WIENER -NN MODEL AND ITS USE IN MODELING OF BIOTECHNICAL PROCESSES

Arto Visala
Automation Technology Laboratory
Helsinki University of Technology
Arto.Visala@hut.fi

Contents

Linear dynamical system

$$y(t)=\int_0^{\infty} h(\tau)u(t-\tau)d\tau \Leftrightarrow Y(s)=H(s)U(s)$$

in discrete case

$$y(t)=\sum_{i=0}^{\infty} h(i)u(t-i) \Leftrightarrow y(t)=H(q)u(t)$$

When impulse response *h* (or transfer function *H*) is approximated in Laguerre basis, whose orthogonal basis functions are l_k (or basis transfer functions L_k) $k=0,1,2,\ldots$

$$h(i)=\sum_{k=0}^{d-1}\theta_k l_k(i), \qquad \theta_k=\sum_{i=0}^{\infty} h(i)l_k(i), \qquad \sum_{i=0}^{\infty} l_j(i)l_k(i)=\begin{cases}0 & k\neq j\\ 1 & k=j\end{cases}$$

then

$$y(t)=\sum_{i=0}^{\infty} h(i)u(t-i)=\sum_{i=0}^{\infty}\left(\sum_{k=0}^{d-1}\theta_k l_k(i)\right)u(t-i)=\sum_{k=0}^{d-1}\theta_k\sum_{i=0}^{\infty} l_k(i)u(t-i)=$$

$$\sum_{k=0}^{d-1}\theta_k L_k(q)u(t)=\theta^T\left[L_0(q)u(t) \quad \ldots \quad L_{d-1}(q)u(t)\right]^T=\theta^T\left[z_0(t) \quad \ldots \quad z_{d-1}(t)\right]^T$$

It can be seen that

1. Output is a *linear* combination of the outputs of Laguerre filters $z_k(t)$
2. $z_k(t)$ are the coefficients of Laguerre expansion of the past input signal $u(t)$ at time instant *t*.
3. Description of history of input realizes the *state* of the system in a very natural way. Output is a linear mapping from the state.
4. Laguerre representation is efficient when the impulse response has basic form suitable for Laguerre basis and Laguerre parameter has been selected so that Laguerre functions cover the essential history.

Reduced Wiener representation

The starting point is Volterra representation for nonlinear systems (SISO)

$$y(t) = (h_0) + \int_0^\infty h_1(\tau_1)u(t-\tau_1)d\tau_1 + \int_0^\infty \int_0^\infty h_2(\tau_1,\tau_2)u(t-\tau_1)u(t-\tau_2)d\tau_1 d\tau_2 + \ldots$$

$$\int_0^\infty \ldots \int_0^\infty h_n(\tau_1,\tau_2,\ldots,\tau_n)u(t-\tau_1)u(t-\tau_2)\ldots u(t-\tau_n)d\tau_1 d\tau_2 \ldots d\tau_n + \ldots$$

When Volterra kernel functions h_p , i.e. generalized impulse responses, are approximated in p-dimensional Laguerre basis

$$h_p(\tau_1,\ldots,\tau_p) = \sum_{n_1=0}^\infty \ldots \sum_{n_p=0}^\infty d_{n_1\ldots n_p} l_{n_1}(\tau_1)\ldots l_{n_p}(\tau_p)$$

$$d_{n_1\ldots n_p} = \int_0^\infty \ldots \int_0^\infty h_p(\tau_1,\ldots,\tau_p) l_{n_1}(\tau_1)\ldots l_{n_p}(\tau_p) d\tau_1 \ldots d\tau_p$$

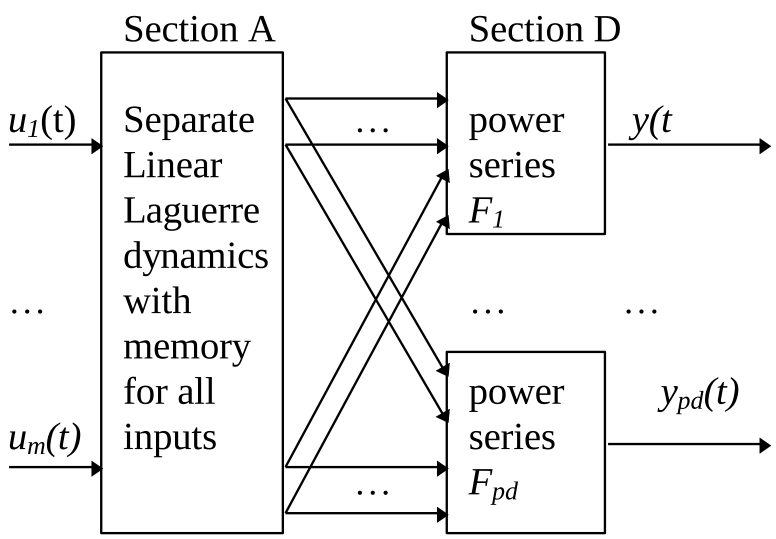
then

$$y(t) = F\left[z_0(t), z_1(t), \cdots z_n(t), \cdots\right] = h_0 + \sum_{p=1}^\infty \left(\sum_{n_1=0}^\infty \ldots \sum_{n_p=0}^\infty d_{n_1\ldots n_p} z_{n_1}(t) z_{n_2}(t) \cdots z_{n_p}(t) \right)$$

