

Probabilistic short-term load forecasting models based on a parametric approach

L.A. Fernández-Jiménez, S. Terreros-Olarte, C. Capellán-Villacín, A. Falces, E. García-Garrido, P.M. Lara-Santillán, M. Mendoza-Villena, E. Zorzano-Alba and P.J. Zorzano-Santamaría

Department of Electrical Engineering E.T.S.I.I., La Rioja University

C/ San José de Calasanz, 3 – 26004 Logroño (Spain)

Phone/fax number: +0034 941 299 488, e-mail: luisalfredo.fernandez@unirioja.es

1. Introduction

Why is it interesting to predict the electric load?

Minimize production cost

for the balance supply and demand

for the Generation dispatch planning

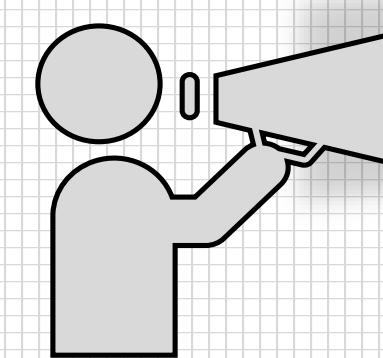
for Scheduled maintenance

for grid operation

for the management of power purchases and sales within electricity market

MAIN OBJECTIVE:

- to present a parametric probabilistic model designed for hourly load forecasting over the 24 hours of the subsequent day which is refined through a multi-objective algorithm optimization process that identifies explanatory variables from a specified set and also selects the type of distribution from among those characterized by two parameters.



Short-term load forecasting models

Deterministic models

- They provide the expected load value for the future instant.
- They omitting any information regarding its associated uncertainty.

Probabilistic models

- They furnish information for assessing the uncertainty associated with the forecasted value:
 - Quantiles or,
 - probability density functions (PDF)

PROCESS FOR THE OPTIMIZATION OF THE PARAMETRIC PROBABILISTIC METHOD:

multi-objective genetic algorithm (NSGA-II):

Explanatory variables: $(x_1, x_2, x_3, \dots, x_n)$
(weather variables, hourly load data over the previous days and dummy variables related to the type of day).

+

for a value provided by a deterministic forecasting model: generates a PDF, y

Models to create PDF

2. Selected parametric models

* GAMLSS (Generalized Additive Models for Location, Scale, and Shape) of two parameters:

μ , conditional mean of the response variable, and,
 σ , dispersion or scale of the distribution of the response variable.

* Distribution function of μ and σ parameters are modelled as linear functions of the selected explanatory variables:

$$y \sim D(\mu, \sigma)$$

$$g_1(\mu) = \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \dots + \beta_{m1}x_m$$

$$g_2(\sigma) = \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \dots + \beta_{n2}x_n$$

where:

y , dependent variable (hourly load),

D , family of distributions parametrized by μ and σ ,

β , linear coefficients,

x_i , explanatory variable i ,

m and n , number of explanatory variables selected for each parameter,

g_1 and g_2 , link functions.

D.- EVALUATION: Comparative analysis

Testing dataset	OPTIMIZED MODEL	Benchmark model REF1	Benchmark model REF2
CRPS (kW)	161.17	161.88	164.44
RI	0.06261	0.06409	0.1023
RMSE (kW)	299.47	307.78	313.59
MAE (kW)	226.35	229.21	231.46
MAPE (%)	6.3219	6.3399	6.4099
Explanatory variables	$51(\mu) - 30(\sigma)$	$55(\mu) - 55(\sigma)$	55

