

# Bayesian Estimation for Airborne Hygrometry

Greg CHADWICK

*University of Exeter*

# Contents

<b>Abstract</b>	<b>3</b>
<b>Acknowledgements</b>	<b>4</b>
<b>Chapter 1</b>	<b>5</b>
Introduction . . . . .	5
<b>Chapter 2</b>	<b>7</b>
2.1 Context for the research . . . . .	7
2.2 The hygrometers and their performance . . . . .	7
2.3 The Bayesian approach . . . . .	10
2.4 Dynamic linear models (DLMs) . . . . .	11
2.5 Melding . . . . .	14
2.6 Opinion pooling . . . . .	16
<b>Chapter 3</b>	<b>20</b>
3.1 Framing the problem . . . . .	20
3.2 Application of the DLM framework . . . . .	21
3.3 Application of opinion pooling . . . . .	26
3.4 Reaching a final estimate . . . . .	29
<b>Chapter 4</b>	<b>32</b>
Discussion and further work . . . . .	32
<b>References</b>	<b>33</b>

## Abstract

A significant challenge faced when producing estimates for the true state of an underlying natural process is the combination of information obtained from multiple sources. Water vapour is the most influential of the greenhouse gases, so the means by which measurements from multiple hygrometers are collated and interpreted will have crucial implications when forming a judgement as to the nature and extent of its content in our atmosphere at a given time or place. This report offers a Bayesian approach relating to Dynamic Linear Models and the combination of evidence from alternative sources to produce a framework for producing estimates of the underlying absolute humidity during the course of a selection of FAAM research flights.

## Acknowledgements

Many thanks to Dr Phil Sansom for providing access to the code used in the paper Sansom, Williamson and Stephenson (2019). Thanks also go to both of my supervisors, Dr Alan Vance and Dr Daniel Williamson, for the guidance and support during the course of this project.

This dissertation was written in R version 3.6.1 (R Core Team 2019) and all relevant code used to produce the figures and results in this report is made available at the GitHub repository [github.com/gjc211/thesis](https://github.com/gjc211/thesis).

# Chapter 1

## Introduction

One of the fundamental challenges faced when producing models used in the analysing and tracking of systems in the natural world is that of measurement consistency; in order to understand and act upon information gained from such research, experts must be confident in their estimates reflecting the true state of the environment at a given time or place. As the world continues to face the growing threat of climate change, we must therefore continue seeking to understand in detail how our growing global commercial activities are affecting this rate of change, the specific drivers and where resources must be directed if change can be enacted in the near future. The shifts in weather patterns associated with climate change and global warming can only be monitored effectively through the efficient collation and interpretation of measurements from disparate sources and accounts. These data, whilst seeking to understand the same natural phenomena, may achieve their measurements through entirely different metrics, thereby making their inference a significant obstacle. Converting measurements from alternate sources to common units will therefore often lead to inconsistency as it is improbable that the specific quantities each are measuring will exhibit a perfect correlation given the range of conditions and extraneous factors at play in any evolving natural process. In practice, researchers may be able use their expertise to interpret these appropriately in order to perform statistical inference, but this as an approach is unlikely to be efficient.

In a rapidly evolving commercial landscape, with decision making and the distribution of information increasingly becoming reliant on big data processes, we are faced with greater potential to explore the incorporation of knowledge and expertise from multiple expert sources into statistical models for unmeasurable natural processes. This dissertation will be based on addressing the difficulties currently faced in measuring and monitoring patterns of water vapour in the atmosphere, which has been established as the most significant of the greenhouse gases (Vance *et al.* 2015). The aim of the analysis will be to produce a reliable model which can estimate the ‘true’ absolute humidity given the unique measurements of four different hygrometers in a range of atmospheric conditions. These instruments will be attached to the same aircraft so it would be expected that they would be measuring the same air at any given point, however there are frequent occasions where discrepancies occur, either systematically or with a significant degree of uncertainty. The idea for this analysis is to identify and incorporate into a model the various biases or dependencies that the individual hygrometers might exhibit differently with time dependent factors such as temperature, pressure and humidity, but also to identify where certain instruments are unlikely to be reliable in unique circumstances. Dynamic Linear Models (DLMs) are a widely favoured approach

to time series modelling and have applications in a variety of settings which attempt to understand dynamic physical systems measured alongside the presence of random disturbances (Petrakis *et al.* 2009). The estimations of the underlying system throughout the time series can be obtained through recursive filtering algorithms which make predictions based on previous values and observations. Its parameters can thus be controlled in a variety of ways to influence how these estimates are reached at each point, and this principle will play a large part in the methods used throughout this dissertation to calculate this best estimate of absolute humidity. The DLM approach is therefore Bayesian in its nature as the estimates of the distributions of the unobservable ‘state’ process are sequentially updated with new information throughout the filtering process. This approach would therefore be consistent with the requirements of the problem at hand here; we have four different instruments providing measurements, each of which are giving some indication of the underlying process of atmospheric water vapour.

It would be easy to assume that the state process in any physical system can be estimated to a high level of accuracy as long as the parameters are specified correctly and extraneous variables are sufficiently taken into account, but in practice there is a significant degree of uncertainty throughout the process. The nature of the hygrometers at the focus of this report means that there are many situations where instability in their measurements occur with little precedent. Further, there are instances where a pattern of humidity given by one of the instruments over a time period is plausible when viewed in isolation. It is when presented with the measurements of other instruments that any inference of the true state can and should be made. These are the limitations that we are faced with in this research and is always worth bearing in mind before exercising judgement on the trends or patterns that are shown. The next two sections will be to introduce the four hygrometers at the centre of this research and to discuss their intricacies in the context of the later analysis.

## Chapter 2

### 2.1 Context for the research

As has previously been presented, hygrometry is an area of study that is gaining in importance with growing fears of climate change and global warming. Seeking to understand and effectively monitor water vapour in the atmosphere will be crucial to our understanding of its distribution but also its interaction with atmospheric variables and other significant greenhouse gases. In depth data obtained in a variety of conditions from a number of research flights has provided useful information which can be used in conjunction with previously available satellite data to consolidate methods of numerical weather prediction (NWP). The data that will be included in this research and to formulate models are from numerous flights of the Facility for Airborne Atmospheric Measurement's (FAAM) research aircraft. These flights occur across a number of different countries and in a range of conditions, so any model which is developed generally should be applicable or at least easily adjustable depending on any unique variables that may be present. Each flight collects data in the form of a time series but which represent spatial changes in the atmosphere, dependent on many intricate factors including altitude, temperature, pressure and the speed of the aircraft. There are still a number of variables that cannot easily be quantified, such as the cleanliness of the air, so it is not always possible to understand why some variation in measurements occur over time and what specific shifts in the atmosphere they have detected, if there were any at all. The research of Vance *et al.* (2015) details the performance of the individual hygrometers in the various types of conditions encountered by the research aircraft, in terms of the biases they exhibit in certain situations and where any instability has tended to occur. Their findings will be discussed in this next section along with how these issues have manifested themselves in the data.

### 2.2 The hygrometers and their performance

There are four hygrometers which are attached to the FAAM research aircraft and each has a unique way of measuring water vapour. The conversion required to reach a figure for absolute humidity is therefore achieved through different means and has meant that the 'quality' of data at a given point is generally dependent on the instrument in question. There are two hygrometers that measure the dew or frost point – the temperature required for water vapour to condense as water droplets or form ice on a surface respectively – using the chilled mirror technique. The Buck CR-2 (Buck) is one of these instruments and works by pumping air through a heated pipe onto a mirror surface in a sample chamber where its temperature can be regulated. The appearance

of condensation is monitored optically in which case the temperature is then recorded. The Buck measurements are calibrated annually at the National Physical Laboratory so generally tend to provide an accurate account of the mirror temperatures, but it still has responded poorly in various conditions which will be discussed. The other chilled mirror hygrometer that has been attached to the aircraft is the General Eastern 1011B (GE), which does not possess the same heated inlet and presents differing biases to the Buck, especially relating to temperature. Both chilled mirrors exhibit limitations to their use at low humidities, most notably after a sharp drop which causes problems with determining the point at which the mirror deposit freezes and becomes the frost point. Two of the hygrometers are known as the WVSS-II; these instruments use diode lasers to measure the atmospheric water vapour mixing ratio, essentially by being able to emit a laser and determine the specific concentration of water vapour that would result in the change in wavelength that is observed once the signal is received again. The system has been shown to have a high level of accuracy and precision, along with a long lifetime (Fleming and May 2004), but its lack of calibration has meant the two instruments are subject to errors and biases in certain conditions. The main difference between the two WVSS-II hygrometers is the type of inlet used, known as the flush and Rosemount inlets. They frequently reach differing measurements generally due to the type of heat regulation within their sample chambers and their proximity from the outside of the plane. It is argued as to which is more effective in certain conditions, but the main issues still persist of poor performance at low humidities, especially without proper calibration.

The research in Vance *et al.* (2015) initially sets out the offsets which were calculated between the pairs of hygrometers after taking data from each that were seen to be stable and comparable. This would include times where instrument lag was not a major factor causing discrepancies and where the difference in their measurements was below a certain threshold above which suggested there was a likely a malfunction. The result was that 90% of the available data was rejected in the analysis. This offset was used in the assessments of how the hygrometers differed relatively in their performance at different humidities. The findings illustrated the instability of each of the instruments at lower humidities, with all four generally agreeing reasonably well at higher humidities. Despite the Buck having a greater capacity to measure data at lower humidities, the issues with the mirror temperature readings have meant it has struggled below a temperature of 203K. It was further found that the GE tended to over-read at low humidities. The discrepancy between the two WVSS-II hygrometers was significant as the humidity decreased; the flush inlet reading was generally higher than the Rosemount on average and was put down mainly to the differing designs of the inlets and their positions on the aircraft. After analysis of the data from around 100 flights in a range of conditions, they conclude that the Rosemount WVSS-II agrees well with the GE and Buck in cloud free conditions. The flush inlet on the other hand tended to over-read at lower humidities and was apparent most significantly when in comparison with its WVSS-II counterpart. Part of the research in this report will attempt to define this bias and produce estimates for what the ‘true’ humidity is likely to be, given the fact that each of these instruments



has struggled in drier conditions. To a large extent, knowing that there is a consistency to the bias shown by the flush WVSS-II for instance below a certain threshold is a useful benchmark and will be discussed in more detail during the latter stages of Chapter 3. The issue that will be faced is that the comparisons with the chilled mirror hygrometers may not be useful if they themselves show irregularities or if each other hygrometer struggles in different instances at low humidities.

