A STATISTICAL METHODOLOGY FOR MODEL VALIDATION IN THE ALLAN.™SIMULATION ENVIRONMENT

Maria del Mar IZQUIERDO* **Gilles LEFEBVRE*** **Elena PALOMO***
**Groupe Informatique et systèmes Energétiques EMP/ENPC*
Fabrice BOUDAUD** **Alexandre JEANDEL****
***Gaz de France, Research and Development Division*

Abstract
This paper deals with the model validation methodology used at the Gaz de France Research & Development Division. The primary emphasis is on the latest developments, concerning different statistical methods for model validation and diagnosis. The corresponding computer implementation is called DVM standing for "Diagnostic et Validation de Modèles".

1. Introduction

For many years, the Gaz de France Research & Development Division has been using modelling and simulation techniques as a means to study systems (Cassagne 88). The approach adopted by GDF is based on the capitalization and reuse of models. Major methodological and applicative work has been undertaken in this area (Jeandel 93-91, Favret 88, Boulkroune 93, Givois 95, Ponttiggia 95).

Validation plays a dual role for the modeller :
- as a modelling aid, guiding the choice of an effective model structure and associated numerical values with respect to the model's specific utilization,
- as an aid to model reuse, by simplifying access to models by a third party.

Moreover, in the context of a client/supplier relation between the modeller and the person or body commissioning the study, it should enable the client to accept delivery of the model.

It should be remembered that a model is only meaningful with respect to a particular utilization objective. It is therefore not possible to qualify the model as definitively "valid". It is only possible to attempt a description of its validity domains.

**Groupe Informatique et systèmes Energétiques*
La Courtine 93167 Noisy le Grand - FRANCE
email : lefebvre@cenerg.enpc.fr - tel : 33.1.49143669

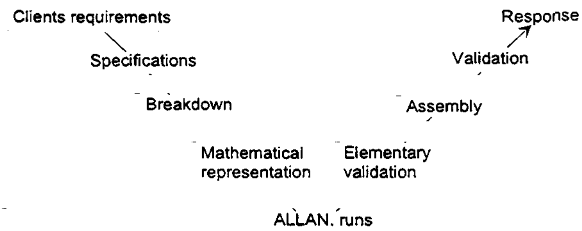
***Gaz de France, Research & Development Division*
93211 La Plaine Saint Denis Cedex - FRANCE
email:alexandre.jeandel@dr.gdf.fr - tel:33.1.49225697

Our work involves the development of working methods, then, if possible, of corresponding utility programs and finally, validation of the implementation of these tools in our ALLAN.™Simulation+ modelling/simulation environment. Statistical methods were so assessed by GISE with the development of computer tools called DVM.

2. Validation strategies

2.1. Validation and modelling approach

Gaz de France has defined a structured working approach. This standard "V" approach includes a certain number of iterations on simple then compound models, repeated until the needs of the study are met.



Our approach comprises a number of key points with respect to validation. We will simply mention these specific elements, without presenting the complete process (Jeandel 93). We call this validation approach "SEPA", standing for "Suivi d'Etudes Pour ALLAN".

+ ALLAN. is a Gaz de France trademark

The table below presents a synthesis of the relationships between the validation elements and our modelling approach.

ALLAN refers to our modelling/simulation tool, "SEPA" to our modelling method, ULM to our neutral modelling language (Jeandel 95), PROTOP to a utility program for optimization of excitation protocols produced by ADERSA and DVM to the utility program developed by GISE.

The validation actions do not need the use of a comparison reference and can be seen as model characterization tools. Results of this work can define a model visiting card which is very helpful for third party model access.

Stage	Validation action	Method	Utility programs	Tools
Requirements and specifications	* Definition of client's simulation requirements	SEPA		
Breakdown	* Choice of models to be qualitatively validated * Choice of models to be experimentally validated	SEPA SEPA		
Mathematical representation	* Definition of qualitative validations * Optimization of experimental protocols * Definition of value domains of variables and parameters * Verification of dimension equations * Choice of modifiable parameters	SEPA SEPA/ PROFORMA SEPA	ULM PROTOP	 ALLAN 3.0 ALLAN 3.0
Elementary validation	* Determination of static and dynamic parameters * Model identifiability * Analysis of I/O relationships * Parametric sensitivity analysis * Frequency analysis * Qualitative validation * Empirical validation	-	DVM DVM DVM DVM DVM	
Assembly validation of compound models	* Assembly of value domains * Assembly of frequency domains * Same approach as for elementary validation	 SEPA		ALLAN 3.0
Validation	* Fitting of parameters * Client simulation runs		ALLAN/MATLAB or ALLAN/BASILE links	

2.2. Types of validation

We have seen that validation is based on the comparison between the behaviour of the model and a reference. Depending on the choice of reference, there are a number of types of validation :

