On the autocorrelation of measurement results for gas volume and calorific value in fiscal metering in gas grids

Federica Gugole[1*], *Meng* Li[2], and *Adriaan M. H.* van der Veen[1]

[1]VSL, Data Science & Modelling Department, Thijsseweg 11, 2629 JA Delft, The Netherlands
[2]TU/e, Eindhoven University of Technology, Department of Mathematics and Computer Science, Eindhoven, The Netherlands

Abstract. The fiscal metering of natural gas is often performed using a flow meter and a gas chromatograph. The flow meter measures the volume flow rate of the gas and the gas chromatograph measures the composition, from which the calorific value is calculated. The energy is then calculated as the product of the volume and the calorific value. The uncertainty of the energy delivered (or received) is an important parameter for operating the gas grid. Currently, the energy values used to compute the total energy are considered mutually independent. However, the underlying processes affecting the gas flow through a pipe are continuous. We analysed the correlation between successive measurements due to temporal variations in the physical process by applying autoregressive moving average (ARMA) models to a set of measurement data taken at a metering station in the Dutch gas grid. The analysis showed that the volume time series can be modelled as an autoregressive model of order 1, while the time series of the calorific value can be described by an autoregressive model of order 2. These correlations lead to an increase of the instrumental measurement uncertainty associated to the total energy of about 50%.

1 Introduction

As part of the energy transition, required to achieve the net-zero greenhouse gas emissions targets for 2050, European gas grids are facing a diversification of the gas entering the grids. Renewable gases such as hydrogen, biogas and biomethane will replace fossil natural gas and the question arises whether the current infrastructure is still adequate. This infrastructure includes the instruments used at a metering station to determine the quantity (i.e., mass or volume) and the quality (i.e., the calorific value) of the gas, including data processing equipment used for calculating the total quantity or energy over a time interval. This equipment uses current modelling practices to calculate the uncertainty of the energy delivered (or received). Usually, the quantity of gas is measured by a flow meter, while a gas chromatograph determines the composition of the gas from which, e.g., the calorific value is calculated. Such measurements are important for financial transactions between buyer and seller, while the combined uncertainty of the measurements is an important parameter to ensure that suppliers can feed the grid and users can draw gas from the grid that is within legislative and contractual specifications.

Fiscal metering of energy gases is described in several standards and guidance documents, such as OIML-R140 [1], EN 1776 [2] and ISO 15112 [3]. The requirements for the measurement equipment to be calibrated and the measurement results to be metrologically traceable are common to all these documents, and so are the models described. The delivered energy is generally calculated as the sum of the energy increments over time, and in the evaluation of the measurement uncertainty, it is usually assumed that the quantities of interest (e.g., volume and calorific value at normal conditions) are mutually independent, and normally distributed. Whereas the latter assumption is usually unproblematic, it was shown that the former is unrealistic, see Van der Veen et al., 2025 [4].

Two sources of correlations between measurement results have been identified. The first cause of correlation is due to the use of the same instrumentation for all measurements, and the second is due to the continuous nature of the underlying physical processes. Whereas the latter effects can be small under steady-state conditions, their contribution can become substantial with a greater diversity of energy gases in gas grids and with larger fluctuations in supply and demand. In an era, where gas grids are facing a further diversification of the gas entering the grid and the fluctuations also increase, the uncertainty due to temporal effects is expected to become more relevant.

In this work, we show how time series models can be used to describe non-random features in gas grid data and quantify the correlation between subsequent measurements due to temporal effects. We then estimate the uncertainty both according to the current modelling practices and including the estimated temporal correlations. In section 2, we outline the methodology used for the analysis. We then present the results (section 3) and our conclusions.

