Eperimental modeling: learning models from data a user point of view

Mario Milanese
Dip. di Automatica e Informatica
Politecnico di Torino

The Logic of Modeling
Alta Scuola Politecnica
Milano, April 21, 2006
Outline

- Models as tools for making inferences from system data
  - prediction, simulation, control, filtering, fault detection

- Model structures
  - physical law based, input-output description, linear, nonlinear

- Model estimation
  - statistical/parametric, set membership, structured

- Model quality evaluation (vs. model validation)

- Application examples
  - Prediction of atmospheric pollution
  - Simulation of dam crest dynamics
  - Identification of vehicles with controlled suspensions
Regression form of system representation

- System $S^o$ produces output signal $y$ when driven by input signal $u$:

\[
\begin{align*}
\mathbf{y}^{t+1} &= f^o(\mathbf{w}^t) \\
\mathbf{w}^t &= [y^t \cdots y^{t-n_y} \, u_1^t \cdots u_1^{t-n_{u1}} \, u_2^t \cdots u_2^{t-n_{u2}} \cdots]
\end{align*}
\]

- Output $y$ is related to input $u$ by the regression function $f^o$:
Regression form of system representation

- Linear system $\Rightarrow f^o$ is linear in $w^t$:

$$y^{t+1} = a_0 y^t + a_1 y^{t-1} + \cdots + a_{n_y} y^{t-n_y} + b_0 u^t + b_1 u^{t-1} + \cdots + b_{n_u} u^{t-n_u}$$

ARMA system

- If $n_y = 0$: MA (FIR) system

- If $n_u = 0$: AR system

- If $f^o$ nonlinear: NARMA, NFIR, NAR systems
Making inferences from data

- It is desired to make an inference on system $S^o$:
  
  prediction, identification, simulation, 
  control, filtering, fault detection

- The system $S^o$ is unknown, but a finite number of noise corrupted measurements of $y^t, w^t$ are available:

  \[
  \tilde{y}^{t+1} = f^o(\tilde{w}^t) + d^t, \quad t = 1, \cdots, T 
  \]
  
  $d^t$ accounts for errors in data $\tilde{y}^t, \tilde{w}^t$

- The inference is described by the operator $I(f^o, w^T)$

  - one-step prediction $\rightarrow I(f^o, w^T) = f^o(w^T)$
  
  - identification $\rightarrow I(f^o, w^T) = f^o$
Making inferences from data

- Problems:
  - For given estimates \( \hat{f} \approx f^o, \hat{w}^T \approx w^T \)
    evaluate the inference error \( \| I(f^o, w^T) - I(\hat{f}, \hat{w}^T) \| \)
  - Find estimates \( \hat{f} \approx f^o, \hat{w}^T \approx w^T \)
    “minimizing” the inference error

- The inference error cannot be exactly evaluated since \( f^o \) and \( w^T \) are not known

Need of prior assumptions on \( f^o \) and \( d^t \) for deriving finite bounds on inference error
Model structures

The model is described by:

\[
\tilde{y}_{t+1} = f(\tilde{w}_t) + d^t
\]

\[
\tilde{w}_t = [\tilde{y}_t \ldots \tilde{y}_{t-n_y} \tilde{u}_1 \ldots \tilde{u}_{1-t-n_{u1}} \tilde{u}_2 \ldots \tilde{u}_{2-t-n_{u2}} \ldots]
\]

Model structure is defined by:

- type of function \( f \)
- type of noise \( d \)
- which inputs \( u_1, u_2, \ldots \)
- lag values \( n_y, n_{u1}, n_{u2}, \ldots \)
**Statistical/parametric approach**

**Model structures**

- Typical assumptions in literature:
  - on system: \( f^o \in F(\theta) = \left\{ f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma_i(w, \beta_i) \right\} \)
    - known lag values \( n_y, n_{u1}, n_{u2}, \ldots \)
  - on noise: iid stochastic noise

- Functional form of \( F(\theta) \) required:
  - derived from physical laws
  - \( \sigma_i \): “basis” function (polynomial, sigmoid,..)