- numerical validation and software intercomparison,
- empirical validation,
- qualitative or analytical validation

Different studies have been carried out in the past in an attempt to establish a methodology for model validation. The first study, undertaken by the US Solar Energy Research Institute (Judkoff 83), resulted in a three-part methodology including analytical tests, inter-model comparisons, and empirical validation. This methodology was futher refined and extended in the second study carried out by four British research teams - the University of Nottingham, Leicester Polytechnic, the Rutherford Appleton Laboratory and the Building Research Establishment (Bloomfield 88). The methodology comprises: theory and source code checking, analytical tests, inter-model comparison, sensitivity analysis, and empirical validation. This was the methodology reviewed and accepted at the commencement of the CEC Concerted Action PASSYS (Jensen 89). In the second phase of PASSYS main emphasis was devoted to empirical model validation. The complete description of the resulting methodology can be found in (Jensen 93) and, in a more condensed form, in (Jensen 93, Palomo 91, Palomo 93).

Numerical validation is generally defined as the comparison of simulation results of a single model run on two different computer programs. It is of very little use and should be avoided. Indeed, it can only be useful if the model is known to be identical in both cases and if the computer implementation and numerical programming of both programs are known to be flawless.

Another type of validation is the comparison of two different models of the same system under the same simulation environement. This is only useful in the case of model reduction from one model to the other. It compares the two validity domains of the models.

In **empirical validation** (Palero 90, Palomo 91, Jensen 93, Palomo 93, Ramdani 94), the model is compared with "reality" or, more exactly, with the measurements on the system. Beyond any technical consideration, it provides a guarantee of client confidence.

It provides a great deal of information, enabling the modeller to :

- improve his understanding of the system he is modelling,
- improve his understanding of his test bench instrumentation,
- modify his model.

This type of validation also has its drawbacks. In particular, it requires specific testing protocols. Upon analysis, we note that standard equipment evaluation tests are not well suited to model validation, tests with regulated values must be avoided, correlated excitations must be avoided, and test lengths should be adapted to the excited time constants.

Moreover, these tests must always be repeated, since experience has shown that two tests performed under apparently identical conditions can give rise to differing results.

Qualitative validation is defined as the comparison of simulation results with the modeller's mental picture of reality. It enables him to ensure he has really modelled what he intended to model. It is the only possible type of validation for equipment design.

In this context, a number of different types of excitation can be considered: standard excitations, nominal or legally enforced conditions, specific solutions calculated by hand or using specific tools, extreme or limit conditions.

Gaz de France systematically validates all its models. Qualitative validation is considered to be essential.

The table below summarizes the relative advantages and drawbacks of empirical and qualitative validation.

Qualitative validation		**Empirical validation**	
+	--	+	--
Modeller's confidence		Customer's confidence	
	Comparison with his own wishes	Comparison with measurements	
	Limited by the modeller's perception of his own model	New phenomena may be detected between measured quantities (analysis of correlations)	
Wider domain of validity	Less accurate results for a given case	Enables model fitting	
Possible in all cases			Specific test equipement required
Wide range of possible scenarios	Risk of extending beyond realistic excitation patterns		Difficult to develop pertinent scenarios : -narrow range of possible excitations -limited to a particular system
Reproducible			Requires numerous test to achieve a confidence interval
Problem well covered			Introduces new sources of error
Possibility of covering and thus defining the validity range of a model : -parametric sensitivity -excitation combinations			Limits the fiels of exploration of a model's validity range
			Transforms the modeller's task into a problem of getting curves to coincide
			Compares with the measured quantity and not the exchange quantity of the physical phenomenon

3. "DVM"

3.1. Working method

Gaz de France asked for help from GISE to assess the application of statistical methods in which they have expert knowledge. The objective is to validate working methods so that they can be brought to an operational phase in the form of computer tools with accompanying training sessions. Part of this work should help to define a model validation standard to improve exchange of models.

In order to test the methods and the software produced, Gaz de France provided GISE with "black box" model couples, a reference model representing reality and a second model to represent its approximation. The two models represented a system unknown to GISE. This enabled us to concentrate on the methods used and to assess their pertinence independently of any problems associated with knowledge and measurement of the reference.

3.2. Applied methodology

The mathematical proposed methodology is based on a separate analysis of the so-called static behaviour and the "purely" dynamic one of the respective models.

Let
$$\dot{X}(t) = F(X(t), U(t); \theta)$$
$$Y(t) = G(X(t), U(t); \theta)$$

be a general dynamic model *M*. *X(t)* is the vector of state variables, *U(t)* the vector of excitation variables (inputs), and *Y(t)* the vector of the model outputs. *F* and *G* are matrices of general non-linear functions, and θ the vector of the model's parameters.