Where significant issues or discrepancies may occur is following sharp changes in atmospheric conditions. The flush WVSS-II shows the most rapid response to a large change in humidity, with the chilled mirrors being the slowest of the four to respond. Generally, a sharp drop in humidity will cause the most issues for the chilled mirrors and often leads to a long ‘stabilisation period’ whereby their measurements will not capture any trends that are shown by the WVSS-IIs. Despite this research, there is still a deal of uncertainty about the extent to which the hygrometers react differently in certain conditions, so the estimates that will be obtained through the models included in the later analysis will only be able to approximate our best estimate of humidity given the current knowledge that is available. Further work in Vance *et al.* (2018) confirmed that a sufficiently accurate standard for these hygrometers to be compared did not exist for the period covered by the research. The GE was used as such in this research, but again is largely incapable of measuring low humidities. The main conclusion to take here was that the WVSS-IIs, in order to provide accurate and useful data for research, especially at low humidities, have to be regularly calibrated.

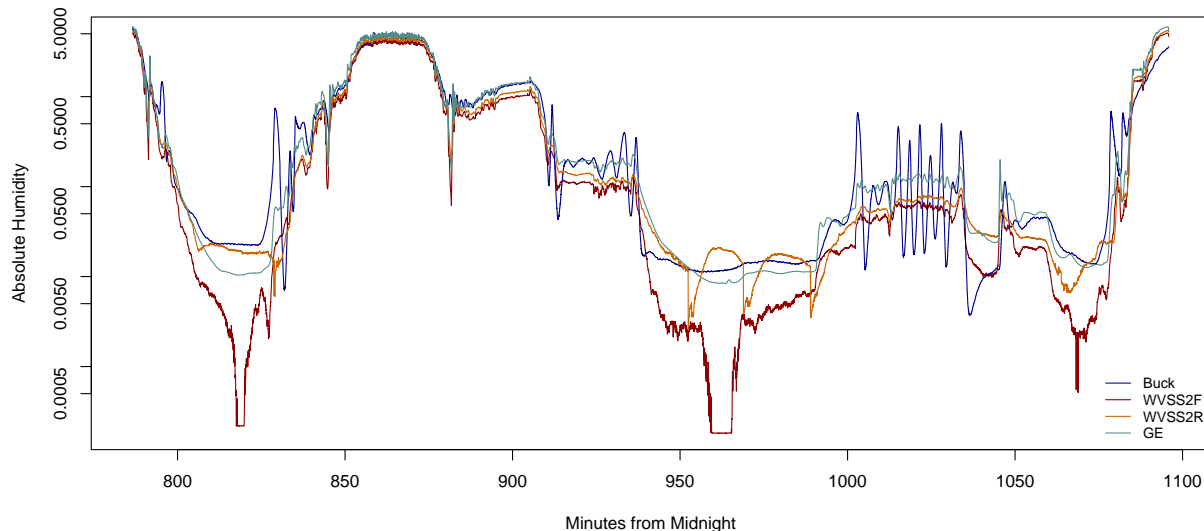


Figure 1: Flight B878

The time series in Figure 1 represents a FAAM research flight from 2014 over the east coast of

Scotland. The absolute humidity is given on a log scale in  $\text{g/m}^3$  to ensure that any trends can be easily observed. The four hygrometers appears to only agree well in the period where humidity was at its highest, around 870 minutes after midnight. In terms of the absolute humidity measurements, the disagreement starts to occur below around  $0.6\text{g/m}^3$ , which is consistent with the findings from Vance *et al.* (2015). The flush WVSS-II clearly starts to struggle significantly when the hygrometers are reporting their lowest humidities, illustrating the lack of reliability in these conditions and when its measurements are not regularly calibrated. Noticeably its WVSS-II counterpart, between 950 and 1000 minutes from midnight, is reporting a pattern that is known could not possibly occur in the atmosphere. Whilst the two chilled mirror hygrometers are reporting more believable measurements around this time, the Buck in particular is subject to periods of significant oscillations where the change in mirror temperature has caused difficulties in its monitoring of the deposit. In particular, rapid changes in atmospheric conditions often cause instability in the chilled mirror measurements and they are also slow to respond. The task will therefore be to use available knowledge of instrument biases in certain conditions and what the other instruments are reporting, in order to estimate where the truth is likely to be. The next two sections will go on to discuss how this problem may be understood by taking a Bayesian approach to deal with some of this uncertainty and using the DLM formulation to control how the underlying state is learnt through a recursive process.

## 2.3 The Bayesian approach

The principle of the Bayesian approach is to treat a statistical model as joint probability distribution consisting of the available data  $y$  and the unknown model parameters  $\theta$ . Whilst the alternative would be to compare the researcher’s internal beliefs about the probability of an event occurring to the outcome from new data, Bayesian statistics uses specified prior beliefs about the unknown event and updates this information with the results from the data to form a posterior belief. This posterior distribution can therefore be specified in terms of the product of the prior distribution of likely values of the unknown quantity  $p(\theta)$  and the likelihood  $p(y|\theta)$ . The prior distribution is therefore a set of subjective beliefs of the likely values of these parameters and the likelihood can be defined as the probability of obtaining the observed data given the current parameter settings. The posterior distribution reached can therefore be specified through Bayes’ theorem as

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}.$$

The marginal likelihood  $p(y)$  is a constant which normalises the product in the numerator so that the probability distribution integrates to one. Due to the complexity of calculating  $p(y)$ , the theorem is often rewritten as

$$p(\theta|y) \propto p(y|\theta)p(\theta) = p(y, \theta).$$

A frequentist approach would only be concerned with the likelihood function, as the results from the data are all that will be used to provide the inference about the probability of the unknown  $\theta$ , given a large enough sample. The Bayesian principle of incorporating some set of prior beliefs in conjunction with any data analysis can therefore be useful in this kind of setting where we wish to make inferences about the true state of a natural process. The aim for this project is to use expert knowledge to make inference about an unknown quantity after considering the available data; the approach therefore offers a solution to the major issue here of data quality. At each time point, we will have some prior beliefs about the true humidity from each hygrometer’s previous measurement but this would also include the understanding of their individual biases or dependencies with other factors and how they vary with each other. The data collected for each instrument will therefore allow inference to be made of the posterior beliefs indicating the underlying state process. This problem represents a unique challenge in that instinctively it seems that four unique models could be defined to provide inference about the state process, but crucially this inference must be made with reference to each of the other measurements at a given time point. This will be addressed later on, but for now the means by which DLMs and recursive algorithms can be used to update our beliefs about the state through time will be discussed.

## 2.4 Dynamic linear models (DLMs)

The state space representation is one that has been widely applied over the last fifty years in many settings such as engineering and economics to model dynamical systems whereby there exists an unobservable state process which is somewhat abstract and evolves over time. A state space model defines how this underlying process evolves over time according to various beliefs about its nature and interaction with other variables, in the attempt to predict the next value of the state in a time series. It defines the dependence structure in place between this latent process and the observable series of measurements; in constructing these models, the deterministic and stochastic or random characteristics of the measurements must be distinguished. The framework is consistent with the Bayesian approach due to the nature of the recursive algorithms used to estimate the underlying state process, which essentially take the prior beliefs obtained from the previous observations  $y_1, \dots, y_{t-1}$  to update beliefs about the likely value of the hidden state using the conditional probability  $p(\theta_t | y_1, \dots, y_{t-1})$ . Considering the simplest interpretation of the problem without the any time-dependency in the state, the value of the unmeasurable quantity can be represented as the value of observations with some added error (Petrus *et al.* 2009):

$$y_t = \theta + v_t, \quad v_t \sim N(0, V_t).$$

We can continually update our predictions of the unknown quantity  $\theta$  when presented with new information at each time step. Now when the element of time dependency in  $\theta$  is included, this

evolution of the underlying state process can be defined by

$$\theta_t = \theta_{t-1} + w_t, \quad w_t \sim N(0, W_t).$$

The posterior density of  $p(\theta_t|y_1, \dots, y_t)$  can therefore be defined by the prior distribution  $p(\theta_t|y_1, \dots, y_{t-1})$  and the likelihood  $p(y_t|\theta_t, y_1, \dots, y_{t-1})$  through Bayes' theorem outlined earlier. The observation and system variance are represented by  $V$  and  $W$  respectively and illustrate how the level of random error in the measurements and the state process vary over time, which can have a significant effect on how the state is learnt. The state process  $\theta_t$  is therefore known as a Markov chain because each of its values are dependent only on its past values up to  $\theta_{t-1}$ . Further, the observed data  $y_t$  are dependent entirely on this underlying process.

This brings the discussion to DLMS, which are a representation of Gaussian linear state space models and are referred to primarily during this project, forming the basis for approaching some of the problems encountered. The principle follows this state space representation of having an unobservable state variable which evolves over time and measurements which are dependent on this underlying process. The DLM equations are represented as follows:

$$y_t = F_t\theta_t + v_t, \quad v_t \sim N(0, V_t),$$

$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t).$$

The inclusion of the matrix coefficients  $F$  and  $G$  means that the nature of this evolution of the state over time can be specified, as well as its influence on the measurements. This time dependency included in these coefficients allows for their values over time to be conditioned on other factors and will prove useful for instance when attempting to correct biases in the measurements. If we know that  $\theta_t$  represents the true absolute humidity at a given time point and that  $y_t$  is known to deviate in some systematic way depending on factors such as temperature, humidity or pressure, this can potentially be encoded through these parameters. The observation and system variance may further be used to represent the elements of stochasticity in both equations; in a practical sense, they may be used to indicate uncertainty in the state evolution or where the instruments exhibit random fluctuations. This flexible framework has meant that DLMS and state space models have been applied to a great variety of practical problems.