It can be seen that

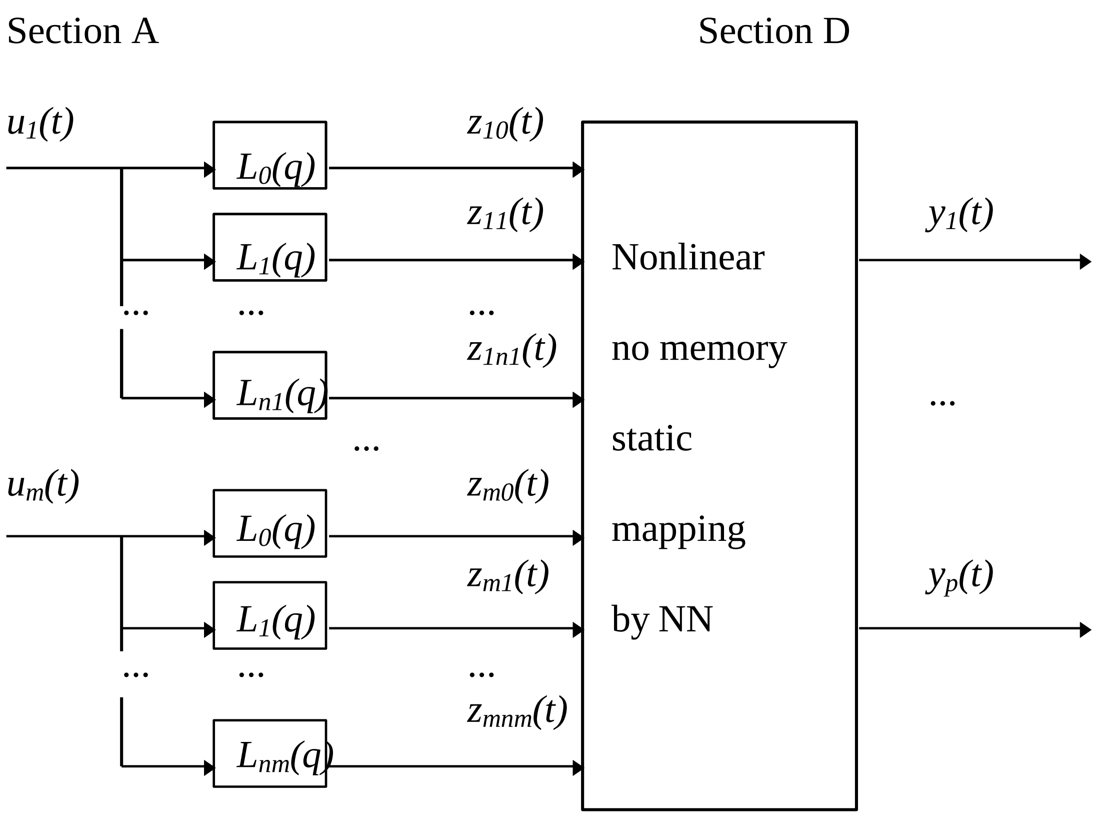
1. Output is now a *non-linear* combination of the outputs of the Laguerre filters $z_k(t)$, power series F.

2. Laguerre representation of the past input realizes the *state* of the non-linear system in a natural way. Output is now a *non-linear* mapping from the state.



’Feedforward’ Wiener-NN model

Barron A. (1993): approximation rate and parsimony of the parameterization of the MLP network are surprisingly advantageous in high-dimensional settings (e.g. compared with finite power series)



’Feedforward’ Wiener NN-model

For modeling of dynamic systems which have finite memory

Model capable for simulation; NN trained as static mapping

State-space model,

- Orthogonal representation of the past inputs as state. Linear Laguerre dynamics as state equation
- NN as static nonlinear measurement equation
- Extended Kalman filter can be used for state estimation

Results of the research on robust identification of linear systems (Mäkilä , Wahlberg *et al*.) can be applied to modeling of non-linear systems

Structural *a priori* information can be included in the model by selecting a suitable basis and parameterizing it roughly with regard to time constants.

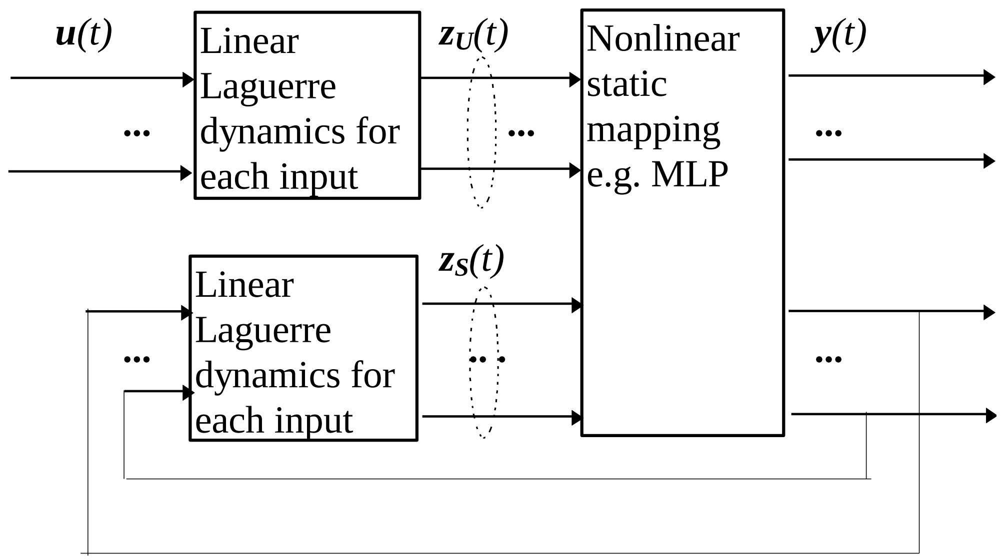
- Laguerre basis suitable for damped and slightly oscillating systems

- Kautz basis for oscillating systems, Meixner etc.