The static behaviour of the system is described by
$$0 = F(X(t), U(t); \theta)$$
$$Y^s(t) = G(X(t), U(t); \theta)$$
and the difference
$$Y^d(t) = Y(t) - Y^s(t)$$
is called the purely dynamic behaviour ("dynamic" behaviour in the following) of the system.

Three main steps can then be distinguished in the methodology :

- **Input design.** Informative enough inputs signal must be built. Their form, spectrum, length, sampling time, ... must be chosen as to give the main characteristics of the system.
- **Simulations.** The response of the models to the above input signals must be obtained.
- **Analysis.** Different methods and tools must be used to characterize the observed differences in the models' behaviour. The aim of the analysis is twofold: a) to determine a domain of application of the simplified model (qualification); and b) to disaggregate the causes of a poor simplified model performance (diagnosis)

Input design and validation methods will depend onthe regime under analysis. In section 3.3, the methodology for the static behaviour analysis is presented, and section 4 includes the description of the methods and procedures proposed for analyzing the "purely" dynamic behaviour. "DVM" is the utility program allowing us to apply such methods and procedures. Different couples of representative models (linear models, non linear models, and non linear models with hysteresis) were analyzed using "DVM". The simplest one (2 inputs - 1 output linear system) is used here to illustrate the proposed methods for model validation.

3.3. Analysis of the static behaviour

3.3.1. Input design

Sequences of step functions are used as inputs to observe the static behaviour of the models.

Let $U(t) = [u_1(t), u_2(t), \ldots\ldots, u_n(t)]^t$
be the vector of excitation variables, each component varying from predefined minima and maxima values (input domain, D_u,),
$$u_i \in [u_{i,\min}, u_{i,\max}]; \quad i = 1, 2, \ldots, n$$

A discrete input domain, $\tilde{D}_u$, is formed by subdividing the above intervals in a number n_i of homogeneous sub-intervals. The allowed values for u_i are noted as
$$u_i^k = \left[u_{i,\min} + \frac{\Delta u_i}{2}\right] + (k-1)\Delta u_i$$
$$\Delta u_i = \frac{u_{i,\max} - u_{i,\min}}{n_i}; \quad k = 1, 2, \ldots, n_i$$

The resulting domain is then composed of $N = \prod_1^n n_i$ points (all the possible combinations among the u_i^k values). *m* points are then selected by a random procedure (uniform distributions are used), leading to a sequence $\{U^1, U^2, \ldots\ldots, U^m\}$ of values for *U*. The vector of excitations is then built linking together *m* step functions :
$$U(t) = \begin{cases} U^{j-1} & t_0 + (j-2)\Delta t_s \le t < t_0 + (j-1)\Delta t_s \\ U^j & t_0 + (j-1)\Delta t_s \le t < t_0 + j\Delta t_s \end{cases}$$

The length Δt_s of the step functions must be chosen to insure, whatever may be the values of the *U*-components, the observability of the models' static behaviour. In the following, we will also make reference to the vector of input increments,
$$\Delta U(t) = U(t) - U(t-1).$$

The continuous domain of input variation in the example is:
$$u_1 \in [-8000, 8000], \quad u_2 \in [0.07, 0.18]$$

These intervals have been subdivided into $n_1 = n_2 = 10$ homogeneous subintervals to define the discrete inputs domain (100 points). The sequence $\{U^1, U^2, \ldots.\}$ of values for *U(t)* has 500 elements.

A value of $\Delta t_s = 200s$ is chosen. The total length of the input vector is thus 100000s.

3.3.2. Simulations and static behaviour determination

Simulations are carried out using the above excitation vector. *Y(t)* is the output vector from the reference model, and $\tilde{Y}(t)$ the one from the simplified model. The static response of both models is then determined as (it is assumed that a pertinent value of Δt_s was chosen):

$$Y^s(U^j) = Y(t_0 + j\Delta t_s)$$
$$\tilde{Y}^s(U^j) = Y(t_0 + j\Delta t_s); \quad j = 1,2,\ldots m$$

If the system represented by the models has hysteresis phenomena, its static behaviour will also depend on $\Delta U^j = U^j - U^{j-1}$. Hence, it is better to write $Y^s(U^j, \Delta U^j)$ and $\tilde{Y}^s(U^j, \Delta U^j)$.

It must be noted that the finer the discretization of the continuous domain of input variation and the random exploration of the resulting discrete domain were, the more the information there will be concerning the static behaviour available at this step.