Inference about the state for the purposes of this project will involve the consideration of past and present values – known as filtering – so that the estimates are updated sequentially at each step using all prior and current information. An alternative would be to also take into account information from future time steps, which results in retroactively smoothing the process; this would not be desired in this case as intuitively we are interested in the subtle changes in atmospheric conditions that the instruments are capturing at each point in time which then update the state estimate.

Filtering is solved through a recursive algorithm, which essentially means that in order to calculate the filtered density at the current time point, it will use the results from previous estimations of the state which were calculated through the same means. In the DLM context, the most prominent filtering algorithm is that of the Kalman filter. There are two stages involved in providing an estimate of  $\theta_t$ , which are ‘predicting’ and then ‘updating’ (Kim and Bang 2018). The predicted state is given by  $\theta_t^{t-1} = \phi\theta_{t-1}^{t-1} + \Upsilon u_t$ , and is achieved through the previous updated state estimate. The superscript on  $\theta$  represents the fact that this is the updated posterior from the previous time step, denoted with the subscript  $t-1$ . The coefficient  $\phi$  is the state transition matrix which controls the state evolution from the previous time point. The coefficient  $\Upsilon$  is the control-input matrix and is applied to the control vector  $u$ , representing any external factors which influence how the state evolves over time. The predicted error covariance is given by  $P_t^{t-1} = \phi P_{t-1}^{t-1} \phi^T + Q$ , which represents any covariance in the random error that is associated with the previous state estimate during the filtering process. The  $Q$  added to the equation is simply the standard deviation of the random error of the state as it evolves to the current time point. Its inclusion here illustrates how the random error, or uncertainty, increases at the prediction stage. At the update stage, the Kalman gain is calculated, given by  $K_t = P_t^{t-1} A_t^T [A_t P_t^{t-1} A_t^T + R]^{-1}$ , which is essentially an optimal estimate for the magnitude of change in our estimate of  $\theta$ . Its value is based on the measurement matrix  $A$  which gives the extent to which the vector of measurements are influenced via the underlying state process, and by  $R$  which is the observation variance matrix illustrating the level of uncertainty in the measurements. The process of estimating the Kalman gain is for the purposes of determining how confident we are in the measurements reflecting the true state when updating the estimate, or whether its value should be more dependent on just the previous estimate  $\theta_{t-1}$ . In order to produce an updated estimate of the state  $\theta_t$ , the Kalman gain is then multiplied by the measurement residual in parentheses resulting in the expression

$$\theta_t^t = \theta_t^{t-1} + K_t(y_t - A_t\theta_t^{t-1} - \Gamma u_t).$$

The measurement residual is the difference between the observed measurements and the predicted measurements obtained from the product of the measurement matrix and the state prediction. The Kalman gain in this final expression therefore represents the correction to the estimate of the state. After this step, the error covariance is updated as  $P_t^t = [I - K_t A_t] P_t^{t-1}$ , illustrating how we are now more confident in our estimate of the state due to its associated random error being smaller. Essentially the process can be controlled by the four matrices from the DLM equations, those being the observation and system matrix coefficients, and the observation and system variance matrices. Using expert knowledge, the nature of how the state process is expected to evolve can be constrained, as well as how the relationship of the measurements with the state changes over time. Further, the likelihood of changes observed in the data being due to random error can be included, along with how volatile the behaviour of the state is expected to be. The Kalman filter is therefore incredibly useful in a variety of applications because of this flexibility. For instance, it has been often been applied to spatial problems whereby there are a series of measurements taken

from a range of locations and inferences are made as to the likely true state given the amount of variability expected in the predicted state estimate and in the actual point observed, as well as the complex dependencies involved in the relationship expected between the two.

For this project, the DLM equations and the Kalman filter will form the main framework of how this unique problem is addressed. The nature of the expert knowledge we are trying to encode into these models in order to obtain an accurate estimate of the state will be related to the intricacies and biases of the individual hygrometers in certain atmospheric conditions. The presence of instrument biases in certain conditions and how the state estimates can be adjusted accordingly will be explored during the analysis section. As has been discussed, some of the potential biases that can occur include the differing speeds to which each instrument responds after a rapid change in humidity, which can be significant. There may also appear to be over or undershooting in the measurements of certain hygrometers after these kinds of sharp changes, along with a potentially extensive stabilisation period required for the two chilled mirrors. In terms of this latter issue of oscillations which occur in particular conditions – primarily in the Buck measurements – this may be controlled through the use of the two variance matrices. The degree of uncertainty in the measurements can be controlled through the  $V_t$  parameter from the DLM equations. In periods where we expect the state to be more variable, this can be specified through the  $W_t$  term; if the variation in the state is expected to be less than that of the observations, the effect of reducing  $W_t$  serves to make the state estimate  $\theta_t$  more reliant relatively on its previous value  $\theta_{t-1}$  as was observed in the state equation. There are however frequent occasions whereby there is no precedent for a period of instability; we simply do not have an explanation for why a particular instrument is struggling. This is why our estimation of the state process should require reference to the information from each of the other hygrometers at any given time point and their individual biases. This challenge will be addressed towards the second half of the analysis section, where the various methods of combining information from differing sources which are attempting to measure the same process will be outlined. Some of the ways that this may be achieved which are consistent with the Bayesian approach will be discussed, starting in the next section with the principles of Bayesian melding.

## 2.5 Melding

The presence of multiple sources of evidence can occur in a variety of settings, but primarily is concerned with the issue of measurement inconsistency. For instance, this can be due to instrument biases or spatial variation in the measurements taken. It can also arise from situations where there are multiple experts with differing opinions about a subject. This can then cause difficulties whereby there is a need to choose appropriate prior distributions or any model parameters in order to conduct Bayesian inference.

Bayesian melding (Poole and Raftery 2000) was developed to address the presence of multiple priors on an unknown quantity and was introduced as a method of making inference about the uncertainty of inputs and outputs of a deterministic model. In its simplest form, the deterministic model  $M$  takes the input parameters  $\Theta$  and relates them to our output parameters  $\phi$ , so that  $\phi = M(\Theta)$ . There will be some pre-model prior distributions for the input parameters  $p(\Theta)$  and also the output parameters  $p(\phi)$ , which are produced before the model is specified. Priors on the input parameters are represented as probability distributions and in the original literature, this model  $M$  effectively acts as a transformation of the input parameters, meaning that the output priors can be induced by the model to become another distribution  $p^*(\phi)$ . Most of the time this induced prior cannot be computed analytically, but generally here can illustrate the concept. These two priors on the outputs,  $p(\phi)$  and  $p^*(\phi)$  are therefore separate distributions on the same quantity but are essentially based on different information. These priors would therefore often be ‘incoherent’, so the next step in the process is to find the common ground that can form the basis of the final Bayesian inference. This is where the idea of ‘melding’ two priors on one quantity is introduced, and there are various means by which this can be achieved depending on the requirements of the individual case. The main application from Poole and Raftery (2000) was a population dynamics model for bowhead whales, and this induced output prior  $\tilde{p}(\phi)$  was produced using just a geometric mean,

$$\tilde{p}(\phi) \propto p(\phi)^{0.5} p^*(\phi)^{0.5}.$$

This method would therefore imply that the information from both output priors were assumed to be equally reliable. There are a number of alternatives that will be appropriate depending on the circumstances, which will be discussed shortly. As summarised by Howes (2019), once this melded output prior has been computed the melded input prior can be converted back to input space as

$$\begin{aligned} \tilde{p}(\theta) &= \tilde{p}(M(\theta)) \left( \frac{p(\theta)}{p^*(M(\theta))} \right) \\ &= p(\theta) \left( \frac{p(M(\theta))}{p^*(M(\theta))} \right)^{1-\alpha}, \end{aligned}$$

where the final expression represents the ratio of the output prior and model induced output prior in terms of the input parameters, which weights the initial input prior according to the term  $\alpha$ . Some of the assumptions outlined here are not straightforward to apply in our case, as there is a deal of uncertainty in terms of how the model will be defined and where the model induced priors fit in to the DLM framework. Its principles however are useful and have great relevance to distinguishing and combining information from a variety of sources. The end goal is to produce melded posterior distributions for the true state of absolute humidity at each point in a time series using inference from each hygrometer  $i$ , as  $p(\theta_t | y_{1_i}, \dots, y_{t_i})$ . This is in line with the principles of Bayesian melding, and its applications to the subject of this dissertation will form the basis for the second half of the analysis chapter in which inconsistent probability distributions will be combined by similar means. In the next section, the ways in which information on an unknown quantity can

be combined from multiple sources will be discussed.

