An Algorithm for Sensor Data Uncertainty Quantification

James T. Meech and Phillip Stanley-Marbell

Department of Engineering, University of Cambridge, Cambridge, CB3 0FA, UK

Student Member, IEEE

Senior Member, IEEE

Manuscript received June 7, 2017; revised June 21, 2017; accepted July 6, 2017. Date of publication July 12, 2017; date of current version July 12, 2017.

Abstract—This article presents an algorithm for reducing measurement uncertainty of one physical quantity when given oversampled measurements of two physical quantities with correlated noise. The algorithm assumes that the aleatoric measurement uncertainty in both physical quantities follows a Gaussian distribution and relies on sampling faster than it is possible for the measurand (the true value of the physical quantity that we are trying to measure) to change (due to the system thermal time constant) to calculate the parameters of the noise distribution. In contrast to the Kalman and particle filters, which respectively require state update equations and a map of one physical quality, our algorithm requires only the oversampled sensor measurements. When applied to temperature-compensated humidity sensors, it provides reduced uncertainty in humidity estimates from correlated temperature and humidity measurements. In an experimental evaluation, the algorithm achieves average uncertainty reduction of 10.3%. The algorithm incurs an execution time overhead of 5.3% when compared to the minimum algorithm required to measure and calculate the uncertainty. Detailed instruction-level emulation of a C-language implementation compiled to the RISC-V architecture shows that the uncertainty reduction program required 0.05% more instructions per iteration than the minimum operations required to calculate the uncertainty.

Index Terms—Sensor, Noise, Bayes, Rule, Uncertainty

I. INTRODUCTION

There are two main types of uncertainty in sensor measurements, aleatoric uncertainty which refers to the uncertainty due to noisy measurements and epistemic uncertainty which refers to the absence of measurements. We will refer to the noise and offset free true value of the physical quantity that we are trying to measure as the measurand. We will refer to any phenomenon such as temperature, pressure, humidity, and so on, measured using a sensor, as a physical quantity. Many noise reduction techniques rely on filtering or averaging the random variation in sensor signals to reduce the aleatoric uncertainty [1]. The lock-in, Kalman, and particle filters are notable exceptions. The lock-in filter attenuates all frequencies outside a narrow band around the signal of interest and must know the expected frequency and phase of the signal [2]. The Kalman and particle filters combine measurements and estimates of the system state to reduce uncertainty [3]. The Kalman filter calculates the proportion of the weight given to the measurement and the estimate using the covariances of the signals. The particle filter calculates the likelihood of individual sensor samples using an estimate of the state [4]. Our method is fundamentally different to the Kalman and particle filters as it relies on correlation in the noise rather than the signals (as it requires sampling faster than it is physically possible for the measurement to change) and doesn’t require the state equations like the Kalman filter or the map of one physical quantity like the particle filter. Our approach differs from related work that measures correlation in sensor signals to reduce uncertainty [5, 6] as we apply constraints to ensure that we are only measuring correlation in the noise.

1) Contribution: This article introduces a new method for reducing the uncertainty of sensor measurements. The method uses the joint distribution of the noise in two signals to reduce the uncertainty in one of the signals. Our method quantifies noise by sampling faster than it is possible for the time-varying measurement signal to change due to physical constraints (the thermal time constant) in the sensor. Our method assumes that the measurands are constant and uses Bayes rule to calculate a new conditional probability distribution for the most uncertain of the two physical quantities. The method checks that each noise distribution is Gaussian with thresholds on skewness and kurtosis.