3.3.3. Methods and procedure for validation

The differences between the static behaviours of the reference model and the simplified model are measured by means of a relative error. For each component in vectors Y^s and $\tilde{Y}^s$, this error is given by

$$\varepsilon_k(U^j) = \frac{y_k^s(U^j) - \tilde{y}_k^s(U^j)}{y_k^s(U^j)} 100\%$$
$$j = 1,2,\ldots,m; \quad k = 1,2,\ldots,q$$

For a given tolerence τ, the domain of application of the simplified model (static behaviour) is defined as

$$U^j \in \tilde{D}_u \quad / \quad \forall k \quad \varepsilon_k(U^j) \le \tau$$

The domain of application (τ = 0.1 %) for the simplified model in the example is represented in figure 1. The crosses are the points of the discrete input domain and the squares, the points belonging to the application domain.

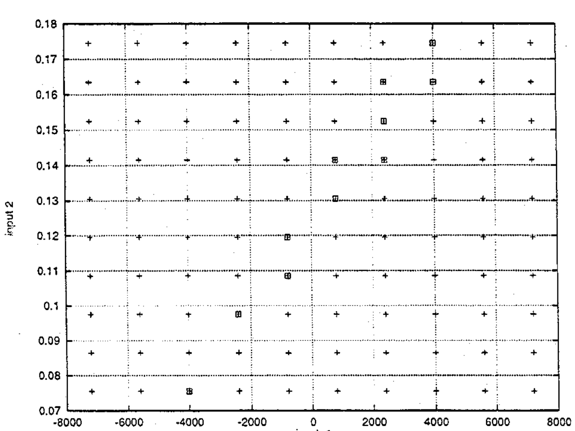


Figure 1 : Application domain of the simplified model

A graphical procedure could then be used to disaggregate the causes of the observed differences. The clouds of points $Y^s(U, \Delta U)$ and $\tilde{Y}^s(U, \Delta U)$ gotten in 3.3.2, are projected on different plans to show the functional nature of the static input-output relationships associated with the models. By visual comparisons of the shape and position of the resulting curves, the nature of the observed discrepancies between the static behaviours of the models can be analysed. The differences concerning shape indicate that a functional problem exists in the simplified model (e.g. a non-linear input-output relationship in the reference model in contrast to a linear one in the simplified one). The differences concerning position, however, indicate that the problem has a parametric nature (e.g. different slopes in linear input-output relationships).

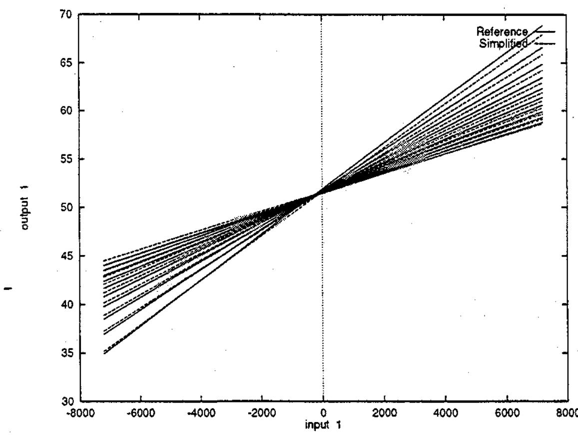


Figure 2 : Static behaviour. Projection on a plane (input 1 - output)

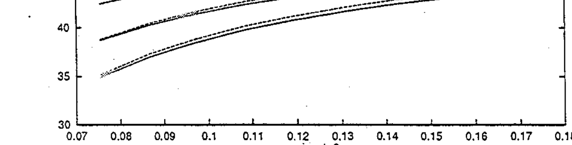


Figure 3 : Static behaviour. Projection on a plane (input 2 - output)

Figures 2 and 3 include the static input-output relationships for the models of the example. Figure 2 comes from the projection of their static behaviour on the plane (y, u_1), and figure 3 is formed by projecting it on the plane (y, u_2). It can be clearly seen that the differences between both models ("reference" = continuous lines; "simplified" = non-continuous lines) have a parametric nature. The (y, u_1) input-output relationships are linear. The slope of the different straight lines in figure 2 (different u_2 values) depends on the u_2 value, which has a multiplier effect on u_1. In contrast, the (y, u_2) input-output relationships are non-linear. They show an exponential behaviour.

3.3.4. ADDITIONAL INFORMATION

Although step functions are not a kind of input specially pertinent for the analysis of dynamic behaviour (the variance is concentrated at low frequencies), some interesting information concerning the functional nature of such regimes can be gotten from the simulations carried out in 3.3.2.