## 2.6 Opinion pooling

The focus of the Bayesian approach in general is to combine any pre-existing knowledge of the probability of obtaining a certain unknown quantity, or an event occurring, with inference that can be made from the data. This is based on determining the parameter settings of a model that will have the greatest chance of producing the observed data. In order to produce the most accurate inference about the true value, it is therefore important to ensure that any relevant prior information that is available be incorporated into forming this posterior distribution. It is often the case however that there is a range of contradictory information available, each of which could be justified as an appropriate prior distribution. There is then uncertainty as to which of these sources of information is likely to be the most accurate, the choice potentially having a significant effect on the resultant posterior distribution. For instance, often the opinions of experts form the basis for choosing any prior distributions, but there may be disagreement among other experts as to the probability of an event occurring or what the true value of a quantity should be at a given time. In terms of forming a model for how an underlying state process evolves over time, there is often uncertainty about the nature of this relationship between the observations and the state. What has already been discussed is how the recursive Kalman filter algorithm can produce estimates for the true state of a dynamic process by following the principles of Bayes' theorem to compute  $p(\theta_t|y_1, \dots, y_{t-1})$ . An inherent limitation of this approach here is that the estimates of the state are only dependent on the values from one set of measurements, thereby taking just one source of information to form inference about the state where there may be others that are directly following the evolution of the same process. For this project, we are presented with four different hygrometers that each have been shown to exhibit unique biases and tendencies. The humidity readings that they produce then are all directly influenced by the evolution of the same state process  $\theta_t$ . Therefore, when applying the Kalman filter algorithm to each of these instruments, they are not likely to produce the same estimates of the state over time. A crucial aspect of forming estimates of  $\theta$  in this project will be to consider the measurements of each of the other hygrometers at every time point, in order to update the individual estimates of  $\theta_t$  in light of this other relevant information. The separate posteriors which are formed for each of the four sets of measurements are therefore often inconsistent, so the process for finding the 'common ground' of the priors in Bayesian melding can be applied here to these four estimates of  $p(\theta_t|y_t)$ . This will be shown in full context to the project and in more detail during the analysis section in Chapter 3.

Opinion pooling is a process that relates to eliciting probability distributions from a set of experts and combining these to form a single suitable prior distribution that reflects the information that



can appropriately be compared to what is observed in the data to form an updated estimate of  $\theta$ . Considering the case where each expert provides a straightforward Gaussian probability distribution, there will be some estimate of the mean as the most likely outcome for  $\theta$ , along with a standard deviation  $\sigma$  which represents the range of values that are deemed plausible, or essentially the confidence in the given estimate of the mean. As will be discussed in Chapter 3, these two metrics will prove central in determining how each of the hygrometers will influence the final estimate of the state process over time. In the context of opinion pooling, each of the four hygrometers will therefore act as separate experts that are basing their judgements on different information and exhibit differing levels of uncertainty in certain atmospheric conditions. Referring back to the DLM equation formulations that will provide the basis for specifying the evolution of each instrument’s individual relationship with the state process, the final posterior estimates will be in the form  $p(\theta_t|y_1, \dots, y_t) \sim N(m_t, C_t)$  (West and Harrison 1997), where  $C_t$  represents the uncertainty of these estimates. There are a number of possible approaches to opinion pooling which can be chosen based on the judgment given by a proverbial decision maker that facilitates this combination of opinions.

If we were to take a wholly Bayesian approach to incorporating information from multiple sources to form a final collective probability distribution, there would have to be detailed knowledge of the extent to which a certain expert’s opinion changes as a result of hearing the probability distributions given by other experts. Under this framework, separate likelihoods would have to be formed as  $p(D|\theta)$ , essentially treating each new distribution as data so that  $D = f_1(\theta), \dots, f_n(\theta)$ . This implies that each expert should have some expectation of what the other experts will say, which will be then used to update their own estimate of  $\theta$ . As O’Hagan *et al.* (2006) have outlined, this is difficult to implement – especially where there is more than one expert – and this final collective distribution can be achieved more effectively through the addition of the beliefs of a single decision maker. They bring attention to the most common methods of opinion pooling: linear and logarithmic pooling. A single ‘consensus distribution’ can be formed most simply through linear opinion pooling as

$$f(\theta) \propto \sum_{i=1}^n \omega_i f_i(\theta).$$

The weighting parameters  $\omega_i$  associated with the separate distributions naturally must sum to one and each can be defined as equal if there is no reason to value one opinion any higher than another. In such a situation, this would essentially just be a weighted average of the probability distributions. As with the DLM terms from before, these weighting terms can be specified with no time-varying aspect if there is no basis to believe that any of the opinions given are more or less reliable in different conditions; for the purposes of this project however, the weightings can effectively be specified as a function dependent on various time-dependent factors if desired. In dealing with instrument reliability in hygrometry we are not likely to be faced with a situation where there are two optimal estimates of ‘true’ humidity at a given point, which as O’Hagan *et al.* (2006) describe

would be problematic. For instance, a reasonable outcome in which there are two instruments giving separate, but equally plausible, readings at a given point is that the ‘true’ humidity will lie somewhere between. There will generally always be some basis to deciding which hygrometer is more believable, based on detailed knowledge we have about the nature of the atmosphere at that point and expert knowledge about which instrument will likely be the most reliable in this situation; the distance between two measurements deemed equally plausible is effectively bounded.

An alternative to linear opinion pooling is that of logarithmic opinion pooling, given by

$$f(\theta) \propto \prod_{i=1}^n f_i(\theta)^{\omega_i}.$$

The weighting parameter can be treated as from before, but now the distributions are instead combined through taking a geometric mean. In comparison to linear pooling which can be seen as more uncertain, this method generally produces more contracted pooled distributions by essentially not giving as much reverence to the opinions of dissenting experts. An example of these two methods applied to Gaussian probability distributions is given in Figure 2, illustrating this increased confidence of logarithmic pooling in producing a distribution which represents the aggregated beliefs of three differing opinions.

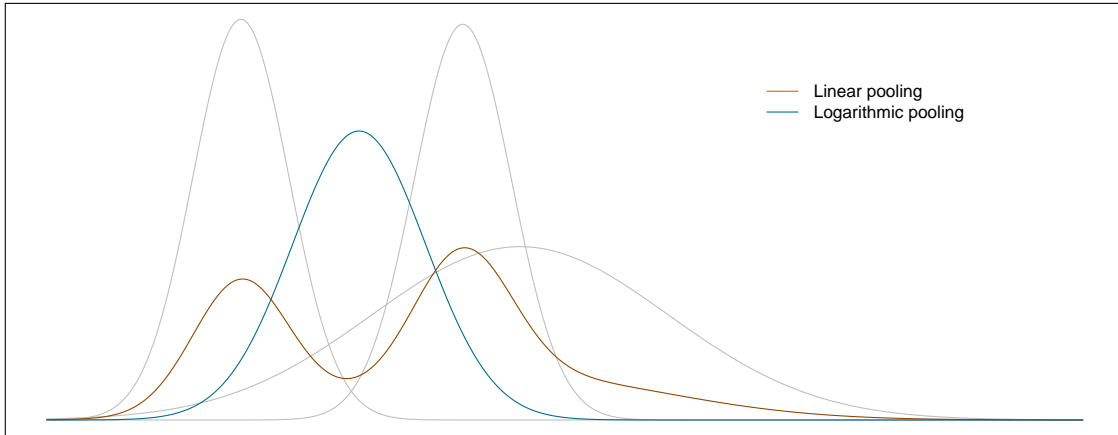


Figure 2: Methods of Opinion Pooling

In essence, logarithmic pooling can be seen to value the group of opinions that are closest in agreement and linear pooling rather values unique opinions. There is therefore a degree of flexibility when selecting a method of opinion pooling and each can have more relevance in certain situations.

For instance, where a decision maker wishes to combine prior information from a variety of sources and there are two very similar probability distributions present, they may not wish to take both into account in the combination as they could both be relying on similar information to form their opinion. It would follow then that logarithmic pooling is more appropriate where the distributions are produced using separate sources of information, as will be the case for this project. The implication in this situation where each expert is seen to be equally reliable is that the weighting parameter would be redundant. For the purposes of the analysis in this report, we will be generally concerned with identifying where there is most agreement in the measurements at each time point and disregarding any unreliable estimates. As will be discussed, logarithmic opinion pooling was found to provide the most appropriate results, and the flexibility of the weighting parameter noted to be crucial in indicating where an instrument's measurements should be given more reverence than what is indicated by how closely it conforms to the most commonly held opinion. A downside to this method however is that if one distribution is deemed implausible by any of the others – indicated by no overlap in their tails – then it will be completely disregarded, irrespective of its weighting. This was largely not covered in the analysis section but represents an area for further development in terms of potentially altering the method of pooling in certain situations. The method overall provides a beneficial level of flexibility because, whilst filtering forms the basis for how the underlying process will be learnt through this ability to include increasingly complex dependencies relating to levels of uncertainty and bias in various atmospheric conditions, the weighting parameter in particular gives us the ability to directly control which estimate of  $p(\theta_t|y_t)$  is unreliable in a given period with minimal adjustments to the model. What has been discussed in this section are the various approaches that can be taken in regard to the mathematical aggregation of probability distributions elicited from multiple expert opinions. Details of how these techniques will be applied to the subject of this dissertation will be discussed through the latter half of the next chapter. Initially we will focus on how the DLM framework and forward filtering algorithm can form the basis for addressing this problem.