- Active research on orthogonal approximations of linear dynamic systems is going on.

Wiener-NN with feedback model

If the system has infinite memory (for example integrating process) or is (partly) autonomous,
feedback is needed.



State-space model, the state vector consists of Laguerre representations of the past inputs and the past outputs

$$\left[z_S(t)^T \quad z_U(t)^T \right]^T$$

Comparision with 'regression' type NN-models

$$\hat{y}(t|\theta) = f_{NN}(\varphi(t,\theta),\theta)$$

The content of regression vector φ, sliding data windows, form the *state*

- NFIR $\quad$ *u(t - k)* regressor,
- NARX $\quad$ *u(t - k)* and *y(t - k)* regressors, $\quad$ predictor
- NOE $\quad$ *u(t - k)* and $\hat{y}_u(t-k|\theta)$ regressors $\quad$ simulator
-

In Wiener-NN models, orthogonal representations of the past signals, calculated on-line by linear filters, are used instead of sliding data windows

When p_L = $2/T$

Laguerre representation =sliding data window

Use and benefits of Wiener-NN with feedback

NFIR Wiener_NN is the ‘feedforward’ Wiener-NN, above

NARX Wiener-NN is suitable for use as *on-line predictor*,

- Training of as static mapping, for instance LM

NOE Wiener-NN is suitable for use as *simulator*.

- Training must base on model predictions with current parameter estimates; feedback must be taken into account during training.
- Extended Kalman filter used for training

  Tuning of the converge by scaling the system covariance matrix

Linear Laguerre dynamics

Linear discrete Laguerre transfer function, $i=0,1,2,\ldots$

$$\Rightarrow L_i(q) = \frac{K}{(q-a)}\left(\frac{1-aq}{q-a}\right)^i \quad \text{where} \quad a = \frac{\frac{2}{T} - p_L}{p_L + \frac{2}{T}}, \quad K = \frac{2\sqrt{2p_L}}{p_L + \frac{2}{T}} = \sqrt{(1-a^2)T}$$

Laguerre dynamics in state-space form

$$z_{ui}(t) = \begin{bmatrix} z_{ui_0}(t) \\ \vdots \\ z_{ui(ni-1)}(t) \end{bmatrix} = Lag_{ni}(u_i(t-1))(t) = \qquad (11)$$

$$F_{ui} z_{ui}(t-1) + G_{ui} u_i(t-1)$$

where $F_{ui}=$

$$\begin{bmatrix} a & 0 & \cdots & 0 & 0 \\ 1-a^2 & a & \cdots & 0 & 0 \\ (-a)(1-a^2) & 1-a^2 & \ddots & \vdots & \vdots \\ \vdots & \vdots & 1-a^2 & a & 0 \\ (-a)^{ni-2}(1-a^2) & \cdots & (-a)(1-a^2) & 1-a^2 & a \end{bmatrix}$$

$$G_{ui} = \begin{bmatrix} K \\ (-a)^1 K \\ (-a)^2 K \\ \vdots \\ (-a)^{ni-1} K \end{bmatrix} \qquad (12)$$

For all signals of certain type, for instance inputs vector

$$z_U(t) = \begin{bmatrix} Lag_{n_1}(u_1(t-d))(t) \\ \cdots \\ Lag_{n_m}(u_m(t-d))(t) \end{bmatrix} = \begin{bmatrix} z_{u1}(t) \\ \cdots \\ z_{um}(t) \end{bmatrix}$$

$$= \begin{bmatrix} F_{u1}z_{u1}(t-1) + G_{u1}u_1(t-1) \\ \cdots \\ F_{um}z_{um}(t-1) + G_{um}u_m(t-1) \end{bmatrix} \qquad (13)$$

$$\Leftrightarrow z_U(t) = F_U z_U(t-1) + G_U u(t-1)$$

where $F_U = \begin{bmatrix} F_{u1} & & \\ & \ddots & \\ & & F_{um} \end{bmatrix}$, $\quad G_U = \begin{bmatrix} G_{u1} & & \\ & \ddots & \\ & & G_{um} \end{bmatrix}$

Wiener-MLP with feedback in state-space form.

$$\begin{cases} z_U(t) = F_U z_U(t-1) + G_U u(t-1) + w_U \\ z_S(t) = F_S z_S(t-1) + G_S Sf(z_U(t-1), z_S(t-1), \theta) + w_1 \end{cases} \qquad (15)$$

$$y\ (t) = f(z_U(t), z_S(t), \theta) + v_1$$

S selects the feedback and has the following form.

$$S = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & 0 & \cdots & 0 \end{bmatrix} \qquad (16)$$

