environmental and earth sciences proceedings

MDPI

Proceeding Paper

Short-Term Probabilistic Forecasting of Water Demand Using GPR: A Case Study in Southern Italy †

Cristian Cappello *, Carla Tricarico, Giovanni de Marinis and Angelo Leopardi

Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio, EUT+—European University of Technology, Via Di Biasio 43, 03043 Cassino, Italy; c.tricarico@unicas.it (C.T.); demarinis@unicas.it (G.d.M.); a.leopardi@unicas.it (A.L.)

* Correspondence: cristian.cappello@unicas.it

† Presented at the 6th International Conference on Efficient Water Systems (EWaS6), Thessaloniki, Greece, 11–14 May 2026.

Abstract

Short-term water demand forecasting is a key issue for the management of smart water networks, particularly in the context of remote control and active regulation. This study analyses a real-world dataset of water demand coefficients, collected at 15 min intervals, from a municipality in Southern Italy serving approximately 73,000 inhabitants. The proposed model, based on Gaussian Process Regression (GPR) with a Rational Quadratic kernel (RQ), is compared with a statistical benchmark constructed using average patterns for each time slot by the application of the Gauss Distribution. The results show a reduction in RMSE and MAE and a better ability to track the daily dynamics of demand using the GPR approach.

Keywords: Water Distribution Network; water demand forecasting; machine learning; gaussian process regression

1. Introduction

Water Distribution Networks (WDNs) management is nowadays moving towards smart water systems in which the remote control of the main components, such as pressure regulations, pump operations, tank levels, etc., are necessary in order to guarantee a fast and reliable supply service to the users. All these settings regulation variations, however, are following the time evolution of the network which varies according to the water demand required by users. The latter is the main cause for system operation and settings variation, and its forecast is a challenge of great interest in research and practical application.

In particular, the availability of reliable short-term forecasting, such as 15–60 min, is of particular interest for supporting system operational decisions.

Water demand as a random variable is difficult to characterise and its study should require a statistical approach which will take into consideration the random nature of the water's requirement and its variation with time.

The recent literature highlights a significant development in data-driven approaches to water demand forecasting. The reviews by [1,2] show how, alongside classical statistical models, machine learning and deep learning methods—including neural networks, support vector machines, random forests and hybrid models—are increasingly being used, although no single approach has emerged as universally superior for all application contexts. Stacking regression models, LSTMs and other advanced architectures demonstrate the strong performance for short-term demand forecasting [3,4]. However, all these methods, even if

Academic Editors: Vasilis Kanakoudis, Evangelos Keramaris, Francesco De Paola, Habib Muhammetoglu and Theodora Papamitsou

Published: 22 June 2026

well trained and tested on numerous data, are not taking into consideration the probabilistic nature of water consumption.

The effectiveness of forecasting urban demand using climatic and socio-economic variables by means of the Gaussian Process Regression (GPR), based on a probabilistic forecasting, has also been demonstrated [5].

Furthermore, most of the residential water demand's short-term forecasting focus on hourly time scales. However, a coarse time interval, even if of interest to the WDS operations which usually could require hourly operations, are not characterising in deep detail the real water requirement, underestimating for instance the peak condition, which is one of the most onerous operative conditions. On the other hand, a 1 s or 1 min time scale could result in time intervals too fine for remote control operation as Water Utilities are not able to manage such a big amount of data. Under this assumption, a good compromise could be 15 min time intervals, in which a sufficiently detailed study of the water demand is allowed and at the same time it is of interest for operation applications.

This study therefore proposes the use of Gaussian Process Regression for the probabilistic forecasting of short-term water demand, applied to a time series of data with a 15 min sampling interval, comparing the model's performance with a statistical benchmark based on the application of the Gauss probabilistic distribution to the same sample data. This methodology was applied to a medium-sized municipality located in Southern Italy, with a population of approximately 73,000 users. The period analysed is the winter months and in particular January-April, in which weekdays and weekends have been considered separately. The data have been further normalised to the averaged daily consumption for the period under analysis, in order to obtain the water demand coefficient, *Cd*. The aim of the present study is therefore to verify whether GPR can offer an advantage not only in terms of point accuracy, but also in terms of probabilistic information useful for future applications of dynamic network management.

