

Introduction à l'identification

Présentation rapide et incomplète de la toolbox *Ident* de Matlab

Vincent MAHOUT

March 13, 2017

Idée de base

- Utilité : fournir des modèles simplifiés (linéaires) de systèmes complexes à partir de données échantillonnées (bruitées).
- L'identification consiste donc en la construction de modèle en ajustant ses paramètres pour qu'ils coïncident au mieux avec les mesures
- La validation d'un modèle consiste à tester l'adéquation de celui ci sur un jeu de données différent de celui utilisé pour fabriquer le modèle
- La toolbox propose 2 approches différentes
 - Une approche basée modèle, qui sur la base du choix d'une famille de modèle, opère à une estimation paramétrique
 - Une approche non paramétrique qui à partir d'une analyse de corrélation (resp. spectrale) recherche la meilleure réponse impulsionnelle ou indicielle (resp. fréquentielle).

Le principe de l'identification

- L'identification d'un système fait intervenir 6 éléments :
 - Le **système réel** : mesures des grandeurs d'E/S
 - Le(s) **modèle(s)** préconisés : choix d'un type de modèle, de l'ordre, de la nature des perturbations...
 - Un **critère** : fonction scalaire qui permet de mesurer la fidélité du modèle supposé
 - Un **optimiseur** : algorithme d'optimisation qui cherche à minimiser le critère utilisé
 - **Incertitudes des paramètres** : il n'existe pas de modèles parfaits
 - Une **analyse critique** des résultats : jusqu'à quel point le modèle est mauvais

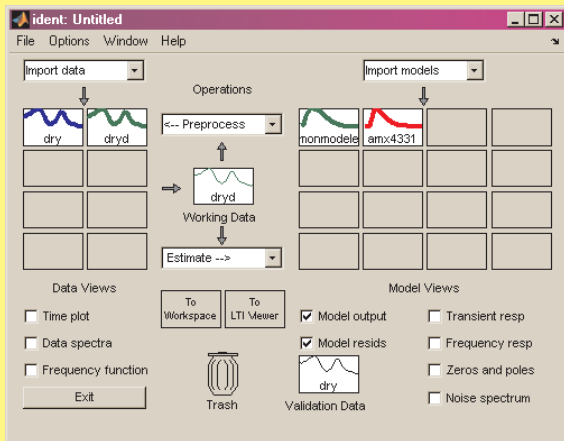
Vocabulaire d'usage

- **Données d'estimation**: jeu de données sur lequel est construit le modèle
- **Données de validation**: jeu de données, différent du précédent, sur lequel est validé le modèle trouvé
- **Structure du modèle**: famille de modèles dont les paramètres sont ajustables
- **Méthode d'identification paramétrique** : méthode qui a pour but d'estimer (au mieux) les paramètres d'un modèle donné (moindres carrés,...)
- **Méthode d'identification non paramétrique** : méthode qui ne nécessite pas obligatoirement la connaissance a priori d'un modèle (corrélation, analyse spectrale,...)
- **Validation de modèle** : test de la confiance que l'on peut apporter à un modèle soit en le simulant à partir de jeux de données différent, soit en analysant les propriétés des résidus

Rôle de la toolbox ident

- Cette boîte à outils propose de nombreuses fonctions qui permettent d'extraire un ou des modèles à partir de différentes campagnes de mesures.
- Il existe également une interface graphique (commande *ident*) qui, bien qu'elle limite l'utilisation des fonctions de base qui tournent derrière, permet d'utiliser la toolbox sans connaître la syntaxe des fonctions qui la constitue.

