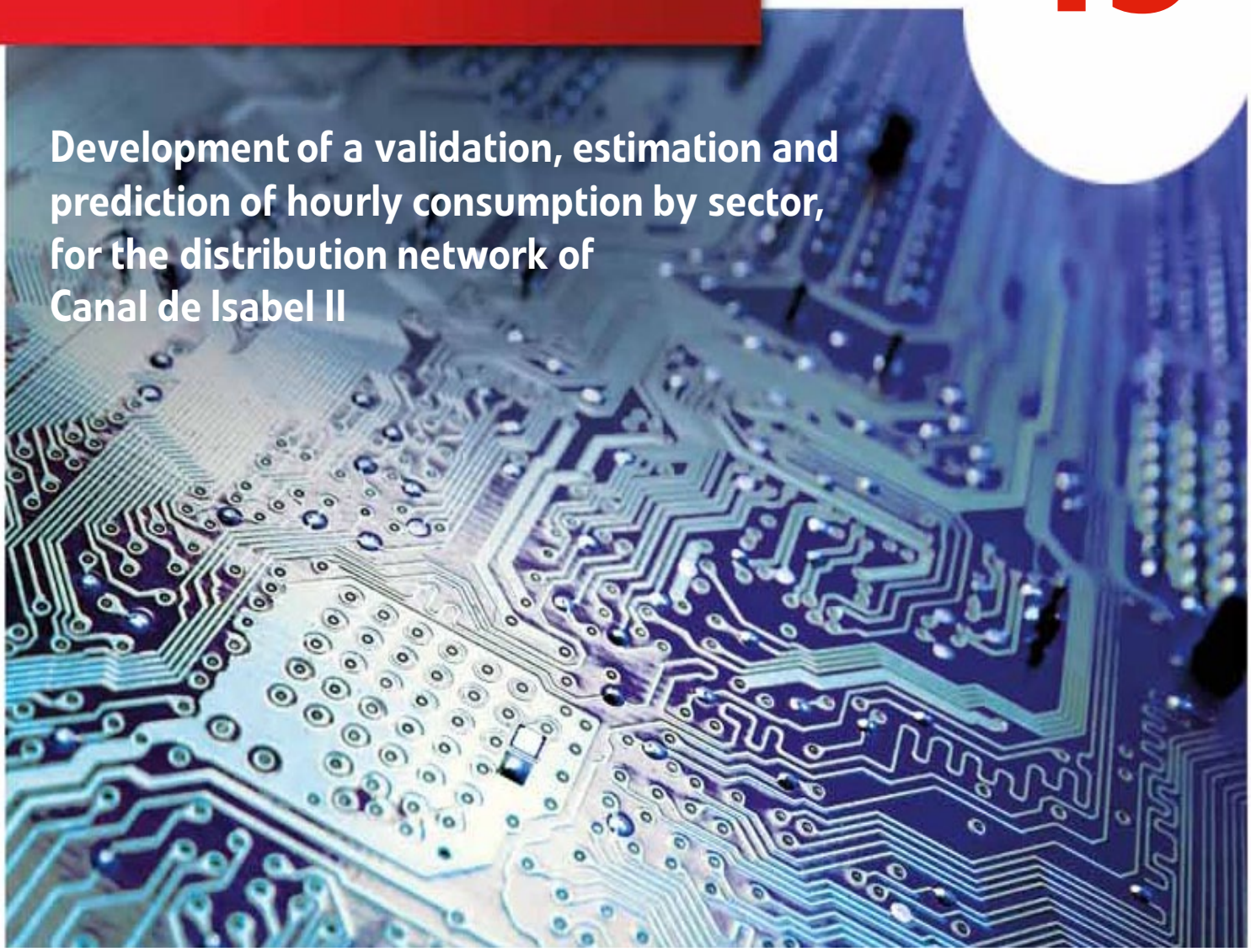
Canal de Isabel II

BOOKLETS

Research + Development & Innovation

13

Development of a validation, estimation and prediction of hourly consumption by sector, for the distribution network of Canal de Isabel II