GAMLSS model without optimization of explanatory variables

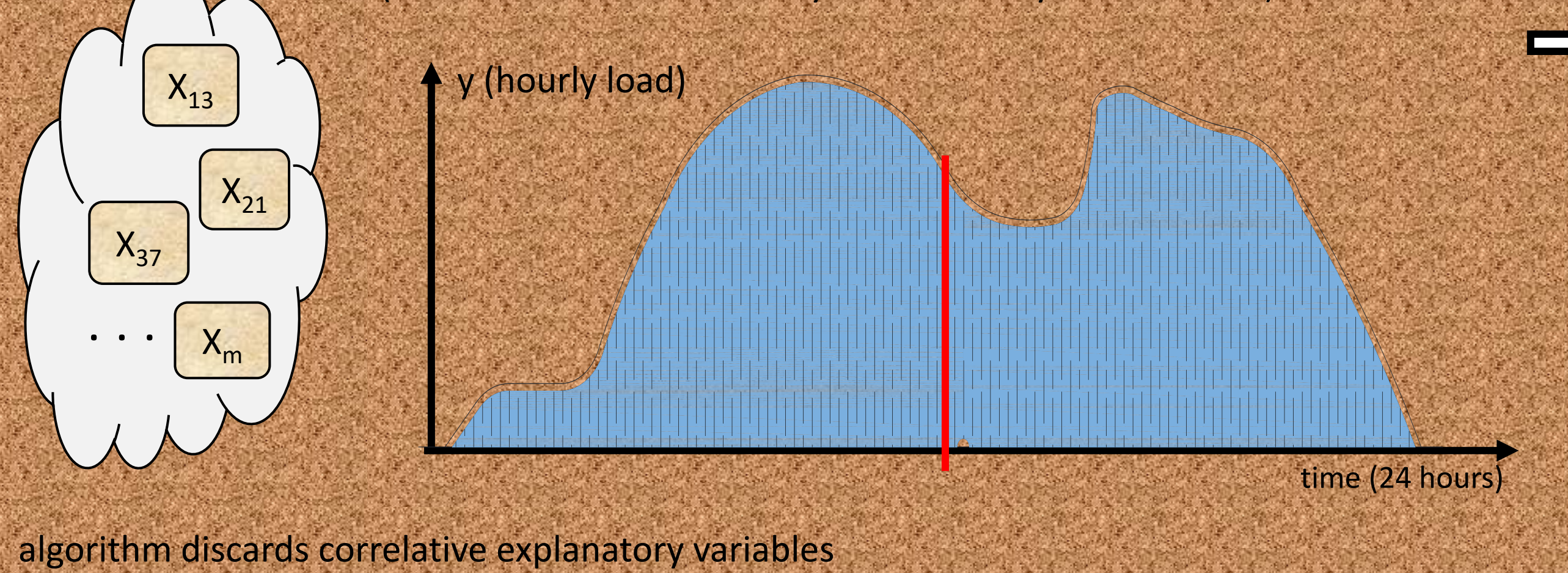
Utilizes a non-parametric approach using the quantile regression technique and incorporating all available explanatory variables

5. Conclusion

A) Our methodology optimizes the selection of explanatory variables for modelling the distribution parameters, enhancing the precision of GAMLSS type models.