II. THEORY AND ASSUMPTIONS

We can model the additive noise in a given temperature measurement as a random variable $T_n$ with a mean of $\mu_{T_n}$ and standard deviation of $\sigma_{T_n}$. Similarly we can model the additive noise in a humidity measurement as a random variable $H_n$ with a mean of $\mu_{H_n}$ and a standard deviation of $\sigma_{H_n}$. Algorithm [1] shows the minimum operations required to measure and calculate the uncertainty in two coherently-sampled sets of temperature and humidity measurements. Figure [1] shows a distribution of real sensor data collected using a state-of-the-art custom embedded system [7] and the histograms of samples from their fitted Gaussians. We used the BME680 [8], a sensor which measures temperature, pressure, humidity, and indoor air quality to
collect the data for Figure 1. Section II-A shows that 99.2 % of the measurements we took using the BME680 satisfy our empirical thresholds for the data to be considered Gaussian distributed. The assumption that $T_n$ (or the noise in any other physical quantity) is Gaussian distributed is only valid if $\sigma_{T_n} > \Delta T$ where $\Delta T$ is the change in the measurement during the time taken to sample the noise distribution [9]. The measurement will follow the value of the measurand with a time lag determined by the time constant of the sensor, $\tau_T$. In practice, $\tau_T$ places an upper limit on the rate of change of the measurement $\frac{\Delta Y}{\Delta t}$. Providing that the reciprocal of the sample rate, $\frac{1}{T_s} << \tau_T$, the measurement cannot change enough to produce a skew in the noise distribution. For the BME680, the humidity sensor time constant is $\approx 8$ s according to the datasheet [8]. Section III-A describes an experiment where we measured the time constant of the BME680 temperature sensor on an embedded system [7] to be $\approx 478$ s. All measurements in Section III satisfy the constraint $\frac{1}{T_s} << \tau_T$. The constraint is an extension of the Nyquist-Shannon sampling theorem. The BME680’s datasheet does not specify the skew in the noise distribution. For the BME680, the humidity sensor probability distribution for the humidity:

$$f_{H,T}(h, t) = \frac{1}{2\pi \sigma_H \sigma_T \sqrt{1 - \rho_{H,T}^2}} e^{-\frac{(t - \mu_T)^2}{2\sigma_T^2} - \frac{(h - \mu_H)^2}{2\sigma_H^2} - \frac{2(\rho_{H,T})(h - \mu_H)(t - \mu_T)}{\sigma_H \sigma_T (1 - \rho_{H,T}^2)}}.$$  \hspace{1cm} (1)

If our assumptions hold and the noise in $H_n$ and $T_n$ is Gaussian distributed, our embedded system can:

1. Take $N$ pairs of samples from the humidity and temperature signals with $\frac{1}{T_s} << \tau_T$.
2. Assume the noise is Gaussian distributed and additive.
3. Check that the skewness and kurtosis of the noise are within a threshold where this assumption is acceptable.
4. Fit the sampled noisy measurements to Equation 1.
5. Use Bayes rule and assert that $t = \mu_T$ to obtain a new conditional probability distribution for the humidity:

$$f_{H,T}(h|t = \mu_T) = \frac{f_{H,T}(h, t = \mu_T)}{f_{H,T}(t = \mu_T)} = \frac{1}{\sqrt{2\pi \sigma_H \sigma_T}} e^{-\frac{(h - \mu_H)^2}{2\sigma_H^2} - \frac{1}{2\sigma_H^2} \frac{\rho_{H,T}^2 t - \mu_T h}{1 - \rho_{H,T}^2}}.$$ 

$$\hat{\sigma}_{H,T} = \sigma_H \sqrt{1 - \rho_{H,T}^2}.$$ \hspace{1cm} (2)

To ensure that the assumptions are valid for each distribution the method checks that kurtosis $< 7$ and that the $|\text{skew}| < 2$ [10]. These thresholds are widely referred to as a good rule-of-thumb check that a set of samples are Gaussian distributed. If the noise violates either of these conditions then the method will report the $\sigma_{H_n}$ instead of the conditional $\hat{\sigma}_{H,T}$. Algorithm 2 shows the new method we propose for calculating the humidity uncertainty at a fixed temperature using Equation 2.

Algorithm 1: The bare minimum set of operations required for uncertainty measurement and quantification.

Result: $\mu_T$, $\mu_H$, $\sigma_T$, $\sigma_H$.

1. Collect $N$ samples from $H_n$.
2. Collect $N$ samples from $T_n$.
3. $\mu_H = \text{Mean}(H_n)$.
4. $\mu_T = \text{Mean}(T_n)$.
5. $\sigma_H = \text{Standard deviation}(H_n)$.
6. $\sigma_T = \text{Standard deviation}(T_n)$.

Algorithm 2: Our new proposed method for uncertainty measurement and quantification at a given temperature.

Result: $\mu_T$, $\mu_H$, $\sigma_T$, $\sigma_H$.

1. Collect $N$ samples from $T_n$ and $H_n$ at sample rate $f_s$.
2. $\mu_H = \text{Mean}(H_n)$.
3. $\mu_T = \text{Mean}(T_n)$.
4. $\sigma_H = \text{Standard deviation}(H_n)$.
5. $\sigma_T = \text{Standard deviation}(T_n)$.
6. if $\text{|skewness| < 2}$ & $\text{(kurtosis < 7)}$ then
   7. $\rho_{H,T} = \text{Correlation}(H_n, T_n)$.
   8. $\hat{\sigma}_{H,T} = \sigma_H \sqrt{1 - \rho_{H,T}^2}$.
   9. else
   10. $\hat{\sigma}_{H,T} = \sigma_H$.
11. end.