2. Methodology

2.1. Problem Setup

In order to apply the forecasting model, the analysis was set up by considering the demand coefficient for the quarter-hour following time t (y_{t+1}) as the target variable, whilst information available up to the previous time step (x_{t-1}) with an uncertainty (ε_t) was used as explanatory variables [Equation (1)]. This formulation allows the problem to be framed as a 15 min one-step-ahead forecast, a configuration particularly suited to potential near-real-time applications in the field of remote control and operational management of water networks.

$$y_{t+1} = f(x_{t-1}) + \varepsilon_t \quad (1)$$

Since water demand patterns vary significantly depending on the type of day, calendar data distinguishing between weekdays and weekends were also taken into account. This is important because residential water consumption patterns generally follow different trends in line with daily activity patterns.

Because of the 15 min time interval, the training was done in three months, January–March, while the testing is validated on the April data. This helped to limit the computational burden of training the GPR, of which the application of very large datasets can be computationally intensive due to the management of the covariance matrix, whose computational cost increases rapidly as the number of observations rises, in the order of O(n^3) [6,7].

2.2. Feature Engineering

For each time step t, the feature vector was constructed by combining autoregressive variables and calendar-based temporal information. In particular, the demand coefficients at times $t-1$, $t-4$, $t-96$ and $t-672$, were considered, corresponding respectively to the previous quarter of an hour, one hour earlier, one day earlier and one week earlier. This choice allows both the short-term memory of the signal and its main daily and weekly periodicities to be represented.

The time variables, associated with the time of the day and the day of the week, were also encoded using trigonometric functions, in order to preserve their cyclical nature and avoid artificial numerical discontinuities. In this way, the model can jointly exploit recent demand dynamics and the regularity of recurring temporal patterns.

2.3. Forecasting Model and Validation Strategy

GPR was selected as the forecasting model due to its ability to approximate non-linear relationships in a flexible manner, without imposing a predefined functional form between input and output. Furthermore, as it is a non-parametric probabilistic approach, unlike other machine learning algorithms, GPR allows an explicit quantification of predictive uncertainty to be associated with the forecast point.

For each input, the model returns an a posteriori Gaussian distribution, described by a mean and a predictive variance. The former is taken as an estimate of the predicted value, whilst the latter provides useful information regarding the local reliability of the forecast.

Among the kernels examined in the preliminary tests, the Rational Quadratic kernel proved to be the most effective in terms of predictive error. Its adoption was consistent with the structure of the problem, characterised by the presence of dynamics acting on different time scales.

The validation procedure was set up using an 80/20 time split, applied separately to each sub-period of the dataset. In this way, the model was trained on the initial portion of each time block and subsequently tested on the final portion, simulating a forecasting setup consistent with the model's operational use.

The decision to allocate 80% of the data to training was made to ensure a sufficient number of observations for learning, whilst still retaining a portion for validation.

2.4. Benchmark and Performance Assessment

A traditional statistical approach, based on the application of the Gauss probabilistic approach with mean equal to the average value of the demand coefficient observed in the historical data for each time slot and relative standard deviations, was adopted as the benchmark. In order to compare the relative results with the GPR methods previously applied, the statistical benchmark was further distinguished between weekdays and weekends.

The choice of this benchmark allows the machine learning model to be compared with a simple, interpretable baseline that was consistent with the methods traditionally used to describe demand patterns. The comparison with GPR was therefore carried out in terms of pointwise forecast accuracy on the test datasets.

To this end, two main metrics were used: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), calculated by comparing observed values with predicted values. The RMSE is more sensitive to larger errors, whilst the MAE provides a direct measure of the mean absolute error.