- Parameters \( \theta \) are estimated by optimizing
  - Least Squares (LS) or Max Likelihood (MS) functionals
Statistical/parametric approach
Model structures

- If possible, physical laws are used to obtain the parametric representation of \( f(w, \theta) \).
- When the physical laws are not well known or too complex, input-output parameterizations are used.

```
“Fixed” basis parametrization
Polinomial, trigonometric, etc.
```

```
“Tunable” basis parametrization
Neural networks, wavelets, etc.
```

often called black-box models
Statistical/parametric approach

Model structures: “fixed” basis

\[ f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma_i(w) \quad \theta = [\alpha_1 \ldots \alpha_r]' \]

\[ \sigma_i(w) : \text{“Basis”} \]

■ **Problem:** Can \( \sigma_i \)'s be found such that

\[ f(w, \theta) \xrightarrow{r \to \infty} f^o(w) \quad ? \]
Statistical/parametric approach
Model structures: “fixed” basis

- For continuous $f^o$, bounded $W \subset \mathbb{R}^n$ and $\sigma_i$ polynomial of degree $i$ (Weierstrass):

$$\limsup_{r \to \infty} \sup_{w \in W} |f^o(w) - f(w, \theta)| = 0$$

Polynomial NARX models
Statistical/parametric approach

Model structures: “tunable” basis

\[ f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma(w, \beta_i) \]

\[ \theta = \left[ \alpha_1 \cdots \alpha_r \beta_{11} \cdots \beta_{rq} \right]^t, \quad \beta_i \in \mathbb{R}^q \]

- One of the most common “tunable” parameterization
  is the one-hidden layer sigmoidal neural network

\[ \sigma(w, \beta_i) = \sigma(w^T a_i + b_i) \]
Statistical/parametric approach 
Model estimation

\[ f^o = f(w, \theta^o) = \sum_{i=1}^{r} \alpha_i^o \sigma \left( w, \beta_i^o \right) \]

\[ \theta^o = [\alpha_1^o \alpha_2^o \cdots \alpha_r^o \beta_1^o \beta_2^o \cdots \beta_r^o] \rightarrow \text{to be estimated} \]

Given \( T \) noise-corrupted measurements of \( y^t, w^t \):

\[ \tilde{y}^2 = f(\tilde{w}^1, \theta^o) + d^1 \]
\[ \tilde{y}^3 = f(\tilde{w}^2, \theta^o) + d^2 \]
\[ \vdots \]
\[ \tilde{y}^{T+1} = f(\tilde{w}^T, \theta^o) + d^T \]

\[ \tilde{Y} = F(\theta^o) + D \]

- Measured output
- Known function
- Unknown residual
Statistical/parametric approach
Model estimation

\[ \tilde{Y} = F(\theta^o) + D \]

Gaussian pdf

Maximum Likelihood – Least Squares estimate

\[ \hat{\theta} = \arg \min_{\theta} R(\theta) \]

\[ R(\theta) = \frac{1}{T} D'D = \frac{1}{T} [Y - F(\theta)]' [Y - F(\theta)] \]

**Problem:** \( R(\theta) \) is in general non-convex
Statistical/parametric approach
Model estimation

“Fixed” basis: \( f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma_i(w) \quad \theta = [\alpha_1 \ldots \alpha_r]' \)

Estimation of \( \theta \) is a linear problem:

\[
\tilde{Y} = L\theta^o + D
\]

\[
L = \begin{bmatrix}
\sigma_1(\tilde{w}_1) & \ldots & \sigma_r(\tilde{w}_1) \\
\vdots & \ddots & \vdots \\
\sigma_1(\tilde{w}_T) & \ldots & \sigma_r(\tilde{w}_T)
\end{bmatrix}
\]

\[
Y = \begin{bmatrix}
\tilde{y}^2 & \tilde{y}^3 \ldots & \tilde{y}^{T+1}
\end{bmatrix}'
\]