* Corresponding author: fgugole@vsl.nl

2 Analysis

2.1 Time series

Data collected (and indexed) over time is called a time series. Just as with other series of data, there are different models that can be employed to describe and summarise the data. A basic time series model is the white Gaussian noise model, i.e., the data is assumed to be independent, identically and normally distributed. Many time series are however not centred around a specific mean, show trends, oscillating behaviour or other features. In gas grids, it is expected that measurement data will also contain such features due to, e.g., changes in supply and demand, blending with different gases or gas qualities or other events. Such phenomena give rise to the situation that the last measurement result depends on the previous one (and so on). Physically, one would expect that the current state at a metering point does not differ (too) much from a previous state, nor will it differ substantially from the next state. It takes time for the gas to respond to phenomena in the grid, and that will be reflected in the data.

If the current state depends on a previous state, then such dependencies affect the uncertainty about quantities evaluated using time series data. This kind of correlation is known as autocorrelation and it is different from, e.g., the correlations arising from the fact that the same instrumentation was used for the entire time series. Describing gas metering data using time series models aids at exploring the features in the time series and evaluating, among others, the autocorrelation between the measurement results in the time series.

An important concept in time series analysis is stationarity, which means that a time series has a constant mean and a covariance function that depends only on the time distance between the data points and not on the specific time at which the data have been collected [5, 6]. Using time series models, the mean and its associated uncertainty can be evaluated. Ignoring the specific features in the time series can lead to a wrong value for the uncertainty.

2.2 Data set

In this study, we analysed data that were collected at a metering station in the Dutch gas grid. The measurements were made with an ultrasonic flow meter, a gas chromatograph and temperature and pressure transmitters. The dataset covers a time window of two days with a time interval of 15 minutes between recorded measurement points.

The dataset consisted of the volume at normal conditions and the volumetric calorific value, also at normal conditions. Corrections for the actual pressure, temperature and calibration of the equipment had already been applied. The reported values for the volume represent the accumulated volume over the 15 minutes interval, while those reported for the calorific value are the average over the interval. The data are displayed in Figure 1.

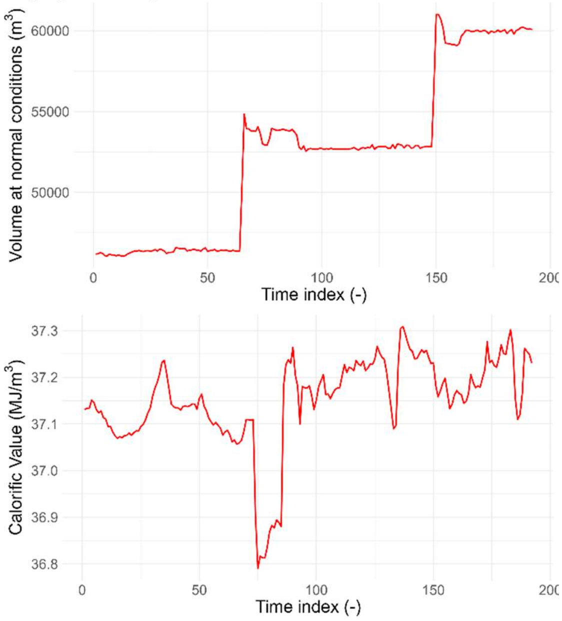

Fig. 1. The available measurement data at normal conditions (top: volume; bottom: calorific value) used in the analysis.

2.3 Time series model

We applied so-called autoregressive moving average (ARMA) models, a widely used class of time series models. ARMA models express the value observed at a particular epoch as a linear combination of a finite number of past values (AR part) plus an error defined as a linear combination of measurement errors affecting the current observation and a finite number of previous observations (MA part). The AR part describes the dependence of one datum with respect to its predecessors. The MA part instead, expresses the dependence of one datum with respect to the current and past instances of a non-identical to itself variable. The autocorrelation function (ACF) measures the correlation between observations at different distances apart. It is well suited to estimate the order of MA processes, since the ACF will be zero for distances larger than the order of the MA process. However, in case of AR processes the ACF alone gives little information. In that case it is helpful to look at the partial ACF (PACF). The idea of the PACF is to compute the correlation between two variables with the linear effect on a third variable removed. It is thus helpful to determine the order of AR processes. We invite the interested reader to check, e.g., Shumway and Stoffer, 2017 [5] for a more detailed explanation.