The state-space model for the NOE-estimation of parameter with Extended Kalman filter

$$\begin{cases} z_S(t) = F_S z_S(t-1) + G_S S f\left(z_U(t-1), z_S(t-1), \theta\right) + w_S \\ \theta(t) = \theta(t-1) + w_\theta \end{cases}$$

$$\Leftrightarrow \left\{ \; x(t) = f_W\left(x(t-1), z_U(t-1)\right) + w \right. \qquad (17)$$

$$y(t) = f\left(z_U(t), z_S(t), \theta(t)\right) + v = f\left(x(t), z_U(t)\right) + v$$

where $x(t) = \left[z_S(t)^T \quad \theta(t)^T \right]^T$

$$Ew_S w_S{}^T = Q_S(t), \quad Ew_\theta w_\theta{}^T = Q_\theta(t), \quad Q = \begin{bmatrix} Q_S & 0 \\ 0 & Q_\theta \end{bmatrix}, \qquad (18)$$

$$Evv^T = R(t)$$

The linearizations around the best current estimates are needed for convariance propagations and updatings and for gain calculation

Extended Kalman filtering can be interpreted as an approximate solution to the minimization of

$$V_N(\theta) = \frac{1}{2}\sum_{t=1}^{N} \left\| y(t) - \hat{y}\left(t \middle| \hat{\theta}(t-1)\right) \right\|_{R^{-1}}^2 + \frac{1}{2}\sum_{t=1}^{N} \left\| \left[w_S(t)^T \quad w_\theta(t)^T \right]^T \right\|_{Q^{-1}}^2$$

$$+ \frac{1}{2} \left\| \left[\left[z_S(0) - z_{S0} \right]^T \quad \left[\theta(0) - \theta_0 \right]^T \right]^T \right\|_{P_0^{-1}}^2$$

- Kalman filter takes into account model predictions with instantaneous parameter estimates, but also corrects model predictions with a robust mechanism towards the real ones towards the real trajectory.

- The model with instantaneous parameter estimates need not be stable.

Procedure for NOE-type parameter estimation

- Rough initial training as a NARX-type model, MLP a static mapping

Training with Extended Kalman filter

  - Several epochs. The estimates of the ordinary state are initialized to right values at the beginning of epochs.

  - The filter is tuned to correct parameter estimates quite vividly at the beginning, and then gradually, to freeze the estimates along with the convergence.

    Elements Q_θ are scaled smaller round by round.

  - The elements of the whole Q can be decreased because
    model predictions become also better as the parameters converge to "right" values.

    Experimentally, it is reasonable to let both Q_θ and Q_S to decrease at the same rate from epoch to epoch.
    For example
    $Q_S(t)$= diagonal(0.1/(10**(epoch-1))),
    $Q_\theta(t)$=diagonal(0.01/(10**(epoch-1)))
    for the measurements R= diagonal(0.01).

  - It is according to the NOE-principles to rely increasingly on the model predictions as the estimation converges.

Case 1: Identification of simulation model for Bakers yeast growth process

$$\dot{x} = \mu(.)\,x - D\,x, \quad D = \frac{F}{V}$$

$$\dot{s} = -\frac{\mu(.)\,x}{Y} + D(S^0 - s), \quad \mu(x,s) = \frac{As}{(B+s)(C+x)}$$

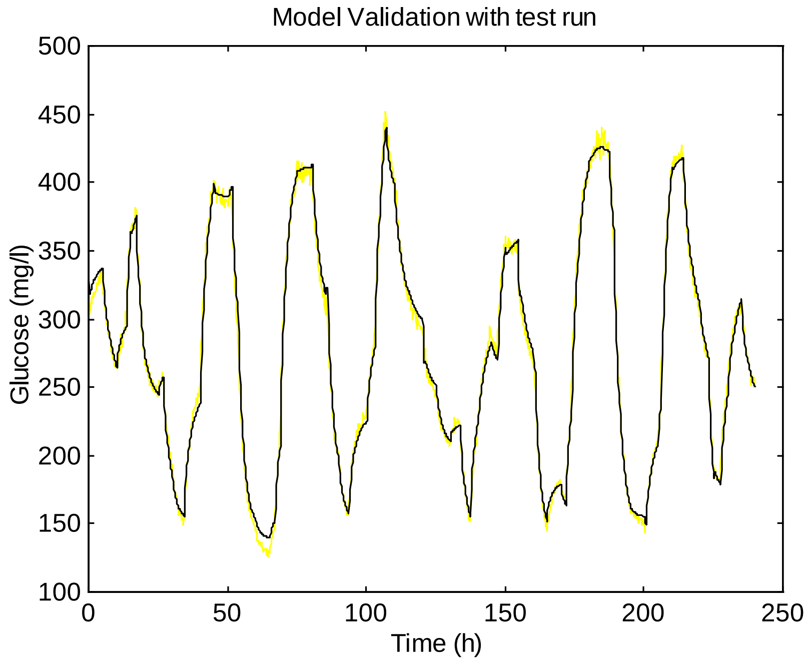
V = 3.7 l, Y= 0.11, (A=28.3, B= 26.5 and C= 6 mg/l).