- If \( D \) is iid gaussian:

\[
\hat{\theta}^{ML} = (L' L)^{-1} L' Y
\]
**Statistical/parametric approach**

**Estimation accuracy**

- For fixed basis and $D$ iid gaussian:
  \[
  \left| \mathcal{G}_i^o - \hat{\theta}_i^{ML} \right| \leq 2 \left[ (L'L)^{-1} \right]_{ii} \sigma_i \quad \text{w.p.} \ 0.95
  \]

- For tunable basis this result holds asymptotically ($T \to \infty$) with:
  \[
  L = \left( \frac{\partial F}{\partial \mathcal{G}} \right)_{\mathcal{G} = \mathcal{G}^o}
  \]

standard deviation of noise component $d^i$
Statistical/parametric approach

Model structures: properties

- Model structure choice:
  - "basis” type $\sigma_j$
  - Number $r$ of "basis"
  - Number $n$ of regressors

- Problem: “curse of dimensionality”
  The number $r$ of basis needed to obtain “accurate” approximation of $f^o$ grows with the dimension $n$ of regressor space

  in the case of “fixed” basis: exponential growth
Using tunable basis:

- Under suitable regularity conditions on the function to approximate, the number of parameters $r$ required to obtain “accurate” models grows linearly with $n$.

- Estimation of $\theta$ requires to solve a non-convex minimization problem.

Trapping in local minima.
Basic to the statistical/parametric approach is the assumption of no modeling error:

\[ \exists \mathcal{G}^o : f^o = f(w, \mathcal{G}^o) \]

\[ d^t = \tilde{y}^t - f(w, \mathcal{G}^o) \]

is a stochastic variable independent of input u.
Statistical/parametric approach
Modeling errors

- Searches for the functional form of unknown $f^o$ are time consuming and lead to approximate model structures.

  $d^t$ is no more a stochastic variable independent of $u$.

- Statistical estimation in presence of modeling errors is a hard problem.

  **Set Membership approach:**
  - no assumption on the functional form of $f^o$
  - no statistical assumption on $d^t$
**Set Membership approach**

- **SM assumptions:**
  - on system: \( f^o \in F(\gamma) = \{ f \in C^1 : \|f'(w)\|_2 \leq \gamma, \forall w \in W \} \)
  - on noise: \( |d^t| \leq \varepsilon^t + \gamma \delta^t, t = 1, \ldots, T \)

- Significant improvements obtained by:
  - use of “local” bound \( \|f'(w)\|_2 \leq \gamma(w) \)
  - scaling of regressors \( w \) to adapt to data
Set Membership approach

- All information (prior and data) are summarized in the Feasible Systems Set:

\[
FSS^T = \left\{ f \in F(\gamma) : |\tilde{y} - f(\tilde{w})| \leq \varepsilon + \gamma \delta, \quad t = 1, \ldots, T \right\}
\]

- \(FSS^T\) is the set of all systems \(\in F(\gamma)\) that could have generated the data

- Inference algorithm \(\Phi\) maps all information into estimated inference:

\[
\hat{I} = \Phi(FSS^T) \approx I(f^o, w^T)
\]
Set Membership approach
Prior assumptions validation

- Prior assumptions are invalidated by data if $FSS^T$ is empty.

- Prior assumptions are considered validated if $FSS^T \neq \emptyset$.

- The fact that the priors are validated by using the present data does not exclude that they may be invalidated by future data.

Set Membership approach
Prior assumptions validation

Define:
\[
\bar{f}(w) = \min_{t=1,\ldots,T-1} (h_t + \gamma \| w - \tilde{w}_t \|_2)
\]
\[
\underline{f}(w) = \max_{t=1,\ldots,T-1} (h_t + \gamma \| w - \tilde{w}_t \|_2)
\]
\[
h_t = \tilde{y}^{t+1} + \varepsilon^t + \gamma \delta^t, \quad \underline{h}^t = \tilde{y}^{t+1} - \varepsilon^t - \gamma \delta^t
\]

Theorem:
Conditions for assumptions to be validated are:

- necessary: \( \bar{f}(\tilde{w}^t) \geq \underline{h}^t, \quad t = 1,\ldots,T \)
- sufficient: \( \bar{f}(\tilde{w}^t) > \underline{h}^t, \quad t = 1,\ldots,T \)
Set Membership approach
Prior assumptions validation

- In space \((\gamma, \epsilon)\) the surface \(\gamma^*(\epsilon) = \inf_{FSS^T \neq \emptyset} \gamma\)
separates falsified values from validated ones

Used for the choice of \(\gamma, \epsilon\) values
Set Membership approach

Error and optimality concepts

- (Local) Inference error:

\[
E(\hat{I}) = E[\Phi(FSS^T)] = \sup_{\tilde{f} \in FSS^T} \sup_{|w^T - \tilde{w}^T| \leq \varepsilon^T + \gamma \delta^T} \| \Phi(FSS^T) - I(\hat{f}, w^T) \| 
\]

- An algorithm \( \Phi^* \) is optimal if:

\[
E[\Phi^* (FSS^T)] = \inf_{\Phi} E[\Phi(FSS^T)] = r \quad \forall FSS^T
\]

\( r \): (local) radius of information

- An algorithm \( \Phi^\alpha \) is \( \alpha \)-optimal if:

\[
E[\Phi^\alpha (FSS^T)] \leq \alpha \inf_{\Phi} E[\Phi(FSS^T)] \quad \forall FSS^T
\]
Set Membership approach

Inference \implies \textbf{Identification}: \( I(f,w^T) = f \)

- Let \( || I(f,w^T)|| = || f ||_p = \left[ \int |f(w)|^p dw \right]^{1/p} \)

- Define \( f^c(w) = \frac{1}{2} [f(w) + \bar{f}(w)] \)

**Theorem:**

1) The identification algorithm \( \Phi_c(FSS^T) = f^c \)

   is optimal for any \( L_p \) norm, \( 1 \leq p \leq \infty \)

2) The radius of information \( r \) is:

\[
E[f^c] = r = \frac{1}{2} \| \bar{f} - f \|_p
\]
Set Membership approach

Inference $\Rightarrow$ Prediction: $I(f, w^T) = f(w^T)$

- Let:

* $||I(f, w^T)|| = |f(w^T)|$

* $B_\delta(\tilde{w}^t) = \{w \in W : \|w - \tilde{w}^t\|_2 \leq \delta^t\}$
Set Membership approach

Inference $\Rightarrow$ Prediction: $I(f, w^T) = f(w^T)$

Theorem:

i) The prediction algorithm $\Phi^c(FSS^T) = f^c(\tilde{w}^T)$ is 2-optimal, with prediction error bounded by:

$$E\left[\Phi^c(FSS^T)\right] \leq \frac{1}{2}\left[f(\tilde{w}^T) - \underline{f}(\tilde{w}^T)\right] + \gamma\delta^T$$

ii) If $B_\delta(\tilde{w}^T) \subset C^T \cap \overline{C}^T$, then prediction $\hat{y}^{T+1} = f^c(\tilde{w}^T)$ is optimal and the radius of information is:

$$E\left[\Phi^c\right] = r = \frac{1}{2}\left[f(\tilde{w}^T) - \underline{f}(\tilde{w}^T)\right] + \gamma\delta^T$$
Structured identification

- In the case of large dimension of regressor space it is often very hard to obtain satisfactory modeling accuracy.

Structured (block-oriented) identification

- The high-dimensional problem is reduced to the identification of lower dimensional subsystems and to the estimation of their interactions
Structured identification

- Typical cases: Wiener, Hammerstein and Lur’e systems
Structured identification

Iterative identification algorithm:

- Initialisation: get an initial guess $M_2^{(0)}$ of $M_2$
- Step k:
  1) Compute $v^{(k)}$ such that $M_2^{(k-1)}[v^{(k)}]=y$
  2) Identify $M_1^{(k)}$ using $u$ and $y$ as inputs, $v^{(k)}$ as output
  3) Identify $M_2^{(k)}$ using $v^{(k)} = M_2^{(k)}[u,y]$ as input, $y$ as output and return to step 1)

