Exercise 2: hair dryer model identification (real data)

Consider an hair dryer whose input is the electric power and output is the air temperature

Problem:

1) Identify from experimental data several ARX, ARMAX and OE models of different orders and delays for the hair dryer

2) Select the "best" model among the ones identified at step 1. Use the following criteria/methods for the selection:
- cross validation
- residual analysis
- Best Fit index (to be maximized), defined as:

$$\textbf{Best Fit} = 1 - \sqrt{\frac{MSE}{\frac{1}{N}\sum_{t=1}^{N}(y(t)-\bar{y})^2}}, \quad \text{with } MSE = \frac{1}{N}\sum_{t=1}^{N}(y(t)-\hat{y}(t))^2 = \text{mean square error}, \ \bar{y} = \frac{1}{N}\sum_{t=1}^{N}y(t) = \text{sample mean of } y$$

Main steps:

In order to practice the System Identification Toolbox under MATLAB version R2014A:
- type the command `iddemo`
- enter `1` for linear model and general toolbox features
- select the example number `1` for a guided tour on the Graphical User Interface *GUI* (required)
- select the example number `2` for building simple models from real laboratory process data (suggested)
- select the example number `3` for comparing different identification methods (suggested)

To solve the hair dryer identification problem:

1) Open the System Identification Toolbox *GUI* typing the command `ident`

2) Import the dryer data (click on *Import data* and select *Example*, then click on *Import* and *Close* in the *Import Data* window): the `Dryer` window will appear in the *Data Views* area and also as *Working Data* and *Validation Data* boxes; the corresponding data may be shown by clicking on the *Time plot* option: the input is the electric power (a Pseudo Random Binary Sequence), the output is the air temperature

3) Remove the mean value form the data (click on *Preprocess* operation and select *Remove means*): the `Dryerd` window will appear in the *Data Views* area; the corresponding data may be shown by clicking on the *Time plot* option

4) Make the data without mean as current working dataset, by clicking the `Dryerd` window in the *Data Views* area and dragging it to the *Working Data* box

5) Partition the whole dataset in two subsets, called estimation dataset (ES) and validation dataset (VS) (click on *Preprocess* operation and choose *Select range*; type `1 500` as Samples for the `Dryerde` dataset and click on *Insert*; type `501 1000` as Samples for the `Dryerdv` dataset and click on *Insert* and *Close*): the `Dryerde` and `Dryerdv` windows will appear in the *Data Views* area and the corresponding data may be shown by clicking on the *Time plot* option

6) Set ES as the current working dataset and VS as the current validation dataset, by clicking the `Dryerde` and `Dryerdv` windows in the *Data Views* area and dragging them to the *Working Data* and *Validation Data* boxes, respectively

7) Perform the order selection for an ARX structure (click on *Estimate* and select *Polynomial Models*, then click on *Order Selection* and *Estimate*) in order to consider and compare different model selection criteria: Best Fit, AIC and MDL, with:

$$\textbf{AIC} = n\frac{2}{N} + \ln(MSE) \qquad \text{(index to be minimized)}$$
$$\textbf{MDL} = n\frac{\ln(N)}{N} + \ln(MSE) \qquad \text{(index to be minimized)}$$

where, in this case and for ARX models (in prediction mode), n = model complexity = $\dim(\theta)$ = `na+nb`

8) Identify several models of different orders and delays using the following structures:
- `ARX(na,nb,nk)`, using `na`=`nb`=`1`,...,`4` and `nk`=`1`,...,`3` (and also the optimal values of `na`, `nb`, `nk` obtained from step 7)
- `ARMAX(na,nb,nc,nk)`, using `na`=`nb`=`nc`=`1`,...,`4` and `nk`=`1`,...,`3`
- `OE(nb,nf,nk)`, using `nb`=`nf`=`1`,...,`4` and `nk`=`1`,...,`3`

by clicking on *Estimate*, selecting *Polynomial Models*, choosing as *Structure* the desired model class, clicking on the *Order Editor* or specifying in *Orders* the model orders and the input-output delay, and finally clicking on *Estimate* in the *Polynomial Model* window: the corresponding window will appear in the *Model Views* area

9) Compare the identified models on the VS dataset considering the Best Fit index, to be maximized (click on the *Model output* option), and the residual analysis (click on the *Model resids* option).
For the whiteness residual test of each model, the Autocorrelation plot has to be considered: the more residual values are inside the confidence interval, the better is the model; the number of residuals sufficiently outside the confidence interval have to be counted and compared with a threshold: if this number is greater than the threshold, then the model is wasted; otherwise, the model is considered for further analyses. A reasonable threshold is 4 or 5, corresponding to a 95% confidence level (note that this threshold may be increased, depending on the case study!!)

10) Select and export the "best trade-off" model (click the corresponding window in the *Model Views* area and drag it to the *To Workspace* box)