L'interface graphique



L'interface graphique

Données d'expérimentation

ident: Untitled

File Options Window Help

Import data

dry dry

Operations

<-- Preprocess

dry

Working Data

Estimate -->

Import models

nonmodele amx4331

Data Views

Time plot

Data spectra

Frequency function

Exit

To Workspace To LTI Mewer

Trash

Model Views

Model output

Model resid

Transient resp

Frequency resp

Zeros and poles

Noise spectrum

Validation Data

dry

L'interface graphique

Données d'expérimentation

Traitement

ident: Untitled

File Options Window Help

Import data

dry dry

Operations

<-- Preprocess

dry

Working Data

Estimate -->

Import models

monmodele amx4331

Data Views

Time plot

Data spectra

Frequency function

Exit

To Workspace To LTI Mewer

Model Views

Model output

Model resid

Transient resp

Frequency resp

Zeros and poles

Noise spectrum

Trash Validation Data

dry

L'interface graphique

Données d'expérimentation

Traitement

Modèles identifiés

ident: Untitled
File Options Window Help

Import data

dry dry

Operations

<-- Preprocess

dry

Working Data

Estimate -->

Import models

nonmodele amx4331

Model Views

Model output Transient resp

Model resids Frequency resp

Zeros and poles

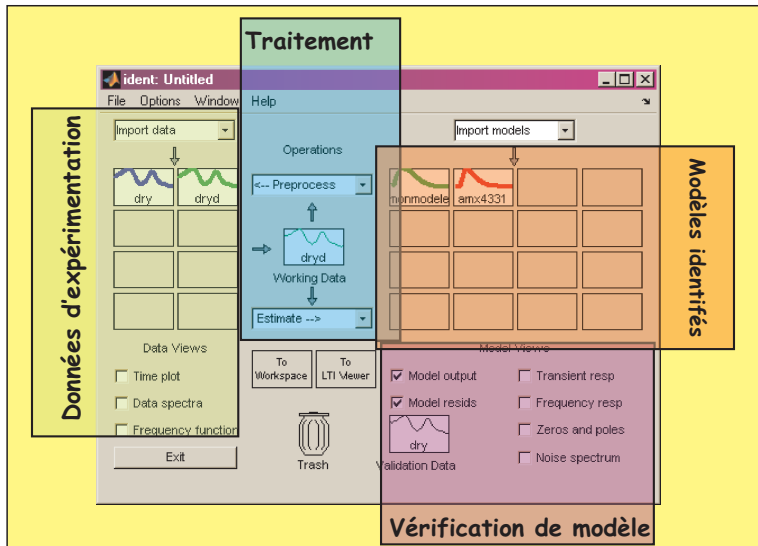
Noise spectrum

To Workspace To LTI Mewer

Trash Validation Data

dry

L'interface graphique

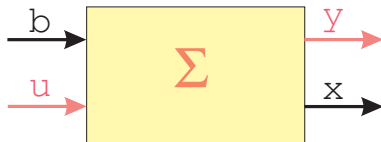


Fonctionnalités de base

- Utilisation de campagne de mesure
- Prétraitement de l'information
- Identification d'une structure de modèle
- Test de validité de l'identification
- Visualisation des réponses (temporelles, fréquentielles,..)
- Conversion dans une structure "connues"

Le système : la campagne de mesure

- Différents process à choisir
- Campagne de mesure à l'aide de *Simulink Real Time*



- Grandeurs accessibles : u et y
- Grandeurs inaccessibles : b bruit en entrée et x variables internes
- Ne pas se contenter d'une seule réponse indicielle
- Utiliser des signaux d'entrée riches en fréquence (SBPA : Séquence Binaire Pseudo Aléatoire)

Les pré-traitements

- Différents pré-traitements permettent d'améliorer les signaux sur lesquels on travaille
- Possibilité de filtrage
- Possibilité de rééchantillonnage
- Possibilité de combiner des expérimentations différentes
- Possibilité d'isoler des entrées et/ou des sorties
- Possibilité d'enlever les valeurs moyennes (offset) : meilleure convergence pour l'optimisation

- **Un prétraitement engendre un nouveau jeu de données**

Les structures de modèle

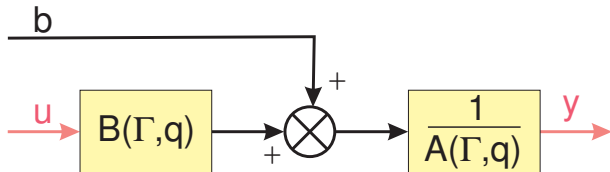
- On ne s'intéressera qu'à l'identification paramétrique (\Rightarrow structure du modèle connue)
- Différents structures de modèles sont couramment utilisées en identification
- La plupart font intervenir la notion de bruit d'entrée
- Tous les modèles sont discrets
- Pour effectuer l'identification, il faut fixer l'ordre du modèle (\Rightarrow problème crucial : lien entre performance et complexité)
- Sous *Matlab* il est possible ensuite de revenir à des formats plus conventionnels pour effectuer les synthèses de correcteurs et la simulation (fonction *tf* et *ss* par exemple)

La structure ARX

- Le bruit intervient directement
- Le système s'écrit

$$A(q)y(t) = B(q)u(t - n_k) + b(t)$$

- Il y a trois paramètres à spécifier n_a la dimension du polynôme $A(q)$, n_b la dimension du polynôme $B(q)$ et le retard n_k