2.4 Automatic segmentation of the data

ARMA models can be used to describe series of observations provided that the series is (statistically) stationary. Many time series are not stationary. To get around the non-stationarity of the entire dataset, it is possible to look for stationary parts in the time series and locate those. To this end, we applied the binary segmentation method (Bai, 1997 [7]) to automatically detect stationary subsets of the available time series. The binary segmentation method is an approximate change point method that uses the mean of the series as criteria to determine the segmentation points. It is therefore well indicated to determine large changes in the time series, but it might not detect more subtle changes. As it can be seen in Figure 2, the binary segmentation method does a good job at detecting the biggest jumps first. It can then be applied recursively until the subseries satisfy the stationarity criteria, e.g., as tested by the augmented Dickey-Fuller test. This test tests the null-hypothesis that a unit root is present in the time series. If that is the case, the time series is non-stationary.

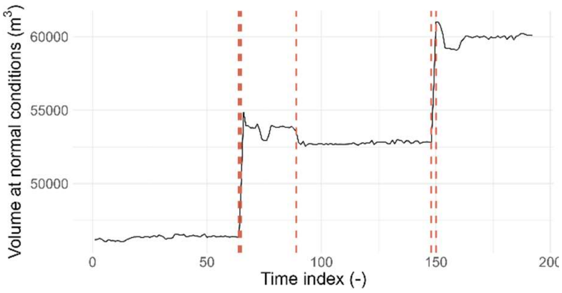

Fig. 2. Example of segmentation results using binary segmentation.

As a result of this analysis, we obtained three stationary subseries for the volume measurement data and two for the calorific value.

2.5 Fitting a time series model

Once stationary subseries of the volume and of the calorific value were identified, we computed the ACF and PACF to determine the order of the ARMA models to be used. The analysis revealed that the volume data are better described by an autoregressive (AR) process of order 1, meaning that the measurement result at time *t* is significantly correlated with the measurement result at time *t*-1. The calorific value time series turned out to be better modelled by an AR model of order 2 (see Figure 3), thus implying that the measurement result at time *t* is significantly correlated with the measurement results at time *t*-1 and *t*-2. Considering the resolution of the time series, *t*-2 corresponds to 30 min.

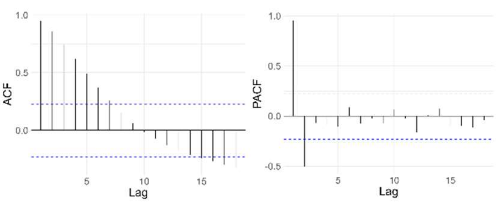

Fig. 3. ACF and PACF of one of the two stationary subseries of the calorific value time series.

The goodness–of–fit of the AR models has been evaluated by simulating 100 000 processes, where each simulation starts from the same initial value(s) (an AR(1) model requires one initial value, while an AR(2) model requires two starting points) as the series of the recorded measurement results and the following observations are simulated using the fitted AR models. The experimental data are within the 95 % confidence interval determined by the simulations (see Figure 4), thus the models are an adequate representation of the measurement data.

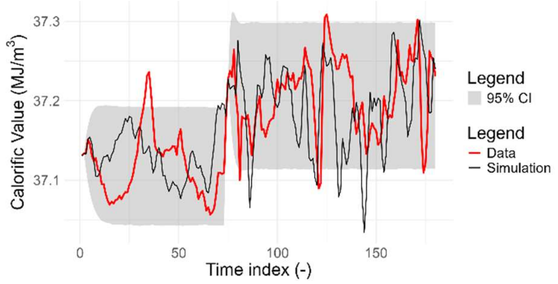

Fig. 4. Comparison between experimental data and simulated data.