A. Applicability to Other Sensors

We conducted an empirical sensor applicability study using an embedded sensor system [7]. The sensors include two 3-axis accelerometers (MMA8451Q, BMX055), a gyroscope (L3GD20H), a magnetometer (MAG3110), and an environmental sensor (BME680). We placed the sensor system on a vibration isolation stage in a thermal chamber set to 25 °C with a sensor supply voltage of 2.5 V, waited for one hour for the temperature to equilibrate, then sampled 10,000 samples from each sensor at a sample rate of 4.9 Hz. Figures 3(a) and 3(b) show the kurtosis and skewness of 333 thirty-sample distributions from each axis or physical quantity from each sensor. We calculated that 75.8 % of the L3GD20H data, 99.6 % of the MAG3110 data, 100 % of the BMX055 data, 99.9 % of the MMA8451Q data, and 99.2 % of the BME680 data are within the threshold. Figure 3(c) shows the 10,000 sample stationary noise distributions from the z-axis of each inertial sensor and the BME680 pressure sensor. The noise distribution of the L3GD20H is a trimodal Gaussian which causes the distributions to fail to meet the skewness and kurtosis thresholds significantly more than the other sensors.
B. How is This Different to Kalman and Particle Filters?

The Kalman and particle filters measure correlation in the sensor signals whereas Algorithm 2 samples faster than it is physically possible for the measured to change and measure correlation in the noise. Algorithm 2 calculates the standard deviation of the noise at a given temperature and allows the noise to depend on the measured and other environmental conditions whereas the Kalman filter does not. Both the Kalman and particle filters require equations that describe the relationship between the sensor signals, our method requires only the set of oversampled temperature measurements. The Kalman filter applies no online checks on the shape of the distribution. In contrast, Algorithm 2 applies checks on the skewness and kurtosis. The particle filter can support multimodal distributions but Algorithm 2 cannot support anything other than a unimodal Gaussian distribution. Unlike the particle filter Algorithm 2 has no predict step and instead measures correlation in the noise.

1) Computational Complexity: Let \(N_0\) be the number of output states in our system and \(M\) be the number of weights. Arulampalam et. al. and Haykin state that the computational and memory complexity of both the extended Kalman filter and the particle filter is \(O(N_0 M^2)\) [11]. We experimentally measured the computational complexity of Algorithm 2 to be \(O(N_0 M)\) by plotting the number of instructions required to run the algorithm for different values of \(N\). This is because we calculate the correlation between each variable and the one we are interested in rather than all correlations between all variables which are required by the Kalman and particle filters. Assuming that the hardware complexity scales with the computational complexity, Algorithm 2 offers considerably lower hardware complexity than the Kalman and particle filters.

2) Comparison Against the Kalman and Particle Filters: We cannot perform a like for like experimental complexity comparison with the Kalman or particle filters for this application. The Kalman filter requires equations relating the humidity and temperature measurements to some shared state. In this application we do not have such equations. We cannot perform a like for like experimental comparison with the particle filter because we have no map of how the humidity will change over time.

C. Sensor Measurement Temperature Compensation

The BME680 [8] uses an internal temperature sensor to compensate its pressure, humidity, and air quality measurements. Let \(K_{T1} - K_{T3}\), \(K_{H1} - K_{H7}\), and \(K_{P1} - K_{P10}\) be calibration constants stored inside the sensor’s digital logic and \(T_{ADC}\) be the value read by the analogue to digital converter (ADC) of the temperature sensor in the BME680. We translated the functions calc_temperature, calc_humidity, and calc_pressure provided by Bosch Sensortech [12] into equations. The converted temperature

\[
T_{Out} = \frac{K_{T2}}{5120} \left( \frac{T_{ADC}}{2^{14}} - \frac{K_{T1}}{2^{10}} \right) + \frac{K_{T3}}{5120 \times 2^{10}} \left( \frac{T_{ADC}}{2^{14}} - \frac{K_{T1}}{2^{10}} \right)^2.
\]

Let \(H_c\) be temperature correction coefficient used to compensate the humidity calculation where

\[
H_c = \frac{K_{H2}}{2^{16}} + \frac{K_{H2} K_{H4} T_{Out} + K_{H2} K_{H5} T_{Out}^2}{2^{38}}.
\]