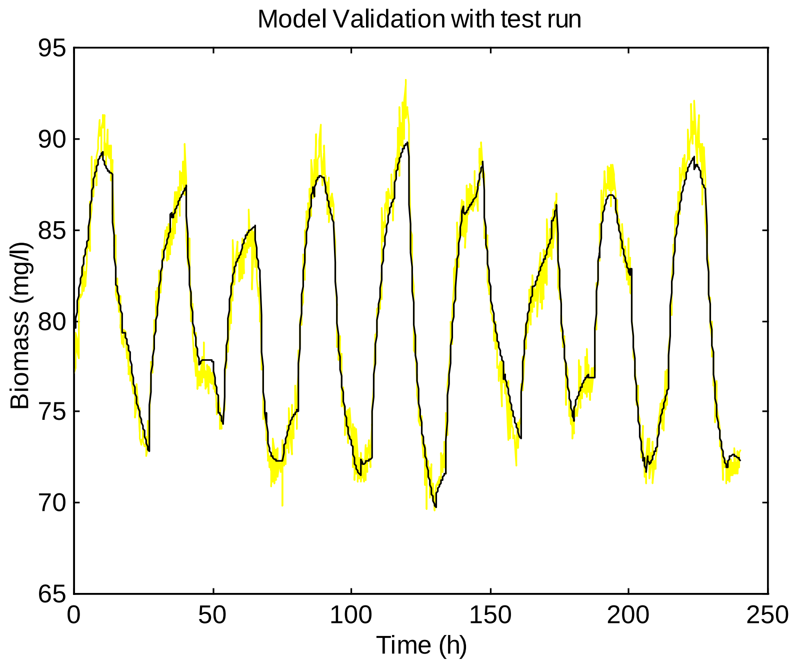
Wiener-MLP with feedback, MLP (6-5-2)

Laguerre representations for biomass x, substrate s (2+2), Laguerre parameter $p_L = 2$.
Laguerre representations for both inputs, D and S^0 (1+1), Laguerre parameters p_L=3.

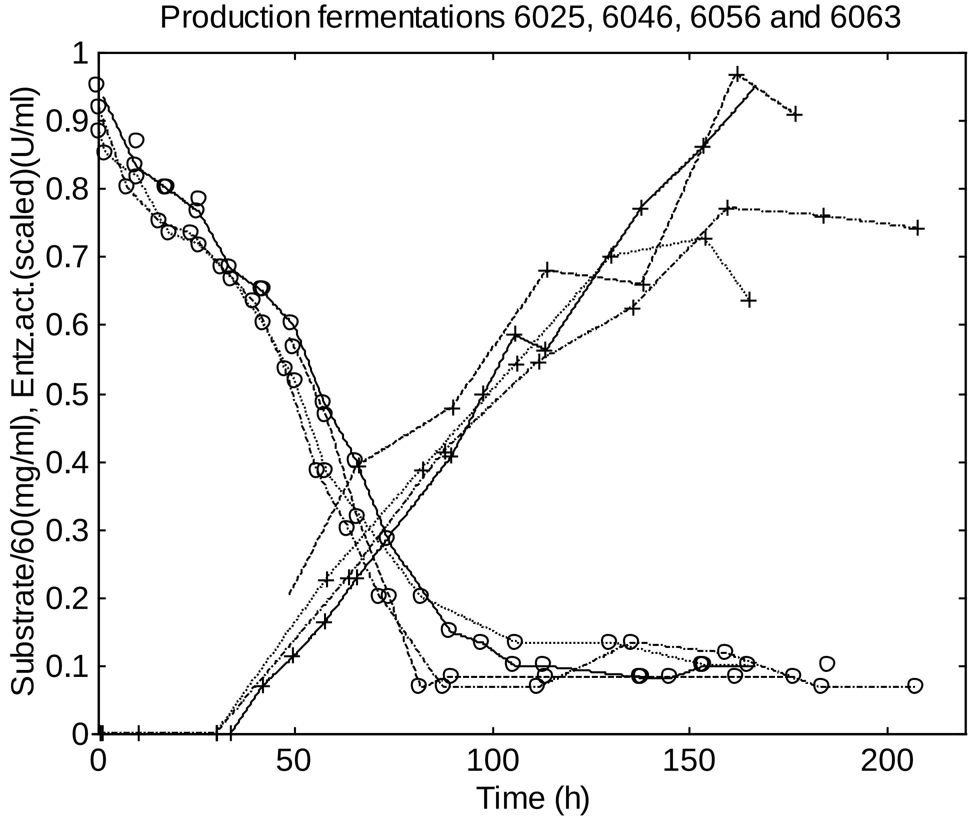
The initial trainig as NARX with LM, 10 epochs.