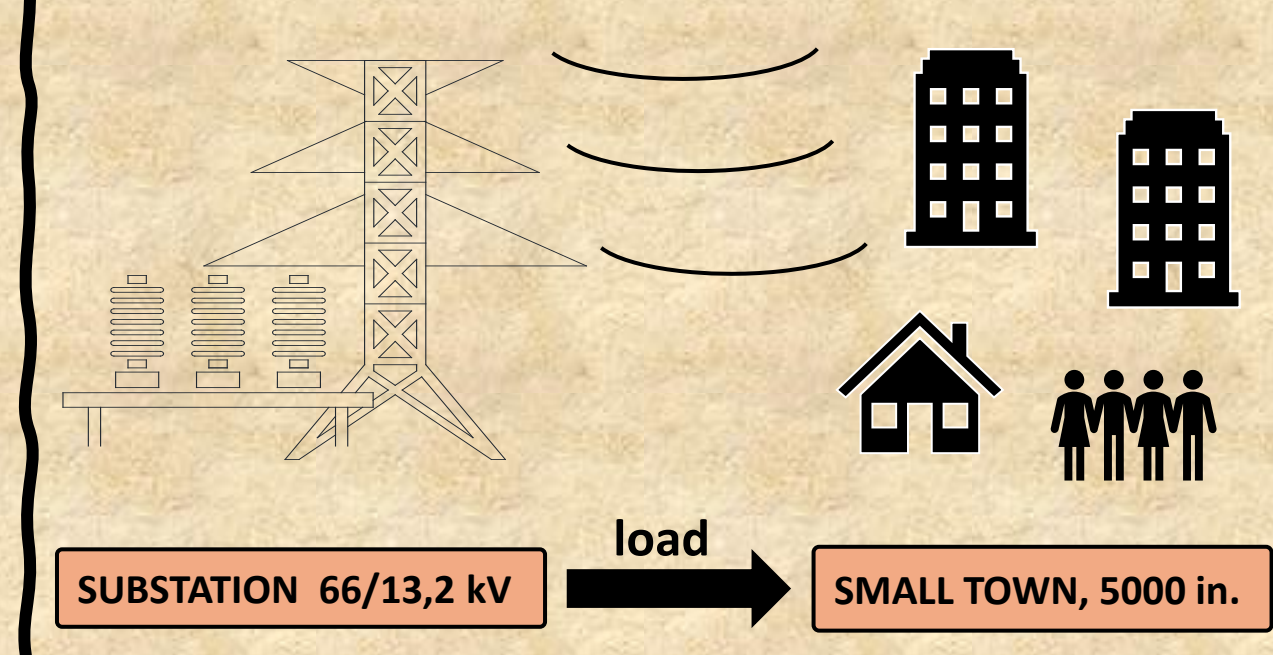
B) The proposed methodology is presently focused to the development of probabilistic short-term net load forecasting models who utilize photovoltaic generation systems installed behind the meter.

Optimized probabilistic model (balances between accuracy and reliability of the model)



4. Case Study

A.- OBJECTIVE: obtain a probabilistic forecasting model of the hourly load in a power substation 66/13,2 kV feeding a small town of about 5000 inhabitants (forecast horizon 24 hours)



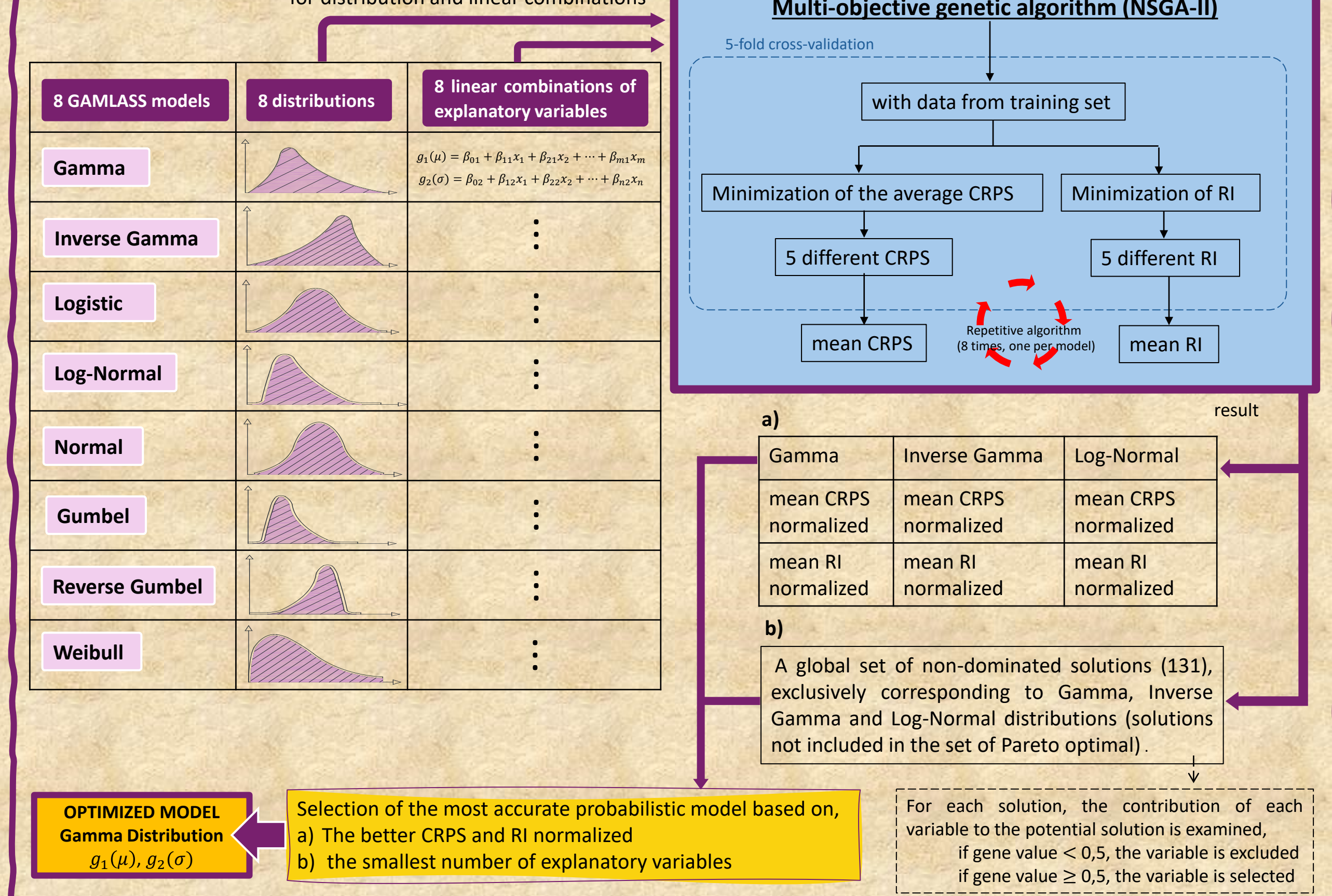
B.- DATASET: hourly load of 30 consecutive months partitioned into,