Key feature:
The identification error is non-increasing for increasing iteration.
Model quality evaluation

- The usual approach is to look for model validity
- Model invalidity only can be surely asserted, when the model does not explain the measured data
  \[ |\tilde{y}^t - y_M^t| > \text{expected noise size} \]
- Infinitely many not-invalidated models can be derived
- Even more, infinitely many models exactly explaining the data can be derived
  \[ \text{"overfitting" danger} \]
Model quality evaluation

- Finding models exactly explaining the data

choose \( r \) of basis functions = \( T \) of measured data

\[
L = \begin{bmatrix}
\sigma_1(\tilde{w}_1) & \cdots & \sigma_T(\tilde{w}_1) \\
\vdots & \ddots & \vdots \\
\sigma_1(\tilde{w}_T) & \cdots & \sigma_T(\tilde{w}_T)
\end{bmatrix}
\]

invertible

\[
\hat{g} = (LL)^{-1}LY \\
Y_M = L\hat{g} = L(LL)^{-1}LY = \tilde{Y}
\]
Model quality evaluation

Example:

\[ \tilde{u}^1 = -2 \quad \tilde{u}^2 = 0.5 \quad \tilde{u}^3 = 0.8 \quad \tilde{u}^4 = -0.5 \quad \leftarrow \text{input} \]

\[ \tilde{y}^1 = 0 \quad \tilde{y}^2 = 1 \quad \tilde{y}^3 = -8 \quad \tilde{y}^4 = 0.125 \quad \leftarrow \text{output} \]

\[ M_1(\mathcal{G}) \Rightarrow \quad y_{M1}^{t+1} = \mathcal{G}_1 u^t + \mathcal{G}_2 u^{t-1} \]

\[ M_2(\mathcal{G}) \Rightarrow \quad y_{M2}^{t+1} = \mathcal{G}_1 u^t + \mathcal{G}_2 (u^{t-1})^2 \quad \text{candidate model structures} \]

\[ M_3(\mathcal{G}) \Rightarrow \quad y_{M3}^{t+1} = \mathcal{G}_1 u^t + \mathcal{G}_2 (u^{t-1})^3 \]
Model quality evaluation

Estimation of $M_1, M_2, M_3$

$M_1(\mathcal{I}) \Rightarrow$

$t=2 \rightarrow \begin{bmatrix} -8 \\ 0.125 \end{bmatrix} = \begin{bmatrix} 0.5 & -2 \\ 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = L^1Y = \begin{bmatrix} -2.03 \\ 3.49 \end{bmatrix}$

$t=3 \rightarrow \begin{bmatrix} -8 \\ 0.125 \end{bmatrix} = \begin{bmatrix} 0.5 & -4 \\ 0.8 & 0.25 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = L^1Y = \begin{bmatrix} 0.81 \\ -2.10 \end{bmatrix}$

$M_2(\mathcal{I}) \Rightarrow$

$t=2 \rightarrow \begin{bmatrix} -8 \\ 0.125 \end{bmatrix} = \begin{bmatrix} 0.5 & -8 \\ 0.8 & 0.125 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = L^1Y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$M_3(\mathcal{I}) \Rightarrow$

$t=2 \rightarrow \begin{bmatrix} -8 \\ 0.125 \end{bmatrix} = \begin{bmatrix} 0.5 & -2 \\ 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = L^1Y = \begin{bmatrix} -2.03 \\ 3.49 \end{bmatrix}$

$t=3 \rightarrow \begin{bmatrix} -8 \\ 0.125 \end{bmatrix} = \begin{bmatrix} 0.5 & -4 \\ 0.8 & 0.25 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = L^1Y = \begin{bmatrix} 0.81 \\ -2.10 \end{bmatrix}$
Model quality evaluation

- All models $M_1, M_2, M_3$ explain exactly the given data $y$ measured by accuracy in simulating data not used for model estimation.