Let \(H_{ADC}\) be the value read by the ADC of the humidity sensor where the temperature-compensated humidity

\[
H_{Out} = H_c \left( H_{ADC} - \frac{K_{H1} T_{Out}}{2} + \frac{K_{H3}}{2} T_{Out}^2 \right) + \left( \frac{K_{H6}}{2^{14}} + \frac{K_{H7} T_{Out}}{2^{21}} \right) H_c \left( H_{ADC} - \frac{K_{1} T_{Out}}{2} + \frac{K_{H3}}{2} T_{Out}^2 \right).
\]

Equations [4] and [5] show that any noise in the temperature measurement will propagate through to the humidity measurement. Let \(P_{C1}\) and \(P_{C2}\) be the temperature correction coefficients used to compensate the pressure calculation where

\[
P_{C1} = \frac{K_{P6}}{2^{10}} (2560T_{Out} - 64000)^2 + K_{P5}(1280T_{Out} - 32000) + 2^{10} K_{P4}
\]

\[
P_{C2} = \frac{K_{P1} K_{P3}}{2^{34}} (2560T_{Out} - 64000)^2 + K_{P1} K_{P2} (2560T_{Out} - 64000) + K_{P1}
\]

Let \(P_{C3}\) be an intermediate variable in the pressure calculation where

\[
P_{C3} = \frac{6520}{P_{C2}} (2^{20} - P_{ADC} - \frac{P_{C1}}{2^{12}}).
\]

Let \(P_{ADC}\) be the value read by the ADC of the pressure sensor where

\[
P_{Out} = \begin{cases} 
P_{C3} \left[ 1 + \frac{K_{P3}}{2^{20}} \right] + 2^{2} K_{P7} , & \text{if } P_{C2} \neq 0 \\[2^{20} - P_{ADC} , & \text{otherwise.} \end{cases}
\]

III. METHODOLOGY

1) Motivation for Experiments: The output of high accuracy metal-oxide gas sensors depends on relative humidity [13]. These sensors need to measure extremely small gas concentrations (10 ppm), decreasing the noise in the humidity measurement will enable more accurate compensation for humidity.

A. Temperature Sensor Experiments

To calibrate the temperature sensor we placed a sensor system that contains a BME680 in a thermal chamber and set the temperature to 0°C. We had the sensor system sample at a constant temperature for one hour before performing the calibration to allow the temperature to reach a steady state. We had the embedded system sample 12,000 values and repeated this at each temperature from 0 to 40°C with a step of 5°C. The thermal time constant is the time required for a system subjected to a step change in temperature to reach 1 - \(e^{-1}\) of its final temperature [14]. We need to measure the thermal time
B. Uncertainty Quantification Experiments

We placed two identical sensor systems in a typical indoor environment running Algorithm 2 at 22.3 Hz and Algorithm 1 at 23.6 Hz for one hour. We had one sensor system calculate the mean and standard deviation of the measurement distribution using Algorithm 1 and the other calculate the uncertainty at the mean temperature using Algorithm 2. We repeated this experiment five times. Figure 4(a) shows the experimental setup that we used to measure the power consumed by each sensor system simultaneously. In a separate experiment we used the Sunflower embedded system emulator to emulate both C-language implementations to obtain estimates of the number of instructions required to run them on a RISC-V processor with a sensor sample rate of 128.3 Hz and a clock speed of 60 MHz.

IV. RESULTS AND DISCUSSION

A. Temperature Sensor Experiments

Figure 4(a) shows the response of the sensor to a step change in temperature from 50 °C to room temperature. We measured the thermal time constant of the temperature sensor to be 478 s. Figure 4(b) shows the result of performing a least-squares regression to map the values from the sensor to the thermal chamber set temperature. We assume that the thermal chamber set temperature is the ground truth for the temperature measurement. We fit a linear model to remove the temperature offset due to the heating from other components. The regression resulted in a best fit line of $T_{\text{meas}} = 1.001 \times T_{\text{sensor}} - 2.179$.

B. Uncertainty Quantification Experiments

Algorithm 2 was able to reduce the standard deviation of humidity measurements by 10.3 % on average and to do this for 95.4 % of the experimental data. Figure 5 shows the distribution of the skewness and kurtosis for each of the 30 sample distributions for both temperature and humidity for a separate five-hour dataset. Figure 5 shows that the data rarely exceeded the thresholds on skewness and kurtosis. The conditions on line 6 of Algorithm 2 are satisfied for 97.9 % of the data. Algorithm 2 requires 2.4 % less energy than Algorithm 1 if we run both algorithms for the same period of time. Algorithm 1 runs faster, samples from the sensor more in a given amount of time and therefore consumes more power on average. The implementation of Algorithm 2 required 5.3 % more time to run per iteration than that of Algorithm 1.