Authors
Alejandro Carrasco Mínguez
Luis García Amor

Project Direction
Juan Carlos Ibáñez Carranza

Acknowledgements

We would like to express our thanks to Guillaume Raimbault and to Bruno Rodríguez Rodríguez for their collaboration and special contribution

ISSN: 2254-8955
Legal Deposit: M-39407-2011

Development of a validation, estimation and prediction of hourly consumption by sector, for the distribution network of Canal de Isabel II

Booklets
Research + Development & Innovation
Canal de Isabel II

13

Introduction

Canal de Isabel II's Research, Development & Innovation Booklets form part of the company's Knowledge Management Strategy and of the development involved in the Research, Development and Innovation Plan.

These Booklets represent an element for diffusion of projects and initiatives that are developed and sponsored by Canal de Isabel II for innovation in those areas related with water service in the urban environment.

A series of different problems that have been undertaken in each project are put forward in the Booklets, along with the results that have been obtained. The intention behind their diffusion by means of these publications is to share the experiences and knowledge that has been acquired with the entire water services sector, with the scientific community and with all those working on investigation and innovation tasks. What is aimed with the publication of these Booklets is to contribute to improvement and efficiency in water management and, consequently, in the quality of service that is provided to the citizens.

The R&D&I booklets published to date are as shown below by their titles in the following table.

Collection Number	Research, Development and Innovation Booklets published
1	Transferences of Water Rights between Urban and Agrarian Demands. The case of the Community of Madrid
2	Identification of Hydrometeorological Runs and Tendencies within the scope of the Canal de Isabel II system
3	Contribution of Canal de Isabel II to the International Demand Management Project (IDMF)
4	Micro-components and Explanatory Factors on Domestic Water Consumption in the Comunidad de Madrid
5	Virtual Water and Hydrological footprint in the Comunidad de Madrid
6	Study on the saving potential of water for residential uses in the Comunidad de Madrid
7	Potentials of efficiency in using dishwashers
8	Accuracy in the measurement of individual water consumption in the Madrid Region
9	Research project to define and assess the applicability of a Bioassay Test to determine the toxicity of water using Zebra Fish embryos
10	Water Use Efficiency in Gardening in the Region of Comunidad de Madrid
11	Remote sensing techniques and geographical information systems for assessing water demand for outdoor uses in the Comunidad de Madrid
12	Cyanotoxin Dynamics Study in two of the Canal de Isabel II's supply reservoirs in the autonomous region of Comunidad de Madrid

Project Outline

Project title	Development of a validation, estimation and prediction of hourly consumption at sectors scope, for the distribution network of Canal de Isabel II
Research line	Guaranteeing a balance between availability and demand
Canal de Isabel II units involved	Research, Development and Innovation Deputy Direction
External participation	Getinsa Ingeniería S.L.
Aim and justification of the project	Canal de Isabel II is currently implementing a sectorization of its distribution network. This measure enables the company to break down the hourly supply records into much greater detail. In this context the need arises to analyse these new series of data, paying special attention to certain behaviour variables that were hitherto considered either to be rather unimportant or had passed unnoticed. The aim is to obtain a series of identifiable patterns that to a large extent can be interpreted, on the basis of the population's habits and activities. At the same time, this scenario change brings with it the need to redesign the Canal de Isabel II's management and analysis tools to adapt them to the new approach.
State of the art contribution	Progress has been made in obtaining an awareness of new consumption patterns that it has been possible to identify, thanks to the distribution network sectorization undertaken. A greater refinement in the divisions has made it possible to group together consumption units, not only on the basis of their geographical location, but also in accordance with their typology. The results obtained from an analysis of the consumption data available can be considered original and novel.
Project development summary and milestones	The research has enabled Canal de Isabel II to develop a mathematical formulation and, from it, to develop an hourly consumption modelling tool that it has been applied to the wide range of demand unit types that Canal de Isabel II has been able to pinpoint in the Comunidad de Madrid supply system.
Obtained results summary	Development of water consumption estimation and prediction models, on an hourly basis, for more than 300 consumption units resulting from the sectorization of the distribution network. Updating the "CHYPRE" software application in order to adapt its functionalities to the management of the consumption data registered by the hydraulic control gauges installed in the different sectors.
Research lines open for continuing the work	Generating an awareness about consumption patterns on the basis of the typological characteristics of the network sectors. The development of the formulation has served to facilitate the process of acquiring knowledge. It is considered that the operating and maintenance stage for the consumption models developed could serve to obtain a greater in-depth working knowledge in this area.