- Training dataset (initial 24 moths)
- Test dataset (subsequent 6 moths)

A set of 55 explanatory variables:
V1-V7: forecast for weather variables (MeteoGalia)
V8: European summer time (logical)
V9-V14: load at the same hour on the preceding 6 days
V15-V37: dummy variables for the hour of the day
V38-V43: dummy variables for day of the week
V44-V54: dummy variables for the month of the year
V55: dummy variable for holidays

load and explanatory variables	load	V1	V2	...	V55
DATE and hour (30 months)	M/D/Y : hour				
...	...	...	...	...	...

C.- PROCESS:



3. Evaluation Indexes

CRPS (Continuous Ranked Probability Score)

(quantifies the discrepancy between the cumulative distribution function (CF) of forecasted probabilities and the CF of observed value).

RESOLUTION of the model

$$* CRPS(CF, y) = \int_{\mathbb{R}} [CF(x) - \mathbb{I}(x \geq y)]^2 dx$$

(by integrating across the entire range of possible outcomes for the forecast variable).

$$* CRPS = \frac{1}{M} \sum_{i=1}^M |q_i - y| - \frac{1}{2M^2} \sum_{i,j=1}^M |q_i - q_j|$$

(case of forecast derived from non-parametric approaches that provide a set of quantiles).

PIT (Probability Integral Transform) histogram

(PIT value, varies from 0 to 1, is derived by computing $CF(x)$ at a specific verification point x and representing the quantiles of the distribution. Comparing PIT value with their corresponding verification points, the level of calibration is assessed).

RELIABILITY of the model

$$* RI = \sum_{k=1}^K \left| k_k - \frac{1}{K} \right|$$

(RI (Reliability Index) represented the frequency with which the actual value of the variable to be predicted falls within one of the intervals represented in the PIT histogram).

Others,

RMSE (Root Mean Square error)

$$RMSE = \sqrt{\frac{1}{N} \sum (\hat{P}(t) - P(t))^2}$$

MAE (Mean Absolute Error)

$$MAE = \frac{1}{N} \sum |\hat{P}(t) - P(t)|$$

MAPE (Mean Absolute Percentage Error)

$$MAPE = 100 \frac{1}{N} \sum \left| \frac{\hat{P}(t) - P(t)}{P(t)} \right|$$

where,

$\hat{P}(t)$, represents the load demand point forecast for hour t ,

$P(t)$, the actual value, and, N , the total number of hours in the evaluation period