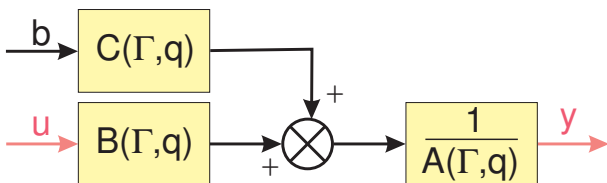


La structure ARMAX

- Le système s'écrit

$$A(q)y(t) = B(q)u(t - n_k) + C(q)b(t)$$

- Il y a quatre paramètres à spécifier n_a la dimension du polynôme $A(q)$, n_b la dimension du polynôme $B(q)$, n_c la dimension du polynôme $C(q)$ et le retard n_k

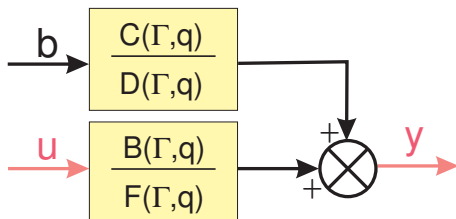


La structure Box Jenkins (BJ)

- Le système s'écrit

$$y(t) = \frac{B(q)}{F(q)} u(t - n_k) + \frac{C(q)}{D(q)} b(t)$$

- Il y a cinq paramètres à spécifier n_f la dimension du polynôme $F(q)$, n_b la dimension du polynôme $B(q)$, n_c la dimension du polynôme $C(q)$, n_d la dimension du polynôme $D(q)$ et le retard n_k

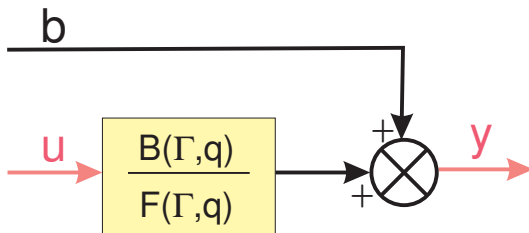


La structure output error (OE)

- Le système s'écrit

$$y(t) = \frac{B(q)}{F(q)} u(t - n_k) + b(t)$$

- Il y a trois paramètres à spécifier n_f la dimension du polynôme $F(q)$, n_b la dimension du polynôme $B(q)$ et le retard n_k



Le critère

- Très généralement le critère est de forme quadratique

$$J = (y - \hat{y})^T Q (y - \hat{y})$$

avec Q une matrice symétrique définie positive, (\hat{y}) la sortie du modèle et y la sortie mesurée.

- Lorsque Q est diagonale le critère peut s'écrire

$$J = \frac{1}{N} \sum_i \sum_{j=1}^N w_i (y_{i,j} - \hat{y}_{i,j})^2$$

où i indice le nombre de sortie et j le nombre d'échantillons. Le critère est généralement divisé par le nombre d'échantillons pour comparer les critères sur des nombres de données différents.

- Le poids relatif w_i attribué aux différentes sorties permet de pénaliser plus ou moins certaines variables.
- Il existe d'autres critères : valeurs absolues, maximum de vraisemblance,...

L'algorithme d'optimisation

- Cherche à adapter les paramètres du modèle en vue de minimiser le critère
- Différents algorithmes : moindres carrés, méthode de Gauss-Newton, filtrage de Kalman,...
- Le choix de l'algorithme dépend de la structure du modèle envisagé, des hypothèses faites sur la nature du bruit et du critère utilisé.
- On ne cherchera pas dans ce micro projet à jouer sur le paramétrage du critère et/ou de l'algorithme d'optimisation

L'incertitude des paramètres - l'analyse critique

- Un modèle obtenu est toujours imparfait
- Le jeu de paramètre obtenu correspond juste à un minimum du critère (des fois local)
- Quelle est la validité du résultat ?
- Première réponse : multiplier les essais, utiliser plusieurs modèles, jouer sur les nombreux paramètres de ceux ci.
- Deuxième réponse : utiliser un jeu de données différent pour valider le modèle.
- Deux analyses sont directement accessibles dans l'interface graphique : la visualisation de la sortie modèle (avec un calcul de % sur la fiabilité du modèle) et l'analyse des résidus (test de corrélation entre l'entrée et l'erreur de sortie).