The "purely" dynamic behaviour of the models $Y^d(t) = Y(t) - Y^s(U(t))$; $\tilde{Y}^d(t) = \tilde{Y}(t) - \tilde{Y}^s(U(t))$ is first integrated over the periods Δt_s of the step functions in *U(t)*

$$I^j = \int_{t_0+(j-1)\Delta t_s}^{t_0+j\Delta t_s} Y^d(t); \quad \tilde{I}^j = \int_{t_0+(j-1)\Delta t_s}^{t_0+j\Delta t_s} \tilde{Y}^d(t)$$

If the dynamic input-output relationships in the models are linear, the above integrals are proportional to the increments $\Delta U^j = U^j - U^{j-1}$. A graphical procedure, similar to the one described in 3.3.3, is then proposed to test the linearity of the dynamic parts of the models. For each input-output couple, e.g. $y_k - u_i$, the integral values associated to y_k^d with $\Delta u_j (j \neq i) = 0$ are projected on the plane (y_k^d - integrals, Δu_i). A visual analysis of the resulting curves allows us to determine the linear/non-linear nature of the dynamic y_k, u_i relationship. As in the previous section, by comparing the shape and position of the curves from the reference models with the ones from the simplified model, the functional/parametric nature of the observed differences between the models can be established.

For the models of the example, the differences between the dynamic behaviour of the models are of a parametric nature. The output y_1 is linearly linked to the input u_1. In contrast, the dynamic relationship between the output and the input u_2 is a non-linear one. Graphs similar to the ones in figures 2 and 3 were obtained.

3.4. Analysis of the models dynamic behaviour

3.4.1. Input design

Because we are interested in the analysis of the "purely" dynamic behaviour of the models (see section 3.2.), and not in their dynamic behaviour, two different kinds of inputs will be used here: ROLBS (Randomly Ordered Logarithmically Binary Sequences) signals and a particular sequence of step functions.

A ROLBS is a binary signal (only two values allowed; e.g. $u_{i,\min}$ and $u_{i,\max}$ with periods

$$\Delta t_k = \frac{\Delta t_0}{a^{k-1}}; \quad k = 1, 2, \ldots; \quad 0 < a < 1$$

randomly ordered. As it will be shown later, one of the advantages of using binary sequences is that they allow us to calculate the "purely" dynamic behaviour of the models. The parametres that determine the spectral properties of a ROLBS signal are: the elementary period, Δt_0; the maximum period, $\Delta t_{\max} = N\Delta t_0$; and the parameter a which is used to discretize the interval $\left[\Delta t_{\max}{}^{-1}, \Delta t_0{}^{-1}\right]$ of frequencies. Selecting adequate values for these parametres (e.g. $\Delta t_{\max}$ and Δt_0 close to the longest and shortest time constants of the system) we can place the validation process in a pertinent range of frequencies. An input vector $U_{ROLBS}(t)$, whose components are decorrelated ROLBS-$(\Delta t_0, \Delta t_{\max}, a)$ signals, is then built.

Because the time constants of the systems are "a priori" unknown, an iterative process {ROLBS design-simulations-analysis} is usually required. The following "closure" procedure is proposed:

a) build a ROLBS signal with a maximum period equal to Δt_s in 3.3.1 (the longest time constant of the system is less than such a value), an elementary period equal to the time step to be used in simulations (time constants less than such a value will not be observed), and a value for the a parameter close to 1 (spectral similarity between the resulting ROLBS signal and a band-limited white noise);
b) perform simulations as indicated in section 3.4.1;
c) estimate the spectrum of the outputs (dynamic part) and determine the shortest time constant of the system (see section 3.4.2);
d) if required, go to a) now using this time a value for the elementary ROLBS period close to the time constant above.

The second input vector, $U_{SF}(t)$, will be a sequence of step functions similar to the one described in 3.3.1. As in that case, the length of such functions is Δt_s. However, the only values allowed for inputs are now their corresponding mimima and maxima values. The sequence of step functions will include all the possible combinations among these values and the ones associated with the allowed input increments. In the case of *n* inputs, a total of 2^{2n} step functions will be in the sequence (16 in the example; see table in section 3.4.1).

3.4.1. SIMULATIONS AND DYNAMIC BEHAVIOUR DETERMINATION

Simulations are carried out using the input vector $U_{SF}(t)$. The static values of the models response associated with the input values in U_{SF} are then obtained by the procedure described in 3.3.2.

$$Y^s \equiv Y^s(U, \Delta U); \quad \tilde{Y}^s \equiv \tilde{Y}^s(U, \Delta U)$$

The results achieved for the reference model in the example are included in the following table. It should be noted that, as there is no hysteresis phenomena in such a model, its static behaviour does not depend on the input increments.