2.6 Evaluation of uncertainty

The uncertainty is evaluated by applying the law of propagation of uncertainty with correlated input quantities [8]:

$$u_c^2(y) = \sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}u(x_i,x_j) \quad (1)$$

In the case of the total energy,

$$E_{tot} = \sum_{t=1}^{N} E_t = \sum_{t=1}^{N} V_t H_t \quad (2)$$

the double summation for the calculation of the covariance simplifies since the only correlation different from zero is between E_t and E_{t-1}, and between E_t and E_{t-2}. Furthermore, due to the stationarity of the time series, the correlation values are independent on the value of the index *t*. We note that the measured values have their own measurement models, describing predominantly the instrumental measurement uncertainty. Due to the fact that the entire time series is obtained from the same measurement equipment, there are correlations between the data due to the instrumentation. These come on top of the correlations identified in the time series analysis and are not considered in this work.

If only the instrumental measurement uncertainty had been considered, then the time series would have been modelled using a white noise model, so a model for independently and identically distributed data. When considering the serial correlation instead we consider the variance of the AR processes (AR(1) for the volume, AR(2) for the calorific value; see Chatfield, 1975 [6] for the formula of the variance of an AR process) and their corresponding correlations. Since the energy is given by the product of two random variables, we used the formula by Bohrnstedt and Goldberger, 1969 [9] to determine the covariance of the total energy, where we assumed that the volume and calorific value measurements are mutually independent. This formula considers all effects arising from the presence of covariances, this in contrast to the law of propagation of uncertainty from the Guide to the expression of Uncertainty in Measurement [8], which carries only the main contributors to the combined standard uncertainty of the output quantity.

3 Results

The effect of adding the serial correlation in the calculation of the uncertainty has been evaluated using two metrics, i.e.,

- the relative difference between the uncertainty of the total energy calculated including the serial correlation $u_{\text{cor}}(E)$ and the uncertainty calculated under the assumption that the measurements are independent and identically distributed $u_{\text{iid}}(E)$:

$$\Delta_{rel} = \frac{u_{\text{cor}}(E) - u_{\text{iid}}(E)}{u_{\text{iid}}(E)} \cdot 100\% \quad (3)$$

- the ratio between the calculated uncertainty under the two scenarios and the total energy (i.e., the relative uncertainty).

These metrics have been calculated over 100 000 simulated data obtained with the time series model fitted on the available measurement data. The results are displayed in Figure 5 and 6. The distributions of the uncertainty values calculated in the two scenarios are clearly separated with no overlap between their 95 % confidence intervals. The inclusion of the serial correlation leads to an increase of the measurement uncertainty by approximately 50 %, with the 95 % confidence interval covering the interval [38 %, 61 %]. Thus, when evaluating the measurement uncertainty without considering serial correlation, the uncertainty is severely underrated, which is expected to be the more relevant the more dynamic (and thus the less stationary) the gas flow is in the pipe.

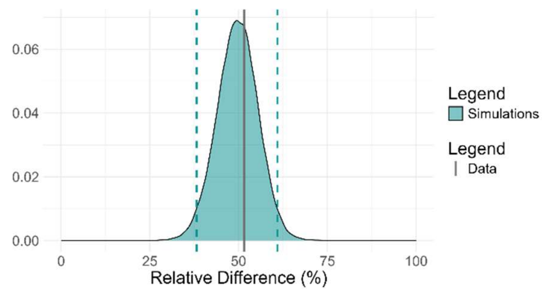

Fig. 5. Relative difference between the uncertainty values obtained including the correlation and the values obtained under IID assumption.