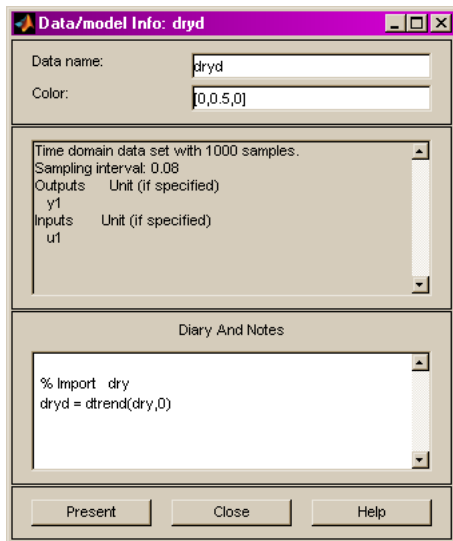
Exemple du tutorial matlab

- Chargement d'un fichier de mesure
 - » *load dryer2*
- Mise au format "ident" des vecteurs d'entrée $u2$ et $y2$
 - » *dry = iddata (y2,u2,0.08)*
- Informations (modifiables) sur la structure de données
 - » *get(dry)*

```
Domain: 'Time'  
Name: []  
OutputData: [1000x1 double]  
y: 'Same as OutputData'  
InterSample: 'zoh'  
OutputName: 'y1'
```

```
ExperimentName: 'Exp1'  
InputData: [1000x1 double]  
u: 'Same as InputData'  
InputName: 'u1'  
Ts: 0.0800  
...
```

Import dans l'interface graphique



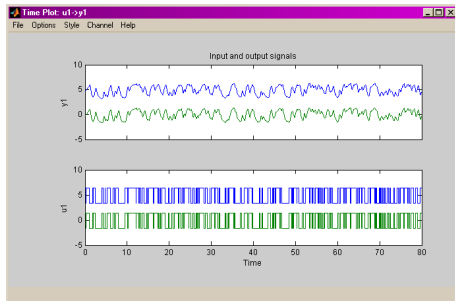
Prétraitement : retrait des valeurs moyennes

The screenshot shows the 'ident' software interface with the following components:

- Title Bar:** 'ident: Untitled' with standard window controls.
- Menu Bar:** 'File', 'Options', 'Window', 'Help'.
- Importers:** 'Import data' and 'Import models' dropdown menus.
- Data Views:** A grid of six icons. The top-left icon shows a blue wave labeled 'dry'. The top-right icon shows a green wave labeled 'dryd'. Below it are four empty icons.
- Operations Menu:** A dropdown menu is open, listing:
 - <-- Preprocess (highlighted)
 - <-- Preprocess
 - Select channels...
 - Select experiments...
 - Merge experiments...
 - Select range...
 - Remove means (highlighted by a mouse cursor)
 - Remove trends
 - Filter...
 - Resample...
 - Transform data...
 - Quick start
- Model Views:** A grid of six empty icons.
- Checkboxes:**
 - Time plot
 - Data spectra
 - Frequency function
 - Model output
 - Model resids
 - Transient resp
 - Frequency resp
 - Zeros and poles
 - Noise spectrum
- Buttons:** 'Exit' button.
- Icons:** A trash can icon labeled 'Trash' and a wave icon labeled 'Validation Data' with 'dry' text below it.
- Status Bar:** 'Data set dryd inserted. Double click on icon (right mouse) for text information.'

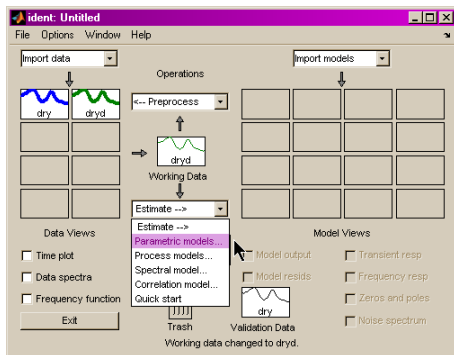
Visualisation des données

- Si l'option est *time plot* est cochée, tous les jeux de données sélectionnés sont affichés dans une fenêtre séparée.
- Un simple clic sur un jeu de donnée permet de basculer le selectionnement
- Un double clic permet d'ouvrir une fenêtre d'information sur ce jeu



Identification paramétrique

- Ouverture du bandeau *Estimate* conduit à une sous fenêtre où l'on effectue le choix de la structure du modèle
- On aura pris soin de placer (drag & drop) dans le *working data* le jeu sur lequel on veut faire l'estimation