Departure	Arrival (+8000, 0.18)	(+8000, 0.07)	(-8000, 0.18)	(-8000, 0.07)
(+8000, 0.18)	59.34	72.29	43.42	31.64
(+8000, 0.07)	59.34	72.29	43.42	31.64
(-8000, 0.18)	59.34	72.29	43.42	31.64
(-8000, 0.07)	59.34	72.29	43.42	31.64

The static behaviour associated with the ROLBS inputs is then calculated as :

$$Y^s(t) = Y^s(U_{ROLBS}(t), \Delta U_{ROLBS}(t))$$
$$\tilde{Y}^s(t) = \tilde{Y}^s(U_{ROLBS}(t), \Delta U_{ROLBS}(t))$$

New simulations, this time using $U_{ROLBS}(t)$ as the input vector, are performed. Let $Y(t)$ and $\tilde{Y}(t)$ be the output vectors from the reference model and from the simplified one. The so-called "purely" dynamic behaviour is then calculated as

$$Y^d(t) = Y(t) - Y^s(t); \quad \tilde{Y}^d(t) = \tilde{Y}(t) - \tilde{Y}^s(t)$$

Finally, residuals are defined as

$$\varepsilon(t) = Y^d(t) - \tilde{Y}^d(t)$$

3.4.2. Methods and procedure for validation

Different statistical analyse of the joint process {inputs, residuals} are the basis of the validation at this step. Their objectives are to determine a domain of application of the simplified model (characterisation or qualification), and to disaggregate the causes of a poor simplified model performance (diagnosis). We will assume that the reader is familiar with the different statistical concepts to be used here. Formal definitions and estimation methods can be found, for example, in (Jenkins&Watts 86), (Priesley 81), or (Cramer 75). To simplify the notation, we will assume in the following single output models.

First of all, the **pertinence of the ROLBS signal** design must be analyzed. The normalized cumulative spectrum of the ROLBS signal, $C_{ROLBS}(f)$, and the one of the simulated dynamic behaviour of the output, $C_{y^d}(f)$, are estimated. They show how the variance of these time series is distributed over frequencies. For instance, a value $C_{ROLBS}(f = 0.1) = 0.6$ indicates that 60% of the ROLBS signal variance (measure of the fluctuations around the mean value) is concentrated in the interval $[0, 0.1]$ of frequencies. This information is used here to estimate a value for the shortest time constant of the system, $\tau_{\min}$. For instance, we can approach $\tau_{\min}$ to the inverse of the frequency at which the cumulative spectrum $C_{y^d}(f)$ starts to be greater than for example 0.90-0.95.

In figure 4, both cumulative spectra were represented: the one of the output from the reference model in the example (crosses), and the one of the input ROLBS signals (squares). It can be seen that more than 95% of the output ("purely" dynamic behaviour) variance is concentrated in the interval $[0, 0.3]$ of frequencies. It can be then assumed that the shortest observable time constant of the system is about 3-4s. The parameters of the ROLBS signals were: $\Delta t_{\min} = 1s$, $\Delta t_{\max} = 3s$ and $a = 0.5$.

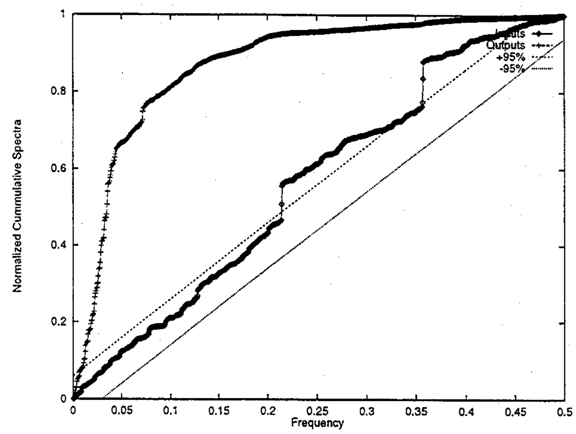


Figure 4 : Normalized cumulative input and output spectra

A test for **qualifying the simplified model** in the frequency domain has been developed. It is based on the use of the following statistics: a) the variance of the residuals, σ_ε^2, and the variance of the output from the reference model, $\sigma_{y^d}^2$; b) the normalized cumulative spectrum of the residuals, $C_\varepsilon(f)$, and the one of the output from the reference model, $C_{y^d}(f)$.

It is clear that the quotient

$$\frac{\sigma_\varepsilon^2}{\sigma_{y^d}^2}$$

gives a preliminary idea of the quality of the simplified model. In the same way, the statistic

$$\eta(f_1, f_2) = \frac{\sigma_\varepsilon^2}{\sigma_{y^d}^2} \frac{C_\varepsilon(f_2) - C_\varepsilon(f_1)}{C_{y^d}(f_2) - C_{y^d}(f_1)}$$

measures the quality of the simplified models on the interval $[f_1, f_2]$ of frequencies. The numerator represents the variance of the residuals concentrated in such an interval, and the denominator is the variance of the output from the reference model in the same interval of frequencies. The spectral domain of application of the simplified model can then be defined as all of the frequencies for which the statistic $\eta(f)$ is less than or equal to a given tolerance.