In addition to assessing point accuracy, the probabilistic component of the GPR model was also analysed through the coverage of the predictive intervals. This indicator measures the percentage of observations in the test dataset that fall within the confidence interval estimated by the model, thereby allowing us to verify the consistency between the predictive uncertainty returned by the GPR and the errors actually observed. Coverage can be expressed as [Equation (2)]:

$$Coverage = \frac{1}{n}\sum_{i=1}^{n} I(y_i \in [L_i, U_i]) \quad (2)$$

where $I(\cdot)$ is the indicator function, equal to 1 if the observed value y_i lies within the predictive interval $[L_i,\ U_i]$, and 0 otherwise. Here, $[L_i,\ U_i]$ denotes the predictive interval associated with the i-th forecast.

3. Results and Discussion

Table 1 shows the forecasting performance of the benchmark and the GPR model, which returns lower RMSE and MAE values than the benchmark for both weekdays and weekends, thus demonstrating a superior ability to forecast short-term water demand. This result confirms that the proposed model outperforms the traditional approach based on average coefficients per time slot. On weekdays, the RMSE value drops from 0.0652 to 0.0268, whilst the MAE falls from 0.0542 to 0.0183. For weekends, the RMSE decreases from 0.0674 to 0.0221 and the MAE from 0.0531 to 0.0163, with a slightly greater improvement observed during weekends, probably due to the numerous sample data of the weekdays in the three months considered.

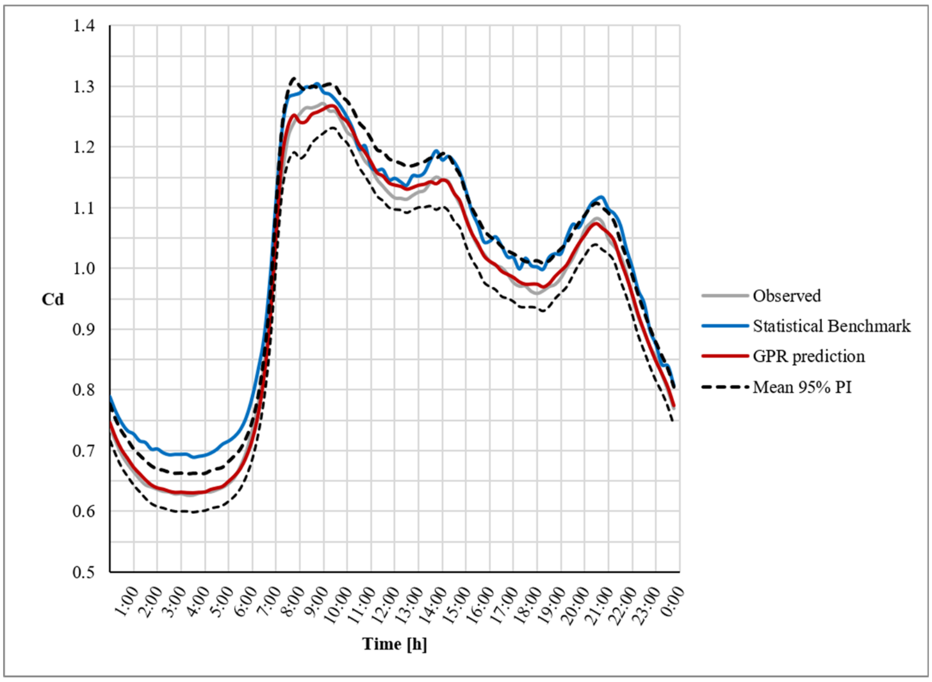
Table 1. Comparison of benchmark and GPR performance across time periods.