How to choose among them?

choose the one with the best “predictive ability”
Model quality evaluation

- Several indexes have been proposed for estimating the predictive ability of models:
  - $FPE = R(\hat{\vartheta}) \frac{T + n}{T - n}$
  - $AIC = \ln R(\hat{\vartheta}) + \frac{2n}{T}$
  - $BIC = \ln R(\hat{\vartheta}) + \frac{n \ln T}{T}$

- They provide quite crude approximations, especially for nonlinear systems

- A simple but effective approach: splitting of data
  - estimation data: estimate candidate models $M_i, i = 1, \ldots, m$
  - calibration data: choose the best one among $M_i$
Model quality evaluation

- Best model among candidate ones $M_i$

minimum simulation error on the “calibration” data

- Example: $M_3$ is the best one among $M_1$, $M_2$, $M_3$
Applications

- Prediction of atmospheric pollution
- Simulation of dam crest dynamics
- Identification of vehicles with controlled suspensions
Prediction of urban ozone peaks

- N2O
- SOx
- CH4
- NH3
- NOx
- CO
- CO2
- VOC

O2 → NO → NO2 → RO2 → O3

smog fotochimico

RH → O → O2

[Image of urban infrastructure and排放源]
Prediction of urban ozone peaks

• Combustion processes and high solar radiation cause high tropospheric ozone concentrations

• Prediction of ozone concentrations is important for authorities in charge of pollution control and prevention

• Studies in the literature show that physical models are not able to reliably forecast the links between precursor emissions (Nox, VOC), meteorological conditions and ozone concentrations


  ➢ Jenkin-Clemitshaw “Ozone and other photochemical pollutants: chemical processes governing their formation”, Atmos. Environ., 1999
Prediction of urban ozone peaks

typical data at Broletto (Bs)

Value to be predicted

Used measurements

\[ [O_3] \text{[µg/m}^3\text{]} \]

day t-1  day t  day t+1
Prediction of urban ozone peaks

• Structure of used models:

\[ y_{t+1} = f^o(w_t) \]

\[ w_t = [y_t, u_{t1}, u_{t2}, u_{t3}, u_{t4}] \]

- \( y_{t_t} \): max O₃ concentration at day t
- \( u_{t1} \): mean NO₂ concentration at 4-8 pm of day t
- \( u_{t2} \): mean O₃ concentration at 4-8 pm of day t
- \( u_{t3} \): max temperature at day t
- \( u_{t4} \): forecast of max temperature at day t+1
Prediction of urban ozone peaks

• Prediction methods tested:
  - **PERS**: $y^{t+1} = y^t$
  - **ARCX**: periodic ARX
  - **NN**: sigmoidal neural net
  - **NF**: neuro-fuzzy
  - **NSM**: nonlinear set membership

• Hourly data measured at Brescia center:
  - **1995-1998**: estimation data set
  - **1999**: calibration data set
  - **2000-2001**: testing data set
Prediction of urban ozone peaks

Indexes measuring the ability to predict concentrations exceeding a given threshold:

<table>
<thead>
<tr>
<th>predicted</th>
<th>observed</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>a</td>
<td>f</td>
</tr>
<tr>
<td>no</td>
<td>m - a</td>
<td>N + a - m - f</td>
</tr>
</tbody>
</table>

- fraction of Correct Predictions: \( \text{CP} = \frac{a}{m} \)%
- fraction of False Alarms: \( \text{FA} = \frac{1 - a}{f} \)%
- Success index: \( \text{SI} = \left[ \frac{(a/m) + ((N+a-m-f)/(N-m))}{1} \right] \)%

*European Environmental Agency, Tech. Report 9, 1998*
# Prediction of urban ozone peaks

## Calibration data set: m=63 exceeded thresholds

<table>
<thead>
<tr>
<th></th>
<th>PERS</th>
<th>ARCX</th>
<th>NN</th>
<th>NF</th>
<th>NSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>65.1</td>
<td>61.9</td>
<td>69.8</td>
<td>63.5</td>
<td>71</td>
</tr>
<tr>
<td>FA</td>
<td>33.9</td>
<td>25</td>
<td>27.9</td>
<td>25.9</td>
<td>27.4</td>
</tr>
<tr>
<td>SI</td>
<td>47.6</td>
<td>51.1</td>
<td>55.7</td>
<td>51.8</td>
<td>51.2</td>
</tr>
</tbody>
</table>