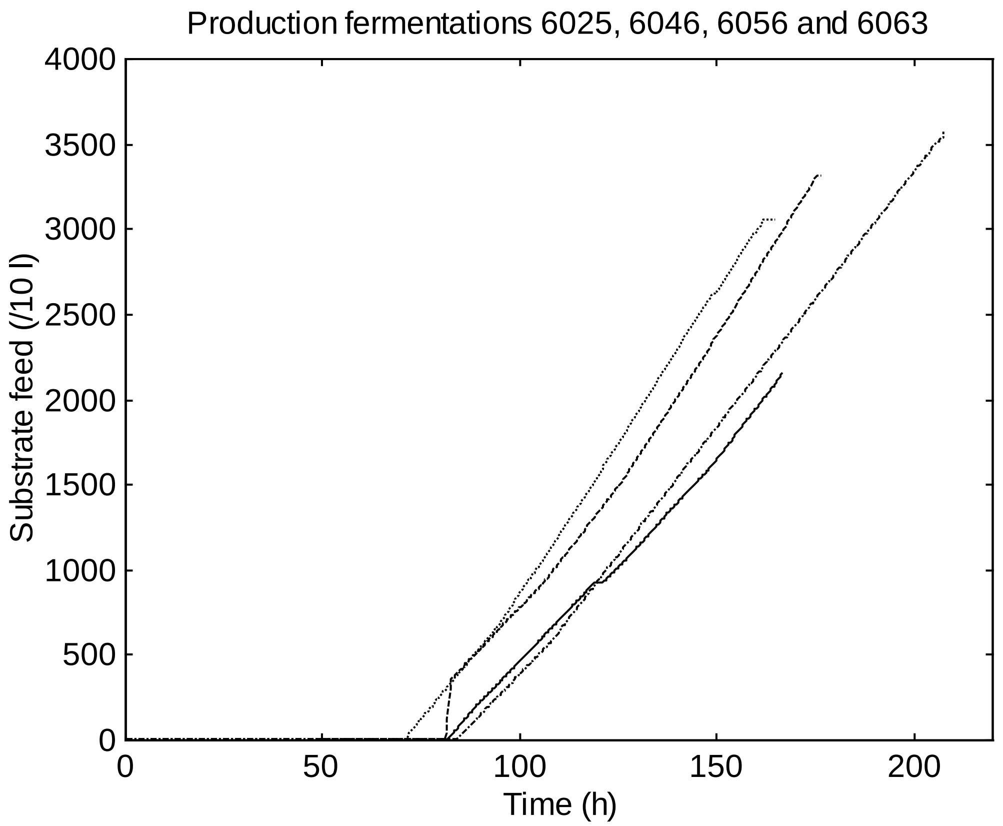
NOE-type estimation with Extended Kalman filter, 5 epochs, covarances above.





Case 2: Simulation model for *Tricoderma* fungi fermentations (150 m^3) producing entzyme





Wiener-MLP with feedback, MLP (6-3-2)

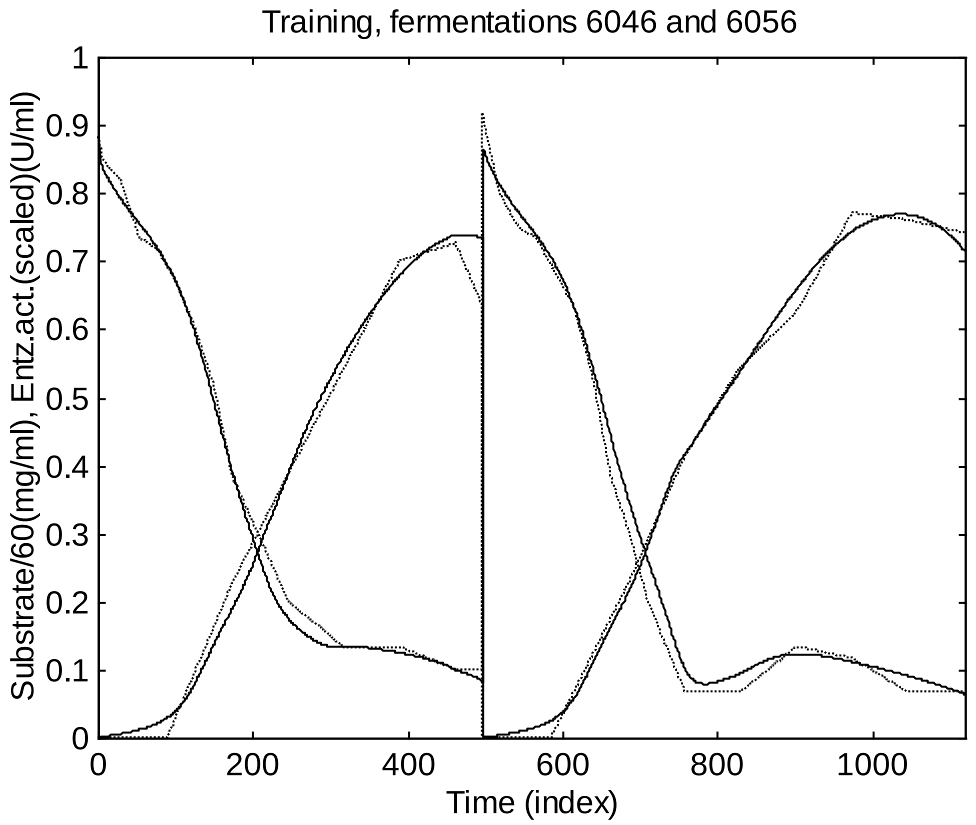
Laguerre representation (2) for input, substrate feed, $p_L = 0.3$.