## Chapter 3

### 3.1 Framing the problem

The data which are at our disposal originate from a variety of FAAM research flights across different climates and which were exposed to a range of atmospheric conditions during their course. Again, any time series figures presented in this report will be on a log scale for readability and to emphasise trends or periods of instability. For the purposes of this project we are presented with two main research flights, the first of which (C007) is out of Dakar, mostly exhibiting a good level of agreement between the hygrometers during its course. The second flight (B878) as was shown in Figure 1 is off the east coast of Scotland, and generally illustrates many of the issues that are encountered in certain conditions by each of the instruments. The various types of biases and tendencies to produce periods of unstable measurements were discussed in Section 2.2 and individually represent unique challenges in terms of identifying where these biases are likely to occur and encoding this structure into a model which can reliably produce estimates of the ‘true’ underlying humidity from the measurements taken from each hygrometer in a given flight. The approach taken in this dissertation will be to utilise the DLM framework to create a structure of dependencies in the DLM coefficients and variance terms which can influence how the estimates are learnt through time, given the different time-dependent atmospheric variables available to us and the readings of each of the other hygrometers. The details have been discussed of how the Kalman filter algorithm produces estimates of  $p(\theta_t|y_1, \dots, y_t)$  using expectations for the state evolution and the nature of the influence of the state on the observations, along with their expected variability. At each time point, four estimates of the ‘true’ humidity given the measurements observed will then be produced, representing each of the hygrometers. As the estimates produced will be for the same quantity, we would hope for consistency if each of the components and dependencies included in the DLM framework are ‘correct’. Due to the nature of the uncertainty associated with each of these instruments and our own uncertainty in the state evolution, these will not be entirely consistent. In past literature – and in relation to the discussion about combining evidence from multiple sources in Section 2.5 – inconsistent probability distributions can be effectively rectified through the appropriate application of opinion pooling. Once we have the four posterior distributions  $p(\theta_{t_i}|y_{1_i}, \dots, y_{t_i})$  for each hygrometer  $i$ , logarithmic pooling can be applied to reach a final estimate of the ‘true’ humidity process. The weighting parameter can be defined similarly to the DLM coefficients and variance terms, but also can allow straightforward adjustments to be made to the final estimate if for instance one of the instrument estimates is exhibiting a trend that we conclude is likely indicative of some instability. One of the limitations inherent in the filtering algorithm that will be applied here is that it achieves its estimates sequentially; only the previous time point has influence on the filtered state estimate at the current time point, meaning that periods where

there may be trends indicating instability cannot easily be identified. This pooling step has further importance when considering that the estimates of the other instruments at a given point will give an indication of whether that estimate is likely reliable; with no alterations to weightings, logarithmic pooling will have this effect of producing a pooled estimate closer to the majority opinion, also discounting estimates that are deemed implausible by the other posteriors. The parameters of the DLM equations can alternatively be conditioned on certain values of the other hygrometer measurements, such as the instrument variance  $V_t$  where there is a large discrepancy. During this analysis, other atmospheric variables and time-dependent factors such as pressure, temperature, humidity and mirror temperatures were primarily used to condition these parameters. The details of these factors and their inclusion in the model structure will be discussed in the next section.

### 3.2 Application of the DLM framework

As was presented in Section 2.4, the DLM equations have been widely applied in a range of practical settings due to their flexibility in producing state estimates which can be controlled in terms of the way the state is expected to evolve, the relationship of the measurements to the state over time, and the expected variability of the two processes over time. The benefit of this framework in a Bayesian setting is that expert opinion can be included in a variety of forms.

The forward filtering code used in this report will follow West and Harrison (1997) to initially produce estimates of the state at the prediction stage so that  $p(\theta_t|y_{t-1}) \sim N(a_t, R_t)$  where  $a_t = G_t m_{t-1}$  and  $R_t = G_t C_{t-1} G'_t + W_t$ . The following one-step-ahead prediction can then be made to reach  $p(y_t|y_{t-1}) = N(f_t, Q_t)$  where  $f_t = F'_t a_t$  and  $Q_t = F'_t R_t F_t + V_t$ . These identities are used to finally produce estimates of the state posterior as  $p(\theta_t|y_t) \sim N(m_t, C_t)$ , where  $m_t = a_t + A_t e_t$  and  $C_t = R_t - A_t A'_t Q_t$ , where  $A_t = R_t F_t / Q_t$  and  $e_t = y_t - f_t$ . The inputs to the function would therefore be the four DLM matrices, the data vector and initial values for  $m$  and  $C$  to start the recursive process taking the prior  $N(m_0, C_0)$ . The intuition behind each of these steps were outlined in Section 2.4 and this is the form of Kalman filter notation that will be therefore used throughout this analysis chapter.

Considering an initial example of a filter with none of this expert knowledge included, this would essentially involve defining the measurements as equal to the state plus some generic level of random error associated with that particular instrument:  $y_t = \theta_t + \epsilon_t$ . The desire in this type of situation would be to estimate an underlying process from a set of ‘noisy’ measurements. To illustrate why we would wish to specify a more appropriate complex variance structure to the instrument variance, this simple example of a filter with no expert knowledge incorporated is included in Figure 3,

extracted from a period of flight C007.

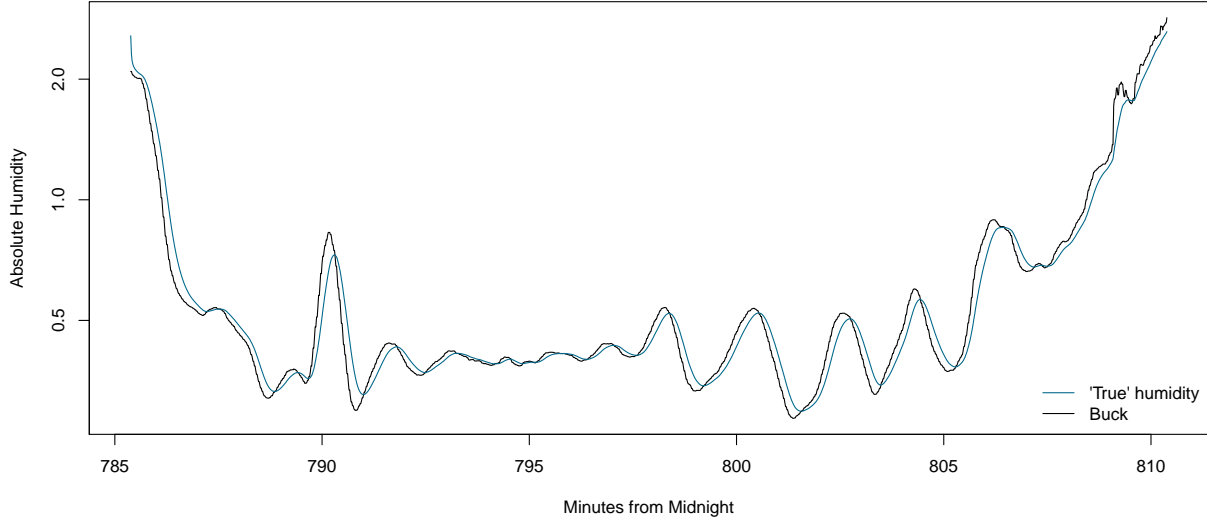


Figure 3: Filtering without the incorporation of expert knowledge

The estimates of the state at each time point essentially take information from previous time points to act as priors which are then updated through the filtering algorithm, so in this sense even this simple filter follows the Bayesian approach. This is why the effect is observed of the estimates seeming to follow the values at previous time lags. Under this approach however, it follows that inference is made strongest through the inclusion of all appropriate prior information that is available, which this simple filter does not achieve. As was discussed in Section 2.2, the chilled mirror hygrometers generally struggle at low humidities and the Buck in particular has a tendency to oscillate following rapid changes in conditions, which is seen here. To start considering the means by which we can deal with such intricacies in the DLM framework to produce stronger inference, we will take this example of the Buck during a particularly dry period of flight C007. Incidentally, this is just below the level of humidity (around  $0.6\text{g/m}^3$ ) where the WVSS-IIs also have a greater tendency to struggle and where instrument instability is more likely to occur.

Generally this situation presented illustrates a period of increased uncertainty in the Buck measurements after a sharp drop in humidity, or once humidity reaches a level below which we would not be as confident that the measurements are accurate. These are two instances that can be identified so that the instrument variance  $V_t$  can be adjusted to be higher in comparison to the state variance  $W_t$ . The mirror temperatures for the Buck and GE can be useful in this regard to indicate where instability might be more likely to occur; the GE in particular has been found to be highly dependent on ambient temperature (Vance *et al.* 2015). This will however be of use

primarily when dealing with instrument bias correction. Whilst it may be challenging to encode the conditions necessary for implying higher uncertainty, in terms of the sharpness of this change in conditions, once the necessary adjustment to  $V_t$  has been defined the modifications to the model may still be minimal. For the purposes of this next example, the ambient pressure was used to condition the instrument variance. For a particular flight, the instrument or state variance can be specified to increase or decrease with decreasing pressure, or a particular threshold chosen below which there is expected to be more uncertainty. Figure 4 illustrates the state estimate given the Buck measurements during this same period, controlling the state variance to illustrate this effect more clearly. Once ambient pressure reaches a sufficiently low level – indicated with further reference to flight B878 but which can be specified as seen appropriate – the state is expected to be more closely related to its past value  $\theta_{t-1}$  than the instrument measurements, which has this effect observed of discounting the instrument variation.

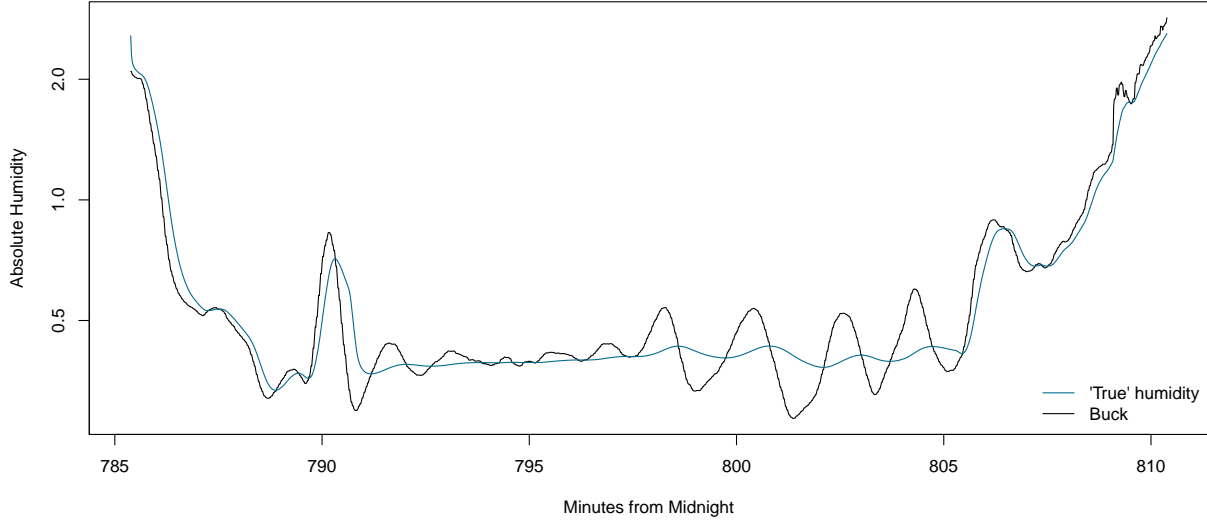


Figure 4: State variance conditioned on ambient pressure

The presence of an uncertainty structure in instruments measuring an underlying process has formed the basis for why the Kalman filter has been widely applied in a variety of settings. An educated estimate of the ‘true’ humidity at each time point can be made based on a variety of factors, but the nature of the state process is that it is unmeasurable so there will always be some level of uncertainty associated with an estimate. Visualising this uncertainty – which is influenced by both  $V_t$  and  $W_t$  – through the inclusion of credible intervals will be an important step and will be discussed more during the ‘pooling’ stage.