## Testing data set: m=39 exceeded thresholds

<table>
<thead>
<tr>
<th></th>
<th>PERS</th>
<th>ARCX</th>
<th>NN</th>
<th>NF</th>
<th>NSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>41.5</td>
<td>35.9</td>
<td>53.8</td>
<td>66.7</td>
<td>71.8</td>
</tr>
<tr>
<td>FA</td>
<td>57.5</td>
<td>51.7</td>
<td>40</td>
<td>44.7</td>
<td>44</td>
</tr>
<tr>
<td>SI</td>
<td>34.4</td>
<td>31.3</td>
<td>49.6</td>
<td>60.2</td>
<td>63.5</td>
</tr>
</tbody>
</table>
Model of Schlegeis Arch Dam

- Model to simulate the crest displacement of the dam as function of:
  - water level
  - concrete temperature
  - air temperature


- Difficulties in deriving reliable physical models

- Models tested: ARX, NN, NSM
Model of Schlegeis Arch Dam

• Structure of used models:

\[ y^{t+1} = f^o(w^t) \]

\[ w^t = [y^t y^{t-1} u_1^{t+1} u_1^t u_1^{t-1} u_2^{t+1} u_2^t u_3^{t+1} u_3^t] \]

- \( y^t \): crest displacement at day \( t \)
- \( u_1^t \): water level at day \( t \)
- \( u_2^t \): concrete temperature at day \( t \)
- \( u_3^t \): mean air temperature at day \( t \)

• Daily data:

- **1992-1996**: estimation data set
- **1997-1998**: calibration data set
- **1999-2000**: testing data set
Model of Schlegeis Arch Dam

- Simulation results on the testing data set:

  ![Graphs showing comparison between ARX model, NN model, and experimental data](image)
Model of Schlegeis Arch Dam

- Simulation results on the testing data set:

![Graphs showing comparison between NN model, experimental data, and NSM model.]
Identification of vehicles with controlled suspensions

GOAL: Derive a model for simulation of chassis and wheels accelerations as function of road profile and damper control

USE: Virtual design and tuning of Continuous Damping Control systems
Experimental setting

- C-segment prototype vehicle with controlled dampers and CDC-Skyhook (Continuous Damping Control system).

- Measurements are performed on a four-poster test bench of FIAT-Elasis Research Center.
Experimental setting

Road profiles:

- **Random:** random road.
- **English Track:** road with irregularly spaced holes and bumps.
- **Short Back:** impulse road.
- **Motorway:** level road.
- **Pavé track:** road with small amplitude irregularities.
- **Drain well:** negative impulse road.

Note: The road profiles are symmetric (left=right).
Experimental setting

Data set: 93184 data, collected with a sampling frequency of 512 Hz, partitioned as follows:

• Estimation data set: 0-5 seconds of each acquisition.
• Calibration data set: 5-7 seconds of each acquisition.
• Testing set: 7-14 seconds of each acquisition.
Structure of vehicles vertical dynamics

Since the road profiles are symmetric, a Half-car model has been considered:
Structured Identification of vehicles vertical dynamics

Structure decomposition:

- **CE**: chassis + engine
- **SWT**: suspension + wheel + tire

Measured variables:
- $p_{rf}$ and $p_{rr}$: front and rear road profiles.
- $i_{sf}$ and $i_{sr}$: control currents of front and rear suspensions.
- $a_{cf}$ and $a_{cr}$: front and rear chassis vertical accelerations.

Note: $F_{cf}$ and $F_{cr}$ are not measured.
Results on testing set of NSM model

Front wheel acceleration: english track road measurements, NSM model
Results on testing set of NSM model

Chassis front accelerations: random road measurements, NSM model
Results on testing set of NSM model

Chassis rear accelerations: random road measurements, NSM model.
Comparison with physical model

Chassis front accelerations: random road measurements, NSM model, physical model