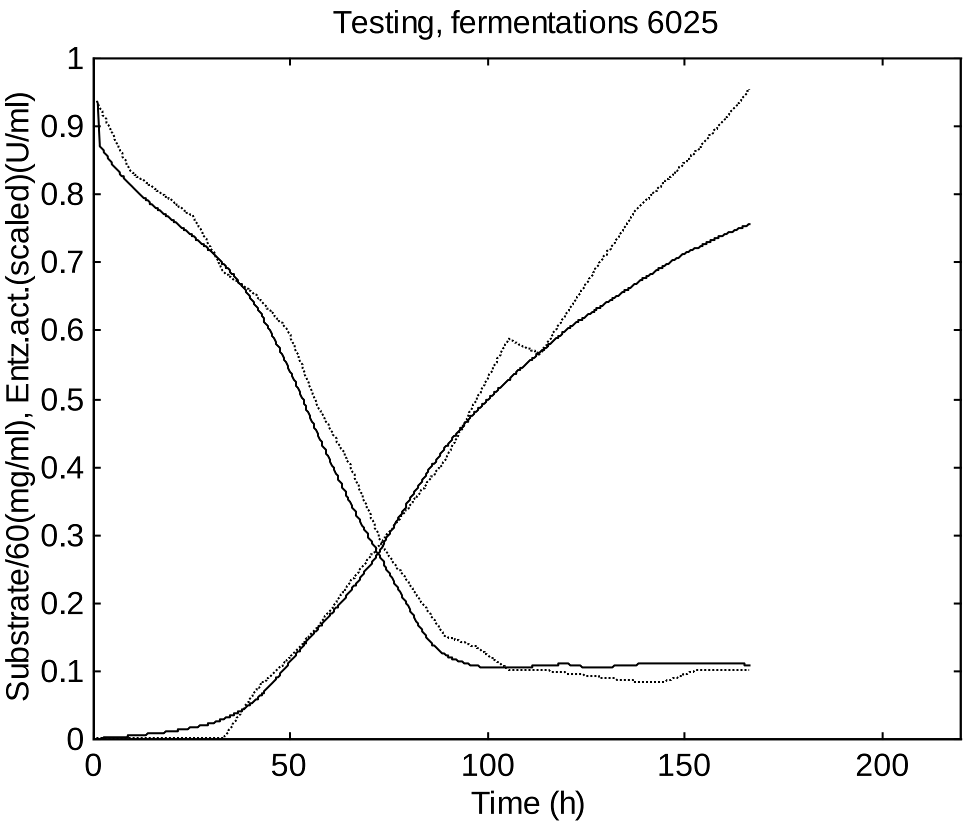
Laguerre representations (2+2) for outputs fed back, *s* and *a*, $p_L = 0.3$.

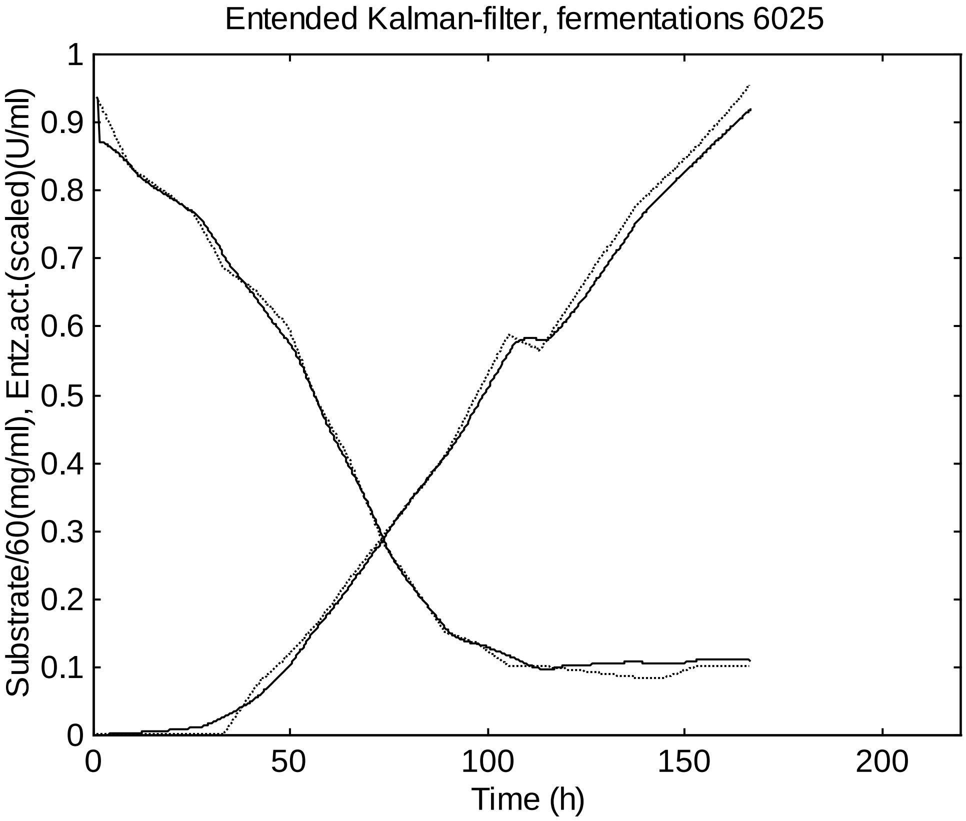
Initial parameters by NARX-type estimation, five epochs.