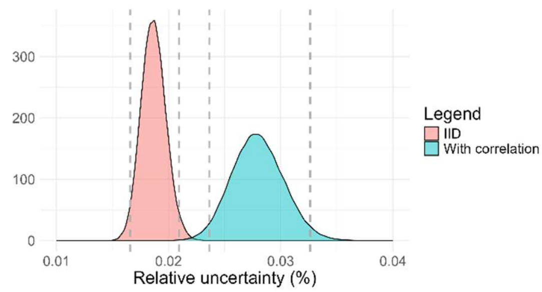

Fig. 6. The distributions of the values of the relative uncertainties associated to the total energy in the two considered scenarios.

4 Conclusions

With an increasing diversity of energy gasses entering the gas grid (e.g., hydrogen and biomethane), it is expected that the temporal effects due to the continuous process of the gas flowing through a pipe will substantially contribute to the uncertainty of the total energy. In this paper the time series of volume and calorific value measurement results from a metering station in the Dutch gas grid has been analysed to quantify the correlation between subsequent measuring results.

Since many time series models require stationary time series, a change point method called binary segmentation has been applied to automatically segment the given series of data into stationary subseries. Autoregressive models have then been applied to the resulting stationary subseries. In particular, it has been found that the volume measurement data could be described using an AR(1) process while the calorific value data are better modelled by an AR(2) process. The fitted AR process have then been validated by confirming that the experimental data are within the 95% confidence interval established by simulating a large number of data.

The simulated data have then been used to calculate the instrumental measurement uncertainty for the case with and without correlations. The results revealed an increase of typically 50 % in the uncertainty when including the temporal correlations. Considering the further diversification of the gas entering the gas grids in the foreseeable future, measurement data should not be treated as mutually independent when calculating the uncertainty to ensure safe and efficient operations of the gas grids. Further work is planned in the project 24GRD10 SmartGasNet to evaluate temporal correlations in case of non-stationary time series.

The authors would like to thank Robert Kruithof, Daniël Bos, Kjetil Folgerø and Astrid Marie Skålvik for helpful discussions during the preparation of this work.

The project 21GRD05 Met4H2 has received funding from the European Partnership on Metrology, co-financed from the European Union’s Horizon Europe Research and Innovation Programme and by the Participating States.

The data used in this work have been shared with the authors by Gasunie for the purposes of the Met4H2 project.

Author contributions
Conceptualization: FG, ML, AvdV
Methodology: FG, ML
Software: FG, ML
Validation: FG
Formal analysis: FG, ML
Investigation: FG, ML
Resources: AvdV
Writing - Original draft: FG, AvdV, ML
Writing – Review & Editing:
Visualization: FG, ML
Supervision: AvdV
Project administration: AvdV
Funding acquisition: AvdV

References

1. International Organization for Legal Metrology, OIML R140 Measuring systems for gaseous fuel. (2007).
2. European Committee for Standardization, EN 1776 Gas infrastructure – Gas measuring systems – Functional requirements. (2015).
3. International Organization for Standardization, ISO 15112 – Energy determination. (2011).
4. A.M.H. van der Veen, K. Folgerø, F. Gugole, Measurement uncertainty in the totalisation of quantity and energy measurements in gas grids. (under review). https://doi.org/10.20944/preprints202412.0011.v1
5. R.H. Shumway and D.S. Stoffer, Time series analysis and its applications: with R examples, (EDP Springer International Publishing, 2017)
6. C. Chatfield, The analysis of time series: theory and practice (Chapman and Hall, 1975)
7. J. Bai, Estimating Multiple Breaks One at a Time, Economet. Theor., **13**(3), pp. 315–352, (1997). https://doi.org/10.1017/S0266466600005831.
8. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, Evaluation of measurement data – Guide to the expression of uncertainty in measurement, JCGM 100:2008, GUM:1995 with minor corrections. (2008).
9. G.W. Bohrnstedt, A.S. Goldberger, On the exact covariance of products of random variables. J. Am. Stat. Assoc. **64**, pp. 1439–1442 (1969).