The DLM framework however further provides control in terms of how the relationship of the mea-

measurements to the state is expected to evolve over time. The measurements  $y_t$  are directly dependent on the state process  $\theta_t$ , and this dependence is therefore controlled through  $F_t$  from the observation equation. This relationship therefore can provide a useful tool for incorporating expert prior knowledge that will appropriately offset the way estimates are updated sequentially by the data. Instrument bias refers to occasions or periods in which certain conditions can cause instruments to produce systematically inaccurate estimates. The primary forms of bias that this  $F_t$  coefficient may be defined to counteract in the final state estimate are over/under-reading and instrument response rates, which were outlined in Section 2.2. These will be the main instrument biases that will be focussed on in this report, but any situation where it is known that the measurements are deviating from the state in a particular way can be addressed through similar means. Rather than specifying a general level of uncertainty in the measurements, as was controlled through the instrument and state variance, being able to define the tendency of the instruments to deviate from the state in a specific way will lead to much stronger inference. In reference to the research from Vance *et al.* (2015) discussed at the beginning of the report, one of the most notable instances of instrument bias occurs in the GE. There has been found to be a clear tendency for its measurements to progressively over-read as humidity decreases, especially below  $0.1\text{g/m}^3$  or a mirror temperature of 230K. Simply put, it follows that to produce an estimate of the state that corrects for instrument over-reading, the  $F_t$  coefficient can be set as higher than one. With reference again back to the observation equation from the DLM framework, this implies that the estimate of  $\theta_t$  produced will be lower than  $y_t$  when multiplied by a value of  $F_t$  greater than one. So to make simple corrections to over or under-reading, its magnitude and whether it is above or below one will influence how far below or above respectively the state estimate will be at that time point.

The other main form of bias that may be shown by the instruments is a slow response rate where the aircraft moves through to different conditions giving rise to a rapid change in humidity. The chilled mirror hygrometers are generally shown to respond slower to changes in conditions than the WVSS-IIs and this has been exhibited in the data, especially where this decrease is smooth and constant. The following example is from such a period of rapid change in conditions where corrections to general over-reading of the GE and instrument lag can be illustrated. Either through encoding in the model or from a retrospective overview of the flight, periods of smooth decreases in absolute humidity obtained from the GE can be identified or where there is expected to be some instrument lag. An available approach which can allow for slow response rates in the GE state estimate would be a form of quadratic or possibly exponential function applied to the  $F_t$  coefficient, through which the extent of lag correction can be controlled. This would have the desired effect of not being a simple over-reading correction, but effectively allowing the state estimate to decrease at an increasing rate up to a point, before slowly returning to the original proximity to  $y_t$  once the humidity stops decreasing. As illustrated in Figure 5, below this level of  $0.1\text{g/m}^3$  outlined previously there can be more of an over-reading correction allowed which has this effect of a lower state estimate in relation to the measurements. During the period of rapid but smooth decrease in



absolute humidity, a function of a similar form to that of an exponential was applied to  $F_t$  which has had the effect described. This was defined here as  $1/(e^{0.5x/n} - 0.5x/n)$  where  $n$  is the length of the period, and is converted simply to have appropriate upper and lower limits for  $F_t$ , generally having the effect of a downward concave curve that eventually starts to decrease faster back towards a value of one. Again, there is flexibility depending on the expected nature of the instrument over or under-reading. Although it may be difficult to determine the correct level of lag correction or to reflect closely the estimates of  $\theta|y$  obtained from the other hygrometers, it will be shown that when choosing the weightings of each estimate during the latter part of the process, making such appropriate corrections through filtering will be of use in the presence of unreliability in the other instruments.

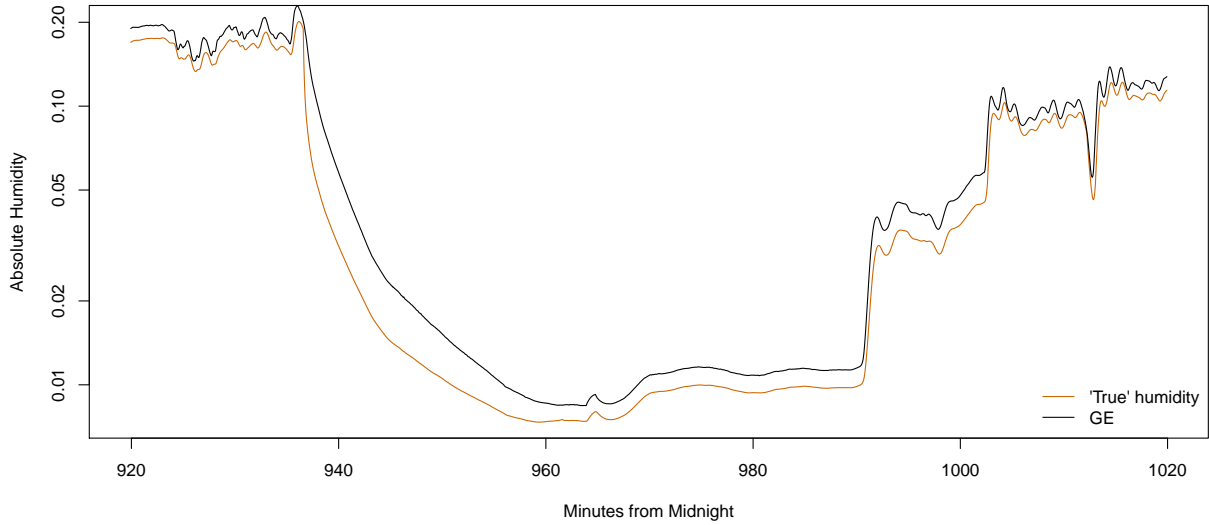


Figure 5: Allowing for instrument bias through F coefficient

There may be further instrument biases that could be addressed through similar approaches, such as the overshooting of the chilled mirrors when the aircraft moves through from dry to more humid conditions. The most likely approach might be to identify these periods where overshooting is expected to occur and set a limit to the resulting state estimate, which would not require an overly complex adjustment to the model. Further, the instrument variance may for instance be allowed to vary directly with an appropriate time dependent factor effectively acting as a proxy for time  $t$ . This may be a situation where there is a gradual increase in instrument uncertainty with pressure, temperature or humidity; an option would then be to define a function to apply to  $V_t$  and then convert the lower and upper value to an appropriate range. Over the course of this section, we have discussed a selection of instances of the kinds of instrument bias and uncertainty that typically can occur over the course of a flight, and how these may be accounted for to reach a state estimate which is more in line with expectations of the ‘true’ humidity process. There are numerous further

dependencies that may be included in the DLM parameters, such as from new research indicating certain conditions that can give rise to instrument uncertainty or bias. The purpose for this report will be essentially to introduce the tools that are at our disposal in terms of the DLM framework and Kalman filter, prompting further investigation into producing further layers to these models that can continue to produce more accurate state estimates and increase the range of conditions in which they can be effective.

### 3.3 Application of opinion pooling

The DLM framework provides a significant level of flexibility in producing state estimates that can take into account instrument uncertainty in a variety of conditions encountered, further allowing for any anticipated instrument biases in the inference of  $p(\theta_t|y_t)$ . Currently, a significant aspect of producing an estimate of ‘true’ humidity is through considering the information gained from other hygrometers at each time point, as this will give an indication to whether the instrument is producing an unreliable estimate. What will have been achieved thus far are four posterior estimates of  $p(\theta_t|y_t) \sim N(m_t, C_t)$  representing the four sets of measurements from each hygrometer, the individual model parameters specified to account for instrument uncertainty and bias. If the conditions of the models have each been set to be ‘correct’, we would hope for these state estimates to be generally consistent; the nature of uncertainty in our inference however implies that there will be a degree of inconsistency throughout. The task at this stage can therefore be framed as a problem concerning the combination of multiple sources of evidence, which was discussed in Section 2.5. The literature on evidence synthesis and more recently the Bayesian application to the problem with Bayesian synthesis (Raftery, Givens, and Zeh 1995) and later Bayesian melding (Poole and Raftery 2000) have presented a flexible approach to combining such information to form appropriate priors for inference. ‘Melding’ here refers to the formation of a single prior distribution from the pre-model and model-induced prior distributions, as was outlined earlier. Whilst Bayesian melding will not be directly applied here, this principle of forming a single probability distribution from a selection of inconsistent distributions – which give expectations for the likely values of the same quantity – will provide the basis for this section where a final estimate of ‘true’ humidity is produced. The disagreement that is present between the model-induced and pre-model priors is therefore akin to this inconsistency found in the four estimates of  $p(\theta|y_t)$  which have been formed. The method originally was applied when eliciting the separate opinions of multiple experts on the likely values of the same quantity or occurrence of an event. The different methods of ‘opinion pooling’ discussed reach this pooled distribution in different ways, each of which effectively seeking to establish common ground that may form the basis for the final posterior mean and standard deviation estimates. These alternative methods, namely linear and logarithmic opinion pooling, were compared during Section 2.6 in terms of the extent to which unique expert opinions, or the presence of agreement between experts, are valued.