NOE estimation, five epochs

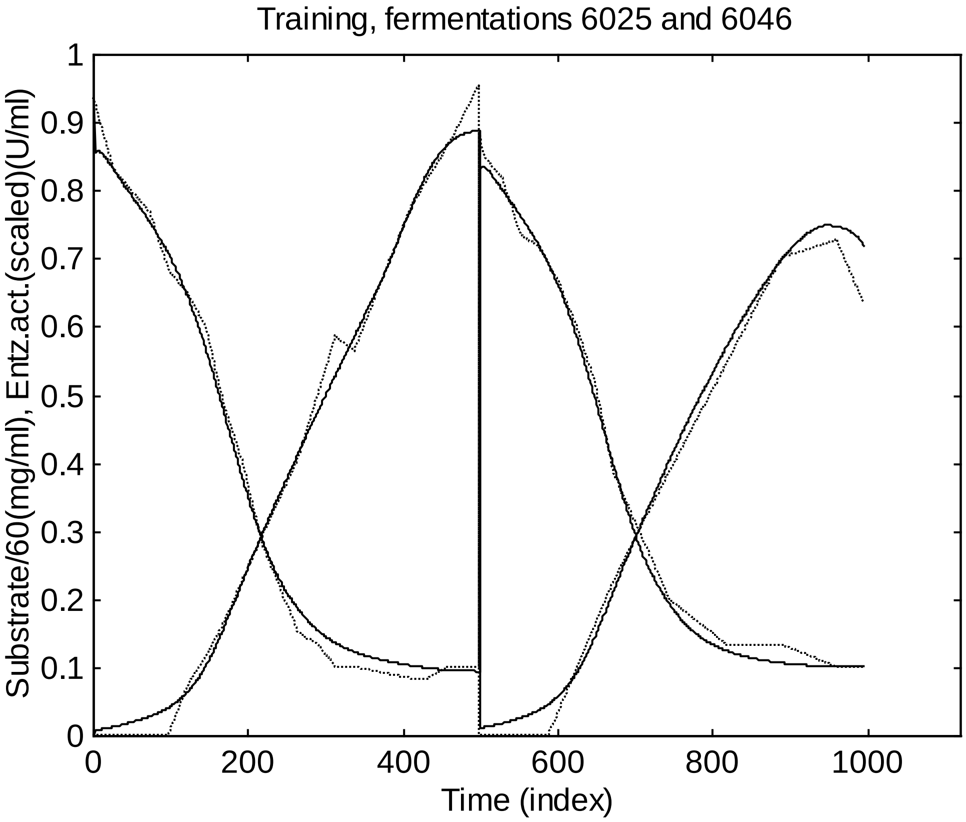
$Q(t)$= diagonal(0.1/(10**(epoch-1))), R = diagonal(0.01).

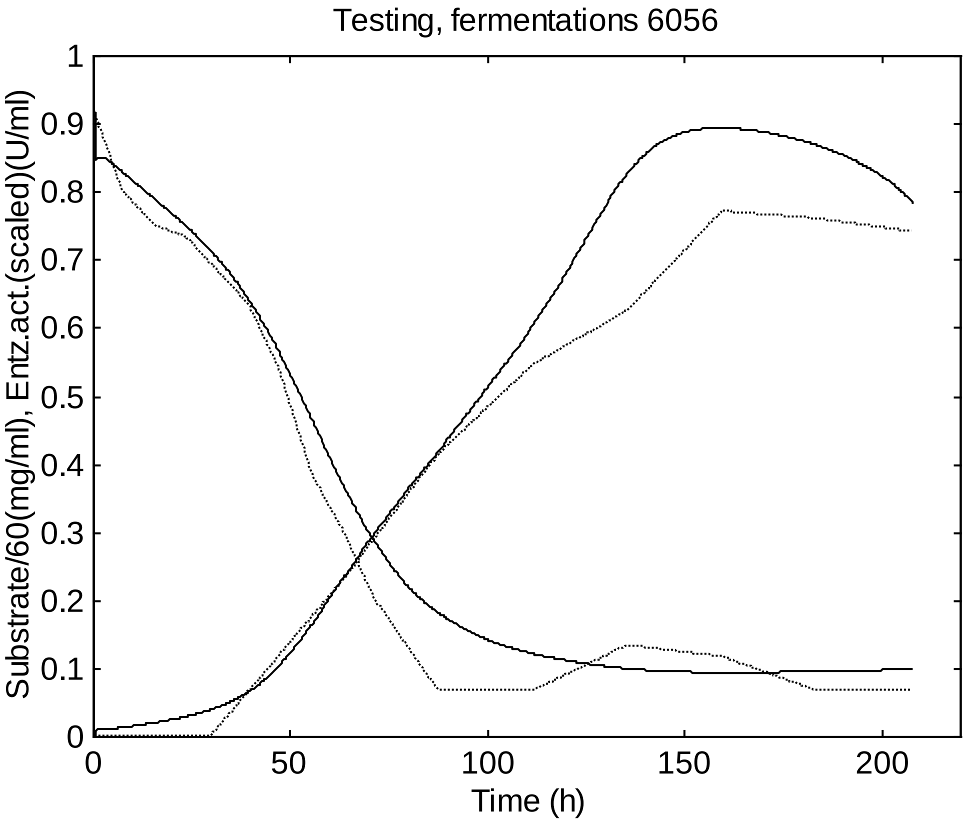



Testing, fermentations 6025
Substrate/60(mg/ml), Entz.act.(scaled)(U/ml)
Time (h)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200


Entended Kalman-filter, fermentations 6025
Substrate/60(mg/ml), Entz.act.(scaled)(U/ml)
Time (h)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200

$Q_S(t)$= diagonal(0.1/(10**(epoch-1))), $Q_\theta(t)$= diagonal(0.05/(10**(epoch-1))),
R = diagonal(0. 1).





Conclusions

The classical Wiener-representation forms mathematically precise basis for dynamic NN-models.

NNs provide a flexible way to realize the Wiener-models.

When feedback is added, also partly or wholly autonomous systems can be modeled.

Structural a priori information can be contained in the model in selecting the basis and its parameter(s). Wiener-MLP models are robust and low-dimensional.

Identification of both the linear dynamic and the static nonlinear mapping in the same time is difficult.

Extended Kalman filter can be well used for parameter estimation of the NOE-type models.

In the NARX case, it does not provide any extra benefit

Extended Kalman filter can be also used as state estimator in the case of different Wiener-MLP models.