After running 100 iterations of each C implementation on the processor emulator for 30 sample distributions the instruction-level program emulation showed that the C-implementation of Algorithm 2 required 0.05 % more instructions to run than that of Algorithm 1. The number of RISC-V instructions executed by the C-language implementations is independent of the system clock speed and therefore generalises to any typical modern architecture. The instruction costs for both Algorithm 1 and 2 scale with $O(N)$. 

V. CONCLUSION

This article presents an algorithm that can calculate the aleatoric uncertainty of sensor measurements at a fixed temperature by sampling faster than it is physically possible for the measurement to change to characterise the noise. Our method (Algorithm 2) reduced the uncertainty in humidity measurements by 10.3 % on average for a speed decrease of 5.3 %. The speed decrease leads to Algorithm 2 requiring 2.4 % less energy to run than the baseline (Algorithm 1) for the same time period. Our approach requires less power than the state of the art which will make it attractive to embedded systems engineers working with power-constrained devices. The performance of Algorithm 2 is limited to an uncertainty reduction of 10.3 % for all applications. In applications where the noise has larger correlation or is sampled at a higher rate we are likely to see larger uncertainty reductions. Stronger or additional constraints on the value of the measurand of one of the physical quantities would lead to further uncertainty reduction. A weakness of the method is the requirement for the sensor noise to be Gaussian. For some sensors such as the L3GD20H this requirement is not satisfied. Extending the method for the case where the sensor noise is not Gaussian but instead some arbitrary $N$ dimensional distribution where $N$ is the number of sensor signals is an open research problem.

REFERENCES

[1] F. Zhou and Y. Chai, “Near-sensor and in-sensor computing,” Nature Electronics, vol. 3, no. 11, pp. 664–671, Nov 2020.
[2] J. H. Scofield, “Frequency-domain description of a lock-in amplifier,” *American journal of physics*, vol. 62, no. 2, pp. 129–133, 1994.

[3] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Trans ASME, J Basic Eng.*, vol. 82, no. Series D, pp. 35–45, 1960.

[4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking,” *IEEE Trans.*, vol. 50, no. 2, pp. 174–188, 2002.

[5] V. Gupta and S. De, “Collaborative multi-sensing in energy harvesting wireless sensor networks,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 426–441, 2020.

[6] S. Silvestri, R. Urgaonkar, M. Zafer, and B. J. Ko, “A framework for the inference of sensing measurements based on correlation,” *ACM Trans. Sen. Netw.*, vol. 15, no. 1, Dec. 2018.

[7] P. Stanley-Marbell and M. Rinard, “Warp: A hardware platform for efficient multimodal sensing with adaptive approximation,” *IEEE Micro*, vol. 40, no. 1, pp. 57–66, Jan 2020.

[8] “BME680 Low Power Gas, Pressure, Temperature & Humidity Sensor,” *Bosch Sensortec*, 2019.

[9] J. T. Meech and P. Stanley-Marbell, “Efficient programmable random variate generation accelerator from sensor noise,” *IEEE ESI*, 2020.

[10] S. West, J. Finch, and P. Curran, “Structural equation modeling: concepts, issues and applications,” *SAGE, Thousand Oaks, CA*, p. 74, 1995.

[11] S. Haykin, *Kalman filtering and neural networks*. John Wiley & Sons, 2004, ch. 2, p. 33.

[12] “BME680 sensor APL,” Aug 2020. [Online]. Available: https://github.com/BoschSensortec/BME680_driver

[13] P. T. Moseley, “Progress in the development of semiconducting metal oxide gas sensors: a review,” *MST*, vol. 28, no. 8, p. 082001, Jun 2017.

[14] A. Pourmovahed and D. R. Otis, “An Experimental Thermal Time-Constant Correlation for Hydraulic Accumulators,” *J. Dyn. Sys., Meas., Control*, vol. 112, no. 1, pp. 116–121, March 1990.

[15] P. Stanley-Marbell and M. Hsiao, “Fast, flexible, cycle-accurate energy estimation,” in *ISLPED’01*. New York, USA: ACM, 2001, pp. 141–146.

[16] “Binder MK56 thermal chamber calibration certificate,” Oct 2020. [Online]. Available: https://github.com/physical-computation-measurements/Binder-MK56-Chamber-Calibration-Certificate