Logarithmic pooling has been found to be the most appropriate here for dealing with unreliable estimates, given by  $f(\theta) \propto \prod_{i=1}^n f_i(\theta)^{\omega_i}$ , where for each hygrometer  $i$ ,  $f_i(\theta) \propto p(\theta_{t_i}|y_{1_i}, \dots, y_{t_i}) \sim N(m_{t_i}, C_{t_i})$ . The appearance of a distribution which is contrasting the most commonly held set of possible values for ‘true’ humidity given by the other hygrometers will indicate an unexpected discrepancy if the DLM parameters have been appropriately set. By the nature of logarithmic pooling, it produces more concentrated or confident pooled distributions as was illustrated by Figure 2. Further, for the purposes of this analysis the Gaussian nature of the pooled distributions obtained from logarithmic pooling is a necessary feature. An example of the logarithmic pooling that will occur at each time point is given in Figure 6, at around 836 minutes after midnight during the flight B878. We have a case here where the two WVSS-IIs are agreeing closely but the Buck for instance is exhibiting a significantly higher mean and standard deviation. The standard deviations  $C_t$  of the individual distributions are influential in reaching the final pooled distribution; the presence of an estimate with a larger standard deviation than the others, and which is weighted above zero, will have the result of a wider pooled distribution. The nature of logarithmic pooling again means that if one estimate is deemed implausible by any of the other distributions  $f_i(\theta)$  – indicated simply by any overlap in their respective tails – it is disregarded from the pooling process. This implies that all other thing being equal, a higher standard deviation will result in the distribution being less likely to be disregarded; it may not necessarily influence the mean estimate but will illustrate this greater uncertainty in the state when interpreting the credible intervals.

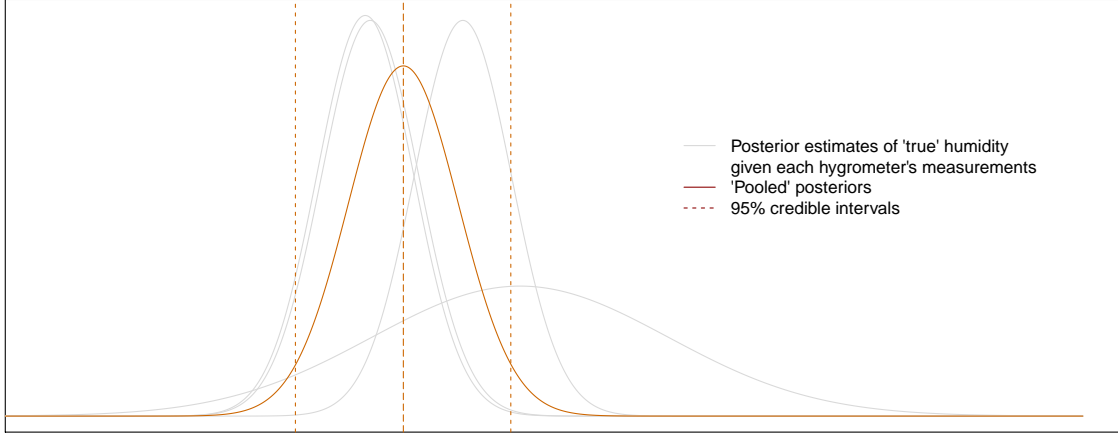


Figure 6: Example of ‘pooling’ that occurs at each time point with equal weighting

The primary advantage of this pooling stage in the final inference of the state is the flexibility given in terms of being able to control the weighting parameter  $\omega$  and allow this to vary over time similarly to the DLM terms. There are various instances in which a distribution has not been

disregarded during logarithmic pooling, but we will be confident that the estimate is not producing an accurate inference of ‘true’ humidity. The adjustments required to select a period in which we are confident the estimates are inaccurate – for instance based on observing an unreliable or uncharacteristic trend, based again on expert knowledge – are minimal, so will be an important final stage in the context of the analysis in this report. This effect of completely removing the weighting of such a distribution is given here in Figure 7. What is observed is that this higher variance in the Buck on the right now has no influence on the overall pooled distribution. The mean value generally has not changed, due mostly to the influence of the two WVSS-II estimates agreeing highly and the GE with its higher confidence indicated by a lower variance in expected values to the Buck. If each distribution had the same variance, it would clearly follow that the weighting parameter would have the effect of shifting the pooled distribution towards those given a higher weighting.

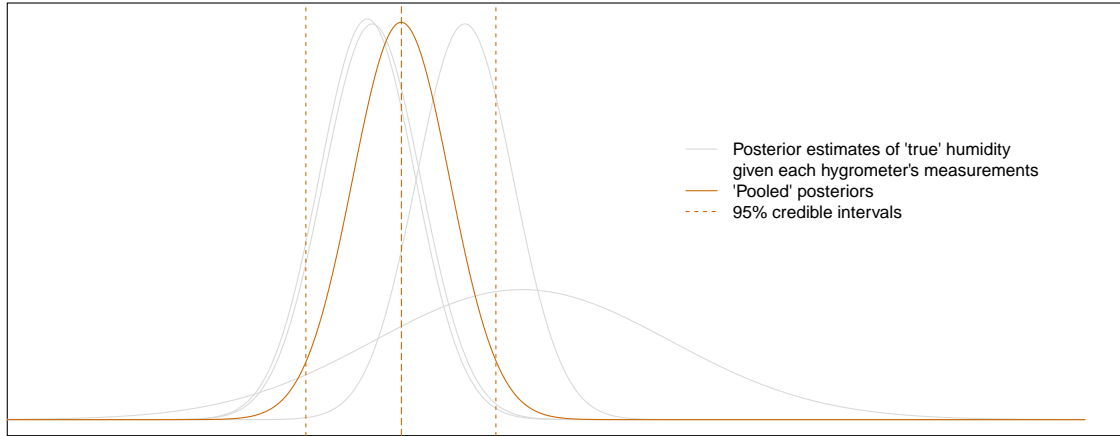


Figure 7: Example of ‘pooling’ that occurs at each time point with unequal weighting

As will be outlined further in the next section, the weighting parameter may be specified to vary through time by a similar means to the DLM coefficients and variance terms from Section 3.2. Rather than just acting to rectify any instrument intricacies that were unable to be accounted for through the DLM framework, we can also emphasise our own confidence in estimates during certain conditions, for instances similarly to where there is higher instrument variance. This may then serve to increase the likelihood of reaching an accurate estimate. In terms of incorporating a time-varying element to the weighting parameter in this analysis, it can primarily be set to decrease gradually with a greater discrepancy from the average of the other estimates. This is due to this limitation of logarithmic pooling not always completely discounting estimates when there is a material discrepancy. We will observe the effect this has on the final estimate in the next section,

where a final result can start to be visualised.

### 3.4 Reaching a final estimate

Once each distribution of  $p(\theta_t|y_t) \sim N(m_t, C_t)$  has been obtained through filtering with appropriately conditioned DLM parameters and logarithmic pooling applied, a final mean estimate of the ‘true’ humidity may be produced along with the credible intervals observed in Figure 6 to illustrate uncertainty in our estimates. As discussed, the four standard deviations  $C_t$  produced influence this final uncertainty, each of which are based on the expected level of instrument variance  $V_t$  in relation to the expected state variance  $W_t$  during certain conditions. The final result which can therefore be obtained will have the appearance of this time series in Figure 8, which represents the result of the process applied to flight B878 from Figure 1. Flight C007 has not been included for illustration here, generally because the instruments exhibited a high level of agreement for the vast majority of the flight and the ideas presented here would not have been clearly visible. In this flight, we have a selection of periods in which certain instruments are acting unreliably; the most notable case arises in the flush WVSS-II around 820 and 960 minutes after midnight, where its measurement drops dramatically and clearly not in line with expectations. To take this example, such a case is difficult to account for through the DLM framework without considering the measurements of other hygrometers at that time point. This is therefore a case whereby logarithmic pooling may be able to deal with such unexpected discrepancies, but in this example brings attention to the limitations of logarithmic pooling not being able to disregard these kinds of outlying values when they are dramatically different to the other estimates. This issue was overcome through specifying a time-varying element to the weighting parameter  $\omega$ , which allowed the weighting of an estimate to decrease gradually with increased discrepancy from the average of the others. In terms of the conditions used in the DLM coefficients, instrument variances were specified to increase from around  $0.6\text{g/m}^3$  and below, with state variance also expected to be lower above this limit. This instrument variance was further allowed to increase below mirror temperatures of 230K for the chilled mirror hygrometers. Lower humidity limits were further used for the WVSS-IIs but the weighting was found to be much more effective than specifying greater instrument variance during these periods in which they tended to struggle. There were further over-reading corrections made to the GE, along with lag corrections during the periods of decreasing humidity around 800 and 940 minutes from midnight. The GE was found to be a useful benchmark when considering the instability of the two WVSS-IIs; where there are definite reported tendencies for an instrument to exhibit over-reading during certain conditions or predictable instances of slow response rates, these can be accounted for effectively by utilising the DLM framework. This means that we can reasonably justify allocating a greater weighting to the GE where the weighting of the WVSS-IIs is allowed to decrease gradually during these periods of high discrepancy. The GE is consistent but may for instance not capture the intricate trends of the WVSS-IIs when they are operating

properly, so during unstable periods it may then act as just a useful benchmark where weightings have to be transferred. These are a selection of some of the adjustments made during the process which has resulted in the estimate generated in Figure 8, and this increased uncertainty has been illustrated at lower humidities. During the periods from around 800 to 840 and 940 to 990 minutes from midnight, the uncertainty intervals generally capture well the range of estimates where the instruments are not considered entirely unreliable.