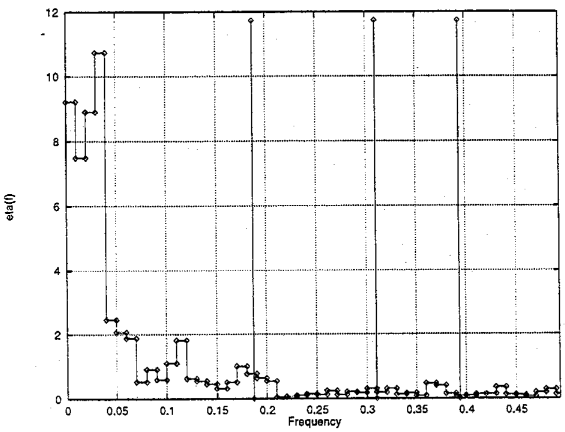


Figure 5 : Spectral domain of simplified model application

Figure 5 shows the results achieved when applying the above test to the models in the example. The statistic η measuring the quality of the simplified model is less than 2% for frequencies greater than $0.05\ s^{-1}$. Up to this frequency, the η values are usually about 8-10%. The simplified model thus shows a poor dynamic performance at low frequencies. It must not be used to simulate the system response to inputs whose variance was concentrated at such an interval of frequencies.

The **model diagnosis** is based on correlation analysis (in the time or in the frequency domains) between the observed residuals and the input signals to the models. We are looking for any causal relationship among such processes. The statistics to be used here are: the cross-correlation functions residuals-inputs, and the so-called coherency spectra.

The cross-correlation function between the residuals and any one of the inputs to the model measures the linear dependency existing between adjacent values in both stochastic processes (local correlation or interactions). If they are statistically independent the estimated cross-correlations must be, for example, inside the interval $\pm 1.96(1/N)^{0.5}$ (N = number of observations in the time series) associated with 95% limits for the assumption of zero cross-correlations (see (Box&Jenkins 76)). However, it must be noticed, that such a test is not strictly applicable to autocorrelated time series (prewhitening is usualy required). The analysis of the cross-correlation functions usually leads to very poor conclusions. Many times, it only serves to determine what inputs are clearly non-correlated with the residuals.

More useful information can be gotten in the frequency domain. The squared multiple coherency (see (Jenkins&Watts 68)) is a normalized measure of the proportion of the residuals spectrum that can be predicted from the explanatory variables of the model (inputs). It takes values from 0 to 1. Zero values mean that no correlations exist between inputs and residuals, unity values mean that residuals could be completely recovered from the inputs, and values between 0 and 1 correspond to situations where residuals can be partially predicted from inputs.

If the assumption of zero squared multiple coherency is unacceptable, the squared partial coherencies could be used to determine how much each one of the inputs contributes to explaining the residuals. Measuring the interaction between residuals and one of the explanatory variables of the model after allowance is made for the effect of the other input variables, leads to the partial correlation function in the time domain, and to the partial squared coherency in the frequency domain. Formal definitions of both statistics can be found in (Jenkins&Watts 68, Cramer 75, Priesley 81). As in the previous case, the squared partial coherency is a normalized mesure of correlation. Zero values mean independence, and unity values indicate a linear deterministic relationship.

In figure 6, the squared coherencies for the models in the example are shown: the multiple coherency (continuous line), the partial coherency with input 1 (crosses), and the partial coherency with input 2 (squares). At very low frequencies ($f < 0.02$) no correlation is detected among the residuals and the inputs to the models. There is no a linear relationship between them. On the other hand, at frequencies $0.02 < f < 0.05$, the multiple squared coherency is about 0.4 (inputs can explain a 40% of the residuals spectrum at such frequencies). However, both inputs contribute equally to recovering the residuals spectrum on this interval of frequencies (see the curves of partial coherencies).

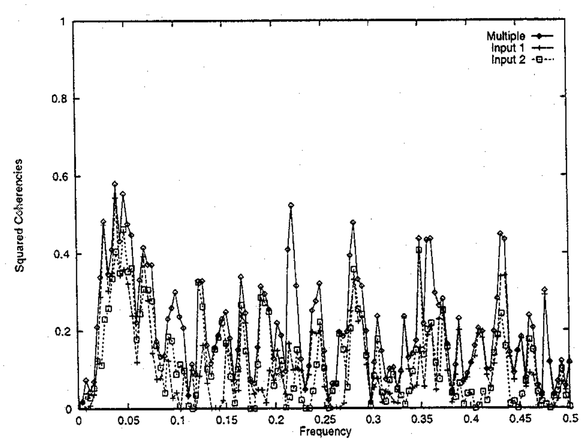


Figure 6 : Results of the diagnosis. Squared multiple and partial coherencies

4. Conclusions

The working method based on the use of black-box model couples greatly amused all the participants in this study. Gaz de France was impressed by the quality and pertinence of the analyses supplied in return. This is due to the statistical methods implemented in DVM for the static and dynamic behaviours.