Day	RMSE Benchmark	RMSE GPR	MAE Benchmark	MAE GPR	Coverage [%]
weekdays	0.0652	0.0268	0.0542	0.0183	89
weekends	0.0674	0.0221	0.0531	0.0163	90

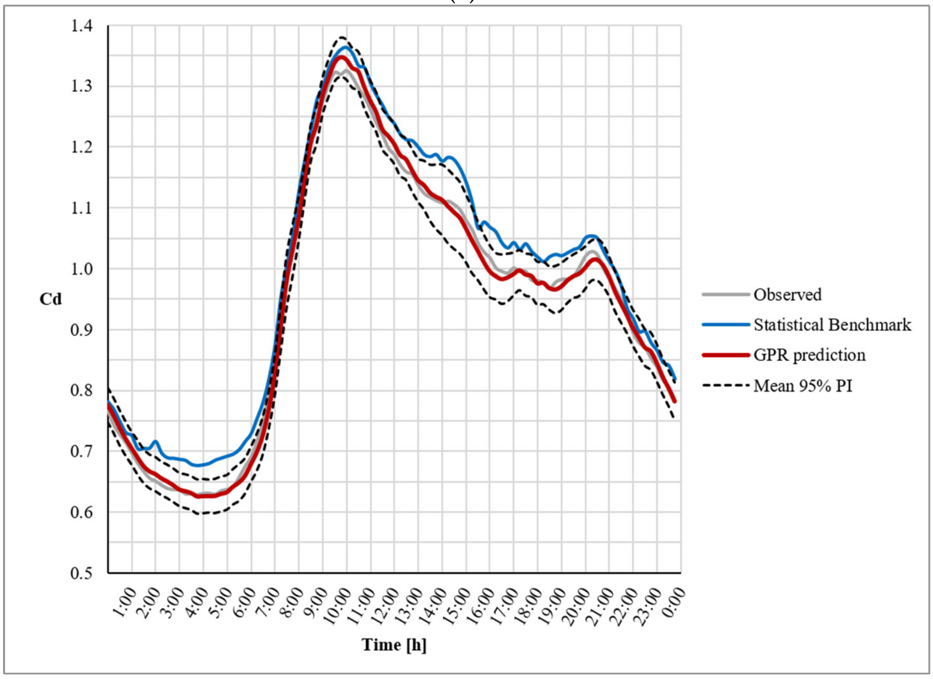
By comparing the GPR-forecasted pattern and the generated statistical benchmarks, for both weekdays (Figure 1a) and weekends (Figure 1b), it is shown that the benchmark reproduces the general shape of the demand profile in a broadly acceptable manner, but exhibits an almost systematic tendency to overestimate the observed demand coefficients, both during night-time hours and at the morning peak and the subsequent evening rise. In this context, the GPR model has a closer fit to the observed profile, largely correcting the benchmark's positive bias and tracking the phases of demand growth and decline with greater consistency, confirming the findings already highlighted by the quantitative metrics. This result could also be explained by the stochastic nature of the water demand during peaks, which is generally not well described by a symmetric distribution like the Gauss model. This latter problem does not affect the forecasting done by the application of the GPR approach.

In addition, a closer inspection of the profiles highlights that the discrepancy between the benchmark and the observed data is not limited to a simple vertical shift, but also affects the shape of the curve, particularly in the transition phases between the morning peak and the mid-day plateau. This effect is more evident during weekends, where the demand profile appears sharper and more irregular, with a more pronounced peak and a steeper decline. In these conditions, the benchmark, being based on average patterns, tends to smooth and misrepresent these variations, while the GPR model is able to better capture both the amplitude and the timing of the demand evolution.

Furthermore, the GPR prediction shows a slightly smoother behaviour compared to the observed data, especially around local fluctuations, suggesting a regularisation effect inherent to the model. However, this smoothing does not significantly affect the overall reconstruction of the demand dynamics, and instead contributes to a more stable and robust representation of the underlying pattern.



(**a**)



(**b**)

Figure 1. Observed and predicted demand patterns for weekdays (**a**) and weekends (**b**).

The probabilistic component of the model was assessed using the 95% empirical coverage of the nominal prediction interval, which was found to be 89% for weekdays and 90% for weekends. These values indicate that the prediction intervals returned by the GPR model tend to be slightly narrower than required by the actual variability of the observed data. In other words, whilst providing accurate point forecasts, the model shows a moderate underestimation of the uncertainty associated with the forecast. The slight difference between weekdays and weekends also suggests that, during the period analysed, the model's probabilistic calibration remains broadly comparable across the two types of days, despite differing demand profiles.