Executive Summary

In 2004, Canal de Isabel II set in motion a plan to split its distribution network into independent sectors limited in extension. This sectorization basically consists of subdividing the network into small zones, isolated from each other, and connected to the main conveyance pipes through a limited number of inflow and outflow points. By equipping each one of these inflow and outflow points with a series of sensors able to provide exact information and at short intervals, associated with hydraulic control over the flows, one is in a situation to be able to diagnose and control the network operations on a small scale.

This project has enabled the Telecontrol system to record incoming flow data continuously, on an hourly basis, for each of the sectors implemented. As historical information from these records has been building up, the daily, weekly, and annual cycles to which most of the time variations in consumption respond have been emerging. In the majority of these cases, this behavior is clearly what is to be expected when one considers the activities, habits and movements of the population. However, some of the variability not accounted for by the aforementioned variables could be associated with meteorological factors, whose action could also be interpreted on the basis of comprehensible mechanisms.

Naturally, the representative pattern of the historical consumption recorded can vary considerably, on the basis of the demand typology or the predominant domestic unit in a given sector. However, such behavior can be understood in terms of there being different responses to the development of one single series of variables, these generally being easy to anticipate.

These precedents have been used in the research in an attempt to develop a mathematical formulation that can be applied to all the sectors implemented, one which can establish a relationship between the flow data obtained and a series of predictors values associated with certain variables on the basis of which it is possible to understand the fluctuations in the historical series of water consumption records.

As a second enterprise of this research, and on the basis of the aforementioned formulation, a methodology has been devised and implemented in a computer application that enables the user to adjust the parameters that define the consumption models for the allocated units.

These hourly-demand models constitute the basis for operating an hourly-consumption validation, estimation and prediction system, individualized for each consumption sector, in Canal de Isabel II supply system, and whose final results have proved to be tangible, and practical for the research work carried out.

To establish this system, the consumption models were developed to comply with a series of characteristics that could be adapted to the required functionalities. Along general lines, these can be summed up as follows:

- Time resolution: hourly
- Spatial resolution: each one of the consumption sectors into which the distribution network will be divided
- Prediction capacity
- Potential for updating

A summary is given below of the steps followed in the research undertaken that has enabled those involved to establish the consumption models and to apply them in practice to the process of flow estimation, validation and prediction.

Describing consumption behavior by analyzing historic data

Obtaining a prior knowledge of consumption behavior and, specifically, interpreting it in terms of the mechanisms or factors that cause it, was a basic activity. As part of this learning, the historical consumption series were analyzed, broken down into individual sectors. By way of a conclusion to this work, it was established that along general lines, the way the consumption patterns evolved could be interpreted as a response to the following factors:

- Calendar variables (effect of bank holidays, not only with respect to differences in urban activity, but also regarding the population movements associated with the days of rest).
- The seasonal nature of consumption (variations in the consumption level that occur between seasons of the year and others that can be attributed not only to holiday movements but also to the seasonal nature of certain water uses).
- The effect of meteorological oscillation, especially rainfall and temperature deviations from the expected values for a particular season
- Phenomena that tend to evolve slowly that can be attributed to long-term variations concerning the population supplied, water uses, consumption habits, etc.

Designing a suitable formulation for the models

The periodical nature of much of the temporal variability affecting water consumption and the clear relationship of dependency that has been found with the aforementioned variables, has made it possible to develop a linear mathematical formulation that, on the basis of a series of predictors associated with these, enables the user not only to suitably reproduce the consumption patterns recorded in the historical series, but also the to easily obtain the parameters involved in that formula.