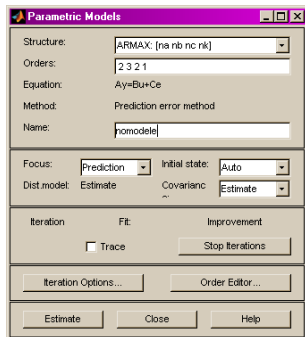
Identification paramétrique

The screenshot shows the 'ident' software interface with the following components:

- Title Bar:** 'ident: Untitled' with standard window controls.
- Menu Bar:** File, Options, Window, Help.
- Import data:** A dropdown menu and a grid of data views. The first row shows two plots labeled 'dry' (blue) and 'dryd' (green).
- Operations:** A central workflow area with a dropdown menu set to '<-- Preprocess'. Below it is a 'dryd' plot and the label 'Working Data'.
- Estimate -->:** A dropdown menu with the following options:
 - Estimate -->
 - Parametric models...** (highlighted)
 - Process models...
 - Spectral model...
 - Correlation model...
 - Quick start
- Model Views:** A grid of model view options, each with a checkbox:
 - Model output
 - Model resids
 - Transient resp
 - Frequency resp
 - Zeros and poles
 - Noise spectrum
- Data Views:**
 - Time plot
 - Data spectra
 - Frequency function
 - Exit button
- Trash:** A trash icon at the bottom center.
- Validation Data:** A 'dry' plot at the bottom right.

Choix du modèle

- On choisit ensuite la structure du modèle envisagé
- On fixe les ordres des différents modèles (on peut donner un nom autre que celui par défaut)
- On lance l'estimation, ce qui crée le modèle dans l'environnement



Choix du modèle

Parametric Models

Structure: ARMAX: [na nb nc nk]

Orders: 2 3 2 1

Equation: $Ay=Bu+Ce$

Method: Prediction error method

Name: nomodele

Focus: Prediction Initial state: Auto

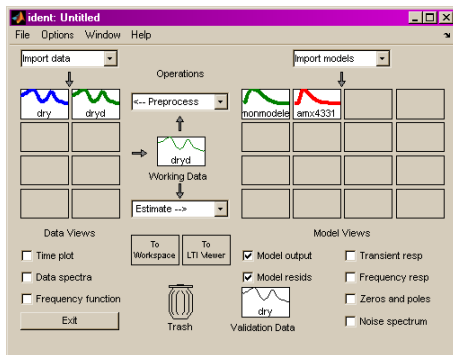
Dist.model: Estimate Covarianc: Estimate

Iteration Fit: Improvement

Trace

La vérification

- A partir des différents modèles créés (pour différentes structures et différents ordres) on peut réaliser des vérifications simples
- Ces vérifications se fait sur le jeu *validation data* qui doit être de préférence différent du jeu d'estimation



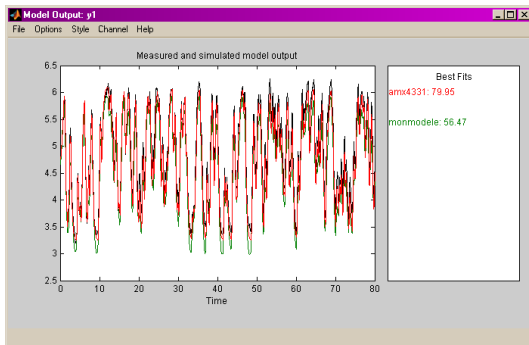
La vérification

The screenshot shows the 'ident' software interface with the following components:

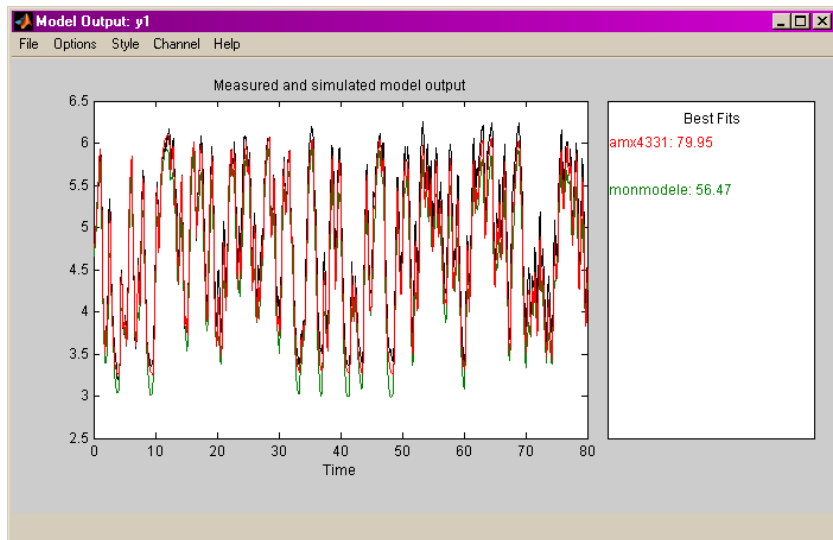
- Title Bar:** 'ident: Untitled' with standard window controls.
- Menu Bar:** File, Options, Window, Help.
- Import Data:** A dropdown menu labeled 'Import data' with a grid of data plots. The first row contains 'dry' (blue wave) and 'dryd' (green wave).
- Operations:** A central vertical flow:
 - '<-- Preprocess' dropdown menu.
 - 'dryd' plot icon.
 - 'Working Data' label.
 - 'Estimate -->' dropdown menu.
- Import Models:** A dropdown menu labeled 'Import models' with a grid of model plots. The first row contains 'nonmodele' (green wave) and 'amx4331' (red wave).
- Data Views:**
 - Time plot
 - Data spectra
 - Frequency function
 - Exit button
- Model Views:**
 - Model output
 - Model residu
 - Transient resp
 - Frequency resp
 - Zeros and poles
 - Noise spectrum
- Workspace:** 'To Workspace' button.
- LTI Viewer:** 'To LTI Viewer' button.
- Trash:** Trash can icon labeled 'Trash'.
- Validation Data:** 'dry' plot icon labeled 'Validation Data'.

La sortie modèle

- Les différents modèles sélectionnés apparaissent en tracé temporel (simulation du modèle sur le jeu de vérification)
- Il apparaît un "taux de recouvrement" exprimé en % qui indique la bonne adéquation

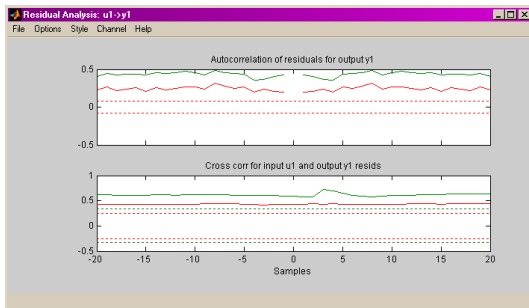


La sortie modèle

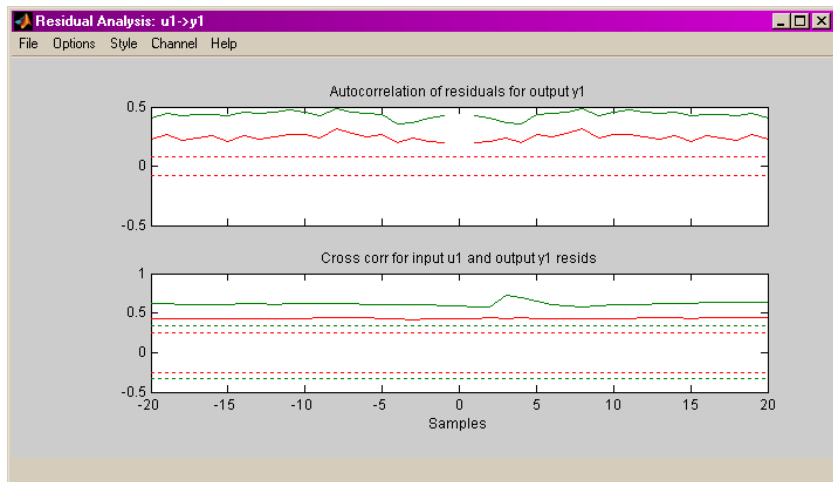


Les résidus

- Le graphique montre s'il existe encore de l'information non exploitée dans le signal de sortie du modèle
- Cela consiste à calculer l'autocorrélation sur le signal d'erreur de sortie ainsi que la corrélation entre l'entrée et le signal d'erreur.
- Ces mesures doivent être le plus proche de 0



Les résidus



Et ensuite

- Il semble raisonnable de sélectionner 2 ou 3 modèles sur lesquels on pourra faire les synthèses
- Il faut penser également que les modèles peuvent n'avoir qu'une validité locale (système non linéaire par exemple pour lequel on travaille autour d'un point de fonctionnement).
- La meilleure vérification est celle des performances en boucle fermée (comparaison des performances attendus et différences avec la simulation).