4. Conclusions

In this study, a Gaussian Process Regression model was proposed for short-term water demand forecasting and its predictive performance was validated against a statistical benchmark derived from a Gaussian distribution applied to the identical dataset. This methodology was applied to a medium-sized municipality located in Southern Italy, with a population of approximately 73,000 users. The period analysed with a time step of 15 min is referring to the winter months and in particular January-April, in which weekdays and weekends have been considered separately.

From an application perspective, the results obtained show that GPR can be a promising tool for forecasting water demand in the very short term within the context of smart water networks. The improvement over the classic statistical benchmark suggests, in fact, a potential use in supporting future strategies for dynamic network management and control.

At the same time, the coverage analysis showed that the probabilistic component of the model is generally consistent with the data structure, although it slightly underestimates the actual variability compared with the nominal 95% level. This does not diminish the value of the results obtained in terms of accuracy point, but rather suggests that the quantification of uncertainty could be further refined. Overall, the model therefore proves promising not only for forecasting demand in the very short term, but also as a basis for future developments aimed at an even more robust probabilistic representation.

A deep characterisation and prediction of the water demand at the probabilistic level is therefore of great interest for future applications of dynamic network management.

Author Contributions: Conceptualization, C.T.; methodology, C.C. and C.T.; software, C.C.; validation, C.C. and C.T.; formal analysis, C.C. and C.T.; investigation, C.C. and C.T.; resources, C.C. and C.T.; data curation, C.C. and C.T.; writing—original draft preparation, C.C. and C.T.; writing—review and editing, C.C. and C.T.; visualisation, C.C., C.T., G.d.M. and A.L.; All authors have read and agreed to the published version of the manuscript.

Funding: Part of the research presented in the paper was funded under the National Grant PRIN2022 n.2022JN9YNJ "SMART REhabilitation of NEtworks with high Water losses (SMART RENEW)". Cristian Cappello was fully funded in the framework of the PhD program in "Metodi, modelli e tecnologie per l'ingegneria" at the University of Cassino and Southern Lazio.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Niknam, A.; Zare, H.K.; Hosseininasab, H.; Mostafaeipour, A.; Herrera, M. A Critical Review of Short-Term Water Demand Forecasting Tools—What Method Should I Use? *Sustainability* **2022**, *14*, 5412. [CrossRef]
2. Ghannam, S.; Hussain, F. Short-Term Water Demand Forecasting: A Review. *Australas. J. Water Resour.* **2024**, 1–19. [CrossRef]
3. Mohamed Hussain, K.; Sivakumaran, N.; Radhakrishnan, T.K.; Swaminathan, G.; Sankaranarayanan, S. Short-Term Water Demand Prediction Using Stacking Regression–Based Machine Learning. *Water Pract. Technol.* **2024**, *19*, 4773–4796. [CrossRef]
4. Kavya, M.; Mathew, A.; Shekar, P.R.; P, S. Short Term Water Demand Forecast Modelling Using Artificial Intelligence for Smart Water Management. *Sustain. Cities Soc.* **2023**, *95*, 104610. [CrossRef]
5. Roushangar, K.; Alizadeh, F. Investigating Effect of Socio-Economic and Climatic Variables in Urban Water Consumption Prediction via Gaussian Process Regression Approach. *Water Supply* **2018**, *18*, 84–93. [CrossRef]

6. Li, J.; Wang, H. Gaussian Process Regression for Uncertainty Quantification: An Introductory Tutorial. *arXiv* **2025**, arXiv:2502.03090.
7. Sun, N.; Zhang, S.; Peng, T.; Zhang, N.; Zhou, J.; Zhang, H. Multi-Variables-Driven Model Based on Random Forest and Gaussian Process Regression for Monthly Streamflow Forecasting. *Water* **2022**, *14*, 1828. [CrossRef]

https://doi.org/10.3390/eesp2026044012