In view of this, a mathematical model of water consumption has been prepared for each consumption sector. The ultimate aim of doing so is to be able to have available a system for estimating and validating hourly consumptions individualized for each sector, in such a way that the data recorded in the flow meters installed at the inflow and outflow points can be controlled.

Furthermore, these models are capable of making predictions, insofar as their constituent mathematical relations have been established in such a way that they simulate behavior observed in the series that respond to causes, with many prospects of continuing in time, in the medium or long term.

The methodology developed in this work followed the criterion that the inter-daily variability in consumption could be treated independently from the variability between days; this criterion led to the simultaneous development of models on two different time resolution levels:

- Hourly model, which considers the way in which daily consumption is distributed into the hours of the day. In this case, the magnitude used to establish the hourly models was the hourly consumption referred to the average hourly consumption of that day.
- Daily model, which simulates the way daily consumption evolves in time scales equivalent to, or greater than one day.

A summary is given below of the main characteristics of the formulation developed in relation to the explanatory variables that it contains:

- In most of the sectors analyzed the consumption pattern shows clear evidence of a weekly cycle, there being a notable difference in the magnitude and consumption pattern for bank holidays / weekends and for working days. In the modeling, this variability has been shown not only where a decrease in activity is concerned, but also with respect to population mobility.
- The seasonal nature of consumption has been obtained through a multiplicative factor in the modeling of daily consumption. The seasonal component in the evolution of the inter-daily pattern has also been taken into account, a predictor affected by this seasonality factor being incorporated into the hourly model.
- A trend-correcting multiplicative factor was also incorporated into this model, not only with a view to represent in the modeling any potential effects arising from long-term evolutionary phenomena, but also to make it possible to use the differences recorded between the observations and the estimations for the immediate past as an input variable; this factor functions adequately on a timescale of work ranging from one to two weeks.
- The effects of the meteorology on consumption are considerable, albeit comparatively less so than the effects of the calendar week. These effects have been incorporated into the modeling of the daily consumption through two related variables, one associated with the deviation between the temperature observed and the temperature expected, for the time of year, and the other associated with the percentile that the recorded rainfall occupies in the historical series. These predictors are, in turn affected by a multiplier factor that incorporates the seasonality attributable to the meteorological component, in view of the fact that it has been observed that, in the summer months, and transition between one season to another, the meteorological factor has greater repercussions on consumption than in the winter.

Adjusting and implementing the models

The modeling process basically consists of obtaining the constants for each one of the linear regression models included in the formula and of establishing seasonality curves that can represent the consumption variables observed among different seasons.

Before these tasks can be undertaken, it is necessary to select the consumption records observed that are to be adjusted. The individualized analysis of the historic series in each sector was used to select the data that could be considered as representative of the consumption pattern for the series. Once this set of reference had been selected, by applying the minimum squares criterion, the modeling constants were established through which the greatest degree of adjustment would be achieved between these consumptions observed and the estimations obtained. As is evident, the goodness of fitting depends, to a large extent, on the quality of the series of consumption data available at the time the sector modeling is done.

The errors in the estimation can be attributed not only to the limitations inherent to the formulation of the model that render it impossible to respond to certain consumption variability components, but also to the effects of phenomena that can be regarded as random, or to errors committed in the records taken. However, in view of the good result obtained in the modeling carried out, it can be concluded that these deviations are not significant, so they can be considered as noise.

These differences between the flows measured and the values provided by the model served to establish the validation parameters, a reasonable limit for being able to diagnose whether the inflowing consumption is sufficiently out of keeping with the expected value for it, to be regarded as non-representative of the real situation that it is intended to measure.

The contribution made by the research in this aspect of the work, included estimating a logical way of calculating the criteria and validation parameters on the basis of which the limit is established above or below which rejection of the inflowing consumption takes place.

The result obtained from the validation carried out on the reference set was established as the basis for comparison when assessing the accuracy of each model. The coefficient of regression (r^2) was selected as being representative of the degree of adjustment and the average quadratic error was selected as the measurement of the error in the model. The degree of adjustment for the models is generally satisfactory. When reference sets are long enough and of sufficient quality, the coefficient of regression takes values close to 0.9, whereas the quadratic error remains around 10% of the average daily consumption measured; these results clearly endorse the validity of the formulation used.