Multi-step functions have been proposed as input signals to analyse the models static behaviour. The observed discrepancies between models are measured by means of a relative error, which serves to define the domain of application of the simplified model. A graphical procedure is then used to determine both the nature of the models static input-output relationships, and the nature of the observed discrepancies.

The models dynamic behaviour is observed using ROLBS signals as inputs. An efficient test has been proposed to determine the spectral domain of application of a simplified model. The diagnostic is now based on the statistical analysis of the joint process {residuals, inputs}. Cross-correlation functions and spectral coherencies have been proposed as statistics to determine the causes of bad model performance.

A certain amount of work is still needed to improve the reliability of "DVM" and to adapt it to specific needs before it enters the operational phase.

In the future, thanks to this collaboration between GDF and GISE, the R & D Division modelling teams will not only be able to further improve the quality of their models, but also to speed up the pace of their studies for technical system improvement and design by the reuse of well characterized models.

5. References

BOX, G.E.P., JENKINS, J.M. (1976), *Time series analysis : forecasting and control*, Holden-Day, San Francisco, USA.

BLOMMFIELD, D. et al. (1988), *An investigation into analytical and empirical validation techniques for dynamic thermal models of building*, Final Grant Report, Science and Engineering Research Council, Polaris House, England.

CASSAGNE, Ph., PALERO, I., LARET, L. (1989), *Gaz de france Research facilities for energy and buildings. Full scale testing and validation of models*, in "CLIMA 2000", Sarajevo, BOSNIA.

CRAMER, H. (1974), *Mathematical methods of statistics*, Princeton University Press.

FAVRET, F. (1988), *ALLAN, a general working tool which liberates the user from programming work*, in the Osaka Gas R&D Forum 88, Osaka, JAPAN, Nov 9-11.

JEANDEL, A., PALERO, I., LARET, L. (1991), *An approach of Thermal Modelling and Simulation of Building at Gaz de France.*, in Proceedings of the 1991 IBPSA Conference, Sophia Antipolis, FRANCE.

JEANDEL, A., FAVRET, F., LAPENU, L., LARIVIERE, E. (1993), *ALLAN.TMSimulation, a general tool for model description and simulation.*, in Proceedings of the 1993 IBPSA Conference, Adelaide, AUSTRALIA.

JEANDEL, A., RAVIER, P. (1995), *ULM 'Un Langage de Modélisation'*, GDF Internal report, (Saint Denis, FRANCE)

JENKINS, G.M., WATTS, D.G., *Spectral analysis and its applications*, Holden-Day, San Francisco, USA.

JENSEN, S. O. (ed.) (1989), *The PASSYS Project Phase 1. Subgroup Model Validation and Development - Final Report 1986-1989*, Commission of the European Communities, DGXII.

JENSEN, S. O. (ed.) (1993), *Model Validation and Development. Research Final Report, PASSYS, Part II*, Commission of the European Communities, DGXII.

JENSEN, S.O., VAN de PERRE, R. (1991), *Tools for whole model validation of building simulation programs. Experience from the CEC concerted action PASSYS*, Building Simulation'91, IBPSA Conference, Nice, France, August 20-22.

MOINARD, S., GIVOIS, E. (1995), *Modelling and Simulation of a Natural gas air-conditioning system*, in Proceedings (to be published) of the international Institute of Refrigeration Conference (The Hague, 20-25 august NEDERLANDS)

PALERO, I. (1990), *Méthodologie de validation du logiciel CSTBât*, GDF Internal report GDF.M.CERUG.90M193774 (Saint Denis, FRANCE).

PALOMO, E., MARCO, J., MADSEN, H. (1991), *Methods to compare measurements and simulations*, Building Simulation'91, IBPSA Conference, Nice, France, August 20-22.

PALOMO, E. (1992) *Desarrollo de una metodología teórico-experimental de caracterización de invernaderos*, Tesis Dr., Universidad Complutense de Madrid.

PALOMO, E., MARCO, J. (1993), *The pertinence and limitations of different statistical tools for empirical validation purposes*, 3er. European Conference on Architecture, Florence, Italy.

PONTIGGIA, P., BRIAND, P., BRACONNIER, E., LOUBERE, N., DUBURCQ, S. (1995) *Toward New Domestic Hot Water Comfort : from Modelisation to Advanced Control*, in Proceedings (to be published) of the International Gas Research Conference (Cannes, nov. FRANCE).

PRIESLEY, M.B. (1981), *Spectral analysis and time series*, Academic Press, San Francisco, USA.

RAMDANI, N. (1994), *Validation expérimentale et analyse de signaux: Développement d'une méthodologie de comparaison modèle/mesures en thermique du bâtiment*, Thèse Dr., Université Paris XII.