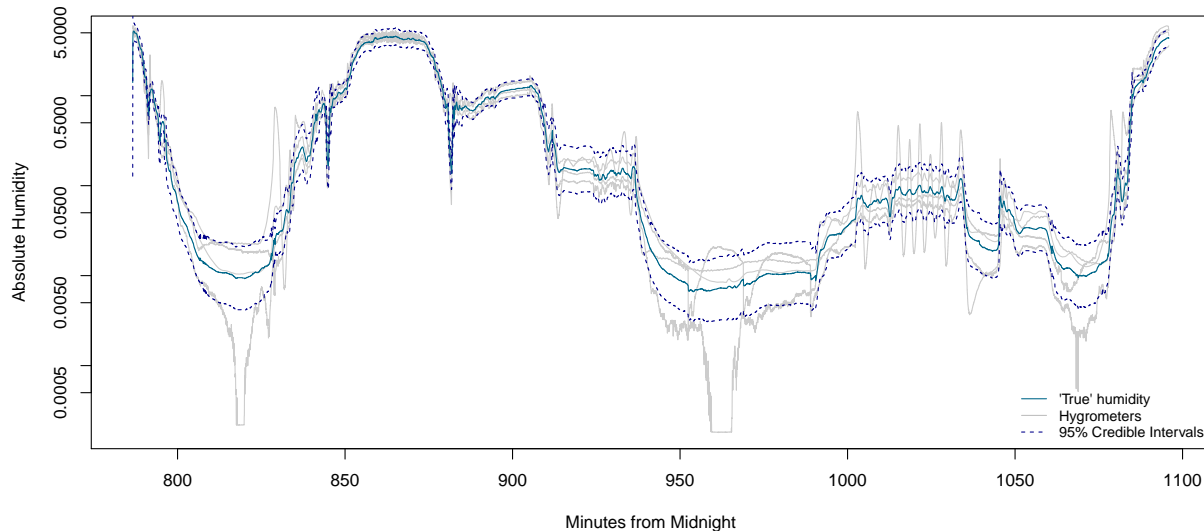


Figure 8: Final state estimate after 'pooling'

As has been emphasised during this analysis, there is still a deal of flexibility available and developments still to be made; there are still instances where a certain estimate is not necessarily unreliable but has not been captured by the credible intervals, such as occasions toward the start and end of the flight. Further layers of instrument bias and uncertainty can be continually included in the process which will serve to keep on improving the accuracy of estimates and the range of conditions taken into consideration. It is important to note that the user of the techniques presented over the course of this report will still have the ability to make simple but effective adjustments through the weighting parameter during the pooling process. There are occasions where there may be no indication of where a period of instability is likely to occur, and which cannot be identified through a recursive process like filtering that sequentially updates state estimates one time point forward. Such simple adjustments are illustrated in Figure 9, where two specific periods were chosen. As can be observed in Figure 8, there is a period between 950 and 1000 minutes from midnight where the Rosemount WVSS-II exhibits a trend that is clearly unrepresentative of true atmospheric conditions. This has not been entirely captured and eliminated through filtering or during the pooling process, so in such a case we have the option to remove its estimate's weighting entirely through a simple adjustment. Further, we have the Buck exhibiting a typical tendency to oscillate around

1000 to 1040 minutes from midnight. Whilst filtering may have reduced the intensity of such variability, we know that this is not a reliable trend so this may be corrected through similar means. This simple adjustment was made to these two periods with the Rosemount WVSS-II and the Buck; it can be observed in Figure 9 that these unreliable trends have been discounted and has therefore resulted in a more accurate final estimate.

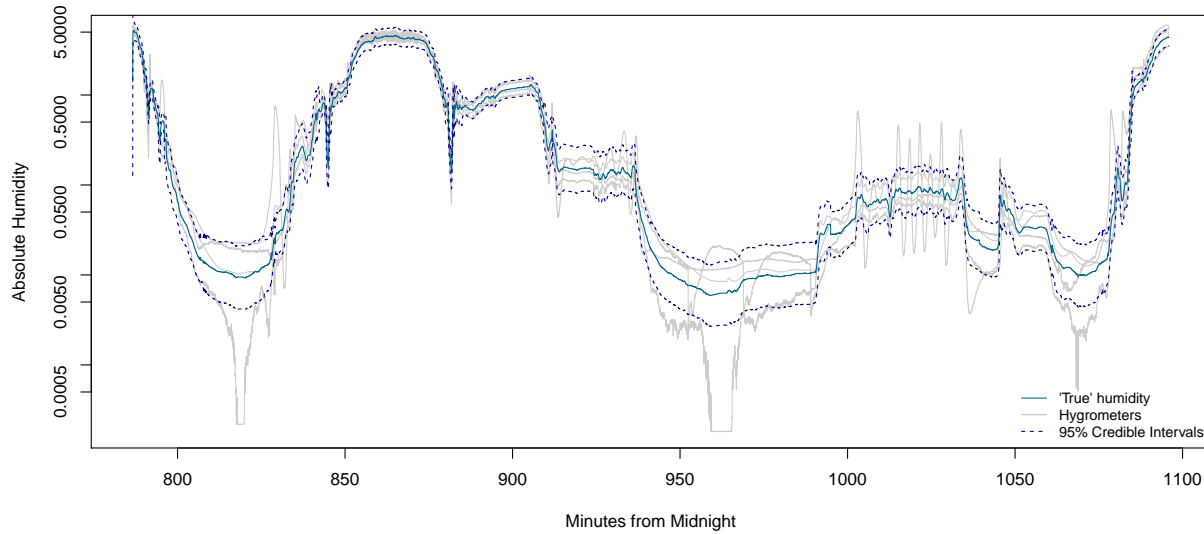


Figure 9: Weighting removed from Rosemount WVSS-II and Buck during periods of instability

Throughout the course of this chapter, the methods undertaken in order to reach these final estimates of ‘true’ humidity have been outlined and the various adjustments that may be made to account for uncertainty and bias introduced. Whilst the conditions and dependencies included in this analysis may not be exhaustive or have led to a final result that is representative of the most accurate estimate possible, the means by which such factors can be accounted for and how this research may develop further has been outlined.

## Chapter 4

### Discussion and further work

There are further areas of research that may be considered for the future which have potentially significant and wide-reaching implications in the estimation of an underlying ‘true’ humidity process. The most interesting consideration for future work may be the application of a machine learning process to aid in the identification of unreliable trends in humidity measurements which currently would require retrospective adjustments to be made where these trends occur over a significant period of a particular flight or are not always of a similar nature each time they occur. For the purposes of the analysis in this report, we have further only been concerned with distributions which are Gaussian in nature with a mean and standard deviation structure. Taking a central estimate of such as distribution is acceptable in producing a mean value with uncertainty represented by its standard deviation, but the reality may be that there are expected to be a wider range of potential values for ‘true’ humidity below the mean value than above. For instance, there may be a range of possible values generated but this does not necessarily imply that the ‘true’ value is at the centre of that distribution. A particle filter may address this problem, whereby we can base our conclusion about where in the distribution our ‘true’ value should be on other factors or variables we have available at that time point that may give some indication, along with what is known about the nature of the hygrometer.

The purpose of this report has been to introduce and outline the various techniques and tools at our disposal in terms of being able to account for the wide-ranging effects that differing atmospheric conditions have on the reliability of the four hygrometers at the focus of this analysis. The Bayesian approach has provided the basis for encoding expert knowledge and research into the inference of  $\theta$  in regard to the expected evolution of the state process as it relates to the measurements taken by each of the instruments and the uncertainty or variability of each instrument. As has been discussed, the primary benefit of the Bayesian approach here is its implication that, as the extent of relevant prior information in the form of expert knowledge is continually improved and developed, the inference made about the ‘true’ humidity after considering the available data is greatly strengthened. It follows then that this analysis should be used to spur on further development and the continual addition of further layers of instrument bias and uncertainty, which will only serve to increase the accuracy of inference and the range of conditions in which we can appropriately apply this framework. The various atmospheric variables and dependencies considered during Chapter 3 therefore represent a selection of those that are available to be researched further.



## References

- Fleming R. J. and May R. D. (2004). *The 2nd Generation Water Vapor Sensing System and Benefits of Its Use on Commercial Aircraft for Air Carriers and Society*. UCAR, Boulder, CO, available at: <https://www.eol.ucar.edu/system/files/spectrasensors.pdf>.
- Howes A. (2019). *Markov Melding*. University of Warwick, available at: <https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/pollock/former/howes.pdf>.
- Kim Y., Bang H. (2018). *Introduction to Kalman Filter and Its Applications*. IntechOpen, available at: <https://www.intechopen.com/books/introduction-and-implementations-of-the-kalman-filter/introduction-to-kalman-filter-and-its-applications>.
- O'Hagan A, Buck C.E., Daneshkhah A, Eiser J.R., Garthwaite P.H., Jenkinson D.J., Oakley J.E., and Rakow T. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons. ISBN: 0-470-02999-4.
- Petris G., Petrone S., Campagnoli P. (2009). *Dynamic Linear Models with R*. Springer-Verlag, New York, DOI: 10.1007/b135794.
- Poole D and Raftery A.E. (2000). *Inference for Deterministic Simulation Models: The Bayesian Melding Approach*. Journal of the American Statistical Association 95 (452). Taylor & Francis Group: 1244–55.
- Raftery A.E., Givens G.H., and Zeh J.E. (1995). *Inference from a Deterministic Population Dynamics Model for Bowhead Whales*. Journal of the American Statistical Association 90 (430). Taylor & Francis Group: 402–16.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sansom P. G., Williamson D. B. and Stephenson D. B. (2019). *State space models for non-stationary intermittently coupled systems*. Journal of the Royal Statistical Society: Series C, in revision, available at: <https://arxiv.org/abs/1711.04135>.
- Vance A. K., Abel S. J., Cotton R. J., and Woolley A. M. (2015). *Performance of WVSS-II hygrometers on the FAAM research aircraft*. Atmos. Meas. Tech., 8, 1617–1625, available at: <https://amt.copernicus.org/articles/8/1617/2015/>.
- Vance A.K., Price H., Woolley A., Szpek K. (2018). *Update on the Performance of the WVSS-II Hygrometers Fitted to the FAAM BAe 146*. OBR Technical Note 93, Met Office, DOI: 10.13140/RG.2.2.31975.60320
- West M., Harrison J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag New York, DOI: 10.1007/b98971.