dynamic behaviour in experiment  $(u_3, y_4)$  and  $(u_4, y_5)$  which it

---

<sup>1</sup>see the mathematical documentation for details on this topic

could before, see Figure 2. Most problematic is the fact that the parameter fit cannot be reproduced by repeating the data acquisition experiment; the scaling will be different.

## References

- [1] Ulli Wolff. “Monte Carlo errors with less errors”. In: *arXiv.org* ().

## A. Model Setup

**Listing 1:** Model

```

1  <?xml version="1.0" ?>
<VectorField Name="ODEmodel11S26P4U" Description="A model for testing purposes">
<Parameter Name="theta_1" DefaultValue="1.0"/>
<Parameter Name="theta_2" DefaultValue="1.0"/>
<Parameter Name="theta_3" DefaultValue="1.0"/>
6  <Parameter Name="theta_4" DefaultValue="1.0"/>
<Parameter Name="theta_5" DefaultValue="1.0"/>
<Parameter Name="theta_6" DefaultValue="1.0"/>
<Parameter Name="theta_7" DefaultValue="1.0"/>
<Parameter Name="theta_8" DefaultValue="1.0"/>
11 <Parameter Name="theta_9" DefaultValue="1.0"/>
<Parameter Name="theta_10" DefaultValue="1.0"/>
<Parameter Name="theta_11" DefaultValue="1.0"/>
<Parameter Name="theta_12" DefaultValue="1.0"/>
<Parameter Name="theta_13" DefaultValue="1.0"/>
16 <Parameter Name="theta_14" DefaultValue="1.0"/>
<Parameter Name="theta_15" DefaultValue="1.0"/>
<Parameter Name="theta_16" DefaultValue="1.0"/>
<Parameter Name="theta_17" DefaultValue="1.0"/>
<Parameter Name="theta_18" DefaultValue="1.0"/>
21 <Parameter Name="theta_19" DefaultValue="1.0"/>
<Parameter Name="theta_20" DefaultValue="1.0"/>
<Parameter Name="theta_21" DefaultValue="1.0"/>
<Parameter Name="theta_22" DefaultValue="1.0"/>
<Parameter Name="theta_23" DefaultValue="1.0"/>
26 <Parameter Name="theta_24" DefaultValue="1.0"/>
<Parameter Name="theta_25" DefaultValue="1.0"/>
<Parameter Name="theta_26" DefaultValue="1.0"/>
<Parameter Name="u1" DefaultValue="0.0"/>
<Parameter Name="u2" DefaultValue="0.0"/>
31 <Parameter Name="u3" DefaultValue="0.0"/>
<Parameter Name="u4" DefaultValue="0.0"/>
<Expression Name="logistic" Formula="1.0/(1+exp(-t))"/>
<Expression Name="U1t" Formula="u1*logistic"/>
<Expression Name="U2t" Formula="u2*logistic"/>
36 <Expression Name="U3t" Formula="u3*logistic"/>
```

```

<Expression Name="U4t" Formula="u4*logistic"/>
<Expression Name="S1" Formula="X4+X5"/>
<Expression Name="S2" Formula="(theta_22)*X9*X7"/>
<Expression Name="A1" Formula="(theta_1)*X2"/>
41 <Expression Name="A2" Formula="(theta_2)*X1"/>
<Expression Name="A3" Formula="(theta_3)*X3*X2"/>
<Expression Name="A4" Formula="(theta_4)*X4"/>
<Expression Name="A5" Formula="(theta_10)*X4"/>
<Expression Name="A51" Formula="(theta_10)*X5"/>
46 <Expression Name="A6" Formula="(theta_13)*X7"/>
<Expression Name="A7" Formula="(theta_14)*S1*X6"/>
<Expression Name="A8" Formula="(theta_19)*X10*S1"/>
<Expression Name="A9" Formula="(theta_20)*X8"/>
<Expression Name="A10" Formula="(theta_22)*X11*X7"/>
51 <Expression Name="A11" Formula="(theta_21)*X9"/>
<StateVariable Name="X1"
               DefaultInitialCondition="1000.0"
               Formula="A1-A2+(theta_5)*(0.1+S2)-(theta_7)*X1" />
<StateVariable Name="X2"
               DefaultInitialCondition="1000.0"
               Formula="-A1+A2-(theta_8)*X2" />
56 <StateVariable Name="X3"
               DefaultInitialCondition="1000.0"
               Formula="-A3+A4+(theta_6)+(theta_12)*U2t-(theta_9)*X3" />
61 <StateVariable Name="X4"
               DefaultInitialCondition="1000.0"
               Formula="A3-A4-A5" />
<StateVariable Name="X5"
               DefaultInitialCondition="0.0"
               Formula="+ (theta_11)*U1t-A51" />
66 <StateVariable Name="X6"
               DefaultInitialCondition="1000.0"
               Formula="A6-A7+(theta_15)+(theta_18)*U3t-(theta_16)*X6" />
<StateVariable Name="X7"
               DefaultInitialCondition="1000.0"
               Formula="-A6+A7-(theta_17)*X7" />
71 <StateVariable Name="X8"
               DefaultInitialCondition="1000.0"
               Formula="+A8-A9-(theta_24)*X8" />
76 <StateVariable Name="X9"
               DefaultInitialCondition="1000.0"
               Formula="+A9-A11" />
<StateVariable Name="X10"
               DefaultInitialCondition="1000.0"
               Formula="-A8+A10" />
81 <StateVariable Name="X11"
               DefaultInitialCondition="1000.0"
               Formula="-A10+A11+(theta_23)+(theta_26)*U4t-(theta_25)*X11" />
<Function Name="Y1" Formula="(X4+X5)"/>

```

```
86 <Function Name="Y2" Formula="(X3+X5)"/>
<Function Name="Y3" Formula="(X7+X6)"/>
<Function Name="Y4" Formula="(X6)"/>
<Function Name="Y5" Formula="(X10+X9+X11+X8)"/>
<Function Name="Y6" Formula="X8"/>
91 </VectorField>
```