Outside the modeling phase, evaluating a model's adjustment level while it is operating is generally complicated and not always representative, because the errors obtained are determined to a great extent by the way the input data evolve. As such data move away from the pattern observed in the reference set, they are liable to be rejected, even when they do not contain any significant anomalies.

Using the models developed in the processes of validating, estimating and predicting consumption

The first result of the work done, made available a series of autoregressive models based on a linear formulation, not only on a daily basis but also on each one of the hourly consumption components.

As has already been pointed out, these models have been developed with a view to serving as the basement for the processes of validating, estimating and predicting consumption data by sectors. The estimation enables one to provide a consumption value regardless of the data measured, using the modeling tool. As has already been stated, this tool is based upon a set of seasonal, calendar and meteorological predictors, and upon the consumptions of this sector recorded (or estimated) at earlier dates.

The validation is used at the model operating phase to diagnose the validity of the consumption data obtained from the Telecontrol systems installed. A record is understood to be valid if it responds, with reasonable precision, to the reality that it is supposed to measure. The errors affecting the data-collection system or the anomalies in the real situation measured yield non-valid data, i.e., unrepresentative data, and so they must be rejected from the historical series. The rejection of data creates a validation gap that is filled in with data estimated by the model for that same day and hour of the day.

As they are relatively simple to formulate, self-restraint where dependence on previous consumption data are concerned, and have a limited number of parameters, the models are generally stable (they do not yield values out of range or sharp fluctuations in the simulated data). In view of these characteristics, they can be used for prediction purposes.

Implementing these individualized consumption models in the more than 300 sectors analyzed until this document was prepared, has served to endorse not only the formulation obtained from the research work conducted, but also the processes used to calculate the modeling parameters by analyzing historical series.

As is well known, practically the entire population of the Comunidad de Madrid is supplied by Canal de Isabel II. As a result, the consumption units supplied vary greatly, which directly accounts for the fact that a wide variety of sector types have had to be analyzed. Although domestic consumption units predominate, there are also sectors located in zones of an industrial nature. There are also many types of consumption units, all of which have their own special characteristics, such as golf courses, camp sites or leisure centers.

At another level of disaggregation, even within the consumptions associated with residential areas, numerous categories can be distinguished on the basis of the representativeness of the sector with respect to the variables that account for consumption: type of dwelling (single family or multifamily), use to which it is aimed (first or second home), income level of the resident population, etc.

As has already been pointed out, the variables on the basis of which the relationship of dependence with the evolution of consumption has been established, are the same in all the sectors analyzed (trend, seasonality, calendar and meteorology). The versatility of the formulation adopted is based on the fact that, regardless of what the characteristics of the sector are, a good correlation has been found between the way the observed consumption has evolved and the variation in these explanatory variables (although the response that is obtained can clearly vary to a great extent, depending on the sector type concerned).

By way of a final conclusion, it should be mentioned that developing theoretical consumption models for each one of the sector units, has not only served to increase awareness about the consumption patterns associated with the different types of demand that have been pinpointed, but has also enabled those concerned to obtain a greater in-depth knowledge about the behavior, on a small scale, of the distribution network. The formula for the model developed is able to hold the knowledge generated about consumption patterns and the interpretations that make it possible to explain them.

Furthermore, thanks to the implementation of the models, the contrast between the consumption estimations prepared by the theoretical models and the actual flows recorded, serve to improve the way distribution network operations are controlled, enabling those involved to detect variations in the pattern of consumption, as well as incidents of any kind (leaks, bursts, works, anomalies in the way the instrumentation is working, etc.).

http://www.canalgestion.es/galeria_ficheros/compromiso-social/publicaciones/Cuaderno13_I+D+i.pdf

Canal de Isabel II

www.cyii. es

Canal de Isabel II

Santa Engracia, 125

28003 Madrid (Spain)