Quality Control in Weather Monitoring with Dynamic Linear Models

Joel Janek Dabrowski*, Ashfaqur Rahman, Ming Li, Quanxi Shao, Shuvo Bakar, Andrea Powell, Brent Henderson

Data61, CSIRO, Australia
firstname.lastname@data61.csiro.au

Abstract

Decisions in agriculture are frequently based on weather. With an increase in the availability and affordability of off-the-shelf weather stations, farmers able to acquire localised weather information. However, with uncertainty in the sensor and installation quality, farmers are at risk of making poor decisions based on incorrect data. We present an automated approach to perform quality control on weather sensors. Our approach uses time-series modelling and data fusion with Bayesian principles to provide predictions with uncertainty quantification. These predictions and uncertainty are used to estimate the validity of a sensor observation. We test on temperature, wind, and humidity data and achieve error hit rates above 80% and false negative rates below 11%.

Introduction

Making decisions based on weather is common practice in agriculture. Government agencies such as the Australian Bureau of Meteorology (BoM) install weather stations across Australia to monitor weather conditions. The data generated from these stations can be used for forecasting extreme heat, snow, frost, and heavy rainfall, which directly affect farmers. Long-term weather forecasts can also be used understand climatic changes and it's affects.

While the BoM weather stations provide sufficient geographical coverage for large-scale weather modelling, the coverage can be insufficient for localised weather modelling. For example, temperature, rainfall, and wind can be highly localised to small geographic regions. Furthermore, BoM weather stations are often located to provide coverage of urban regions, whereas farms are typically located in remote regions. Farmers thus often install off-the-shelf weather stations on their farms; which we refer to as Third Party Weather Stations (TPWS).

Unlike BoM weather stations, TPWS data may not be quality controlled, resulting in uncertainty on the reliability and correctness of the data. For example, low cost sensors may be inaccurate or the weather stations may be incorrectly installed. Incorrect data may cause a farmer take unnecessary costly actions or fail to take actions leading to disastrous consequences.

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

We present an automated approach to detect unusual readings from TPWS data. The approach uses the Dynamic Linear Model (DLM) and the Kalman filter. It performs time series analysis and data fusion using Bayesian principles to provide predictions with uncertainty quantification. A key contribution is that we provide a means to automatically calibrate model. We show that this approach is robust in that the same model, parameters, and hyper-parameters are used across various weather stations and variables to provide accurate results with error hit rates above 80% and false positive rates below 11%. This work is especially relevant to those who seek a solution which is computationally efficient, easy to implement, and easy to maintain.

Related Work

Quality control for weather sensing has been an active research topic for several years (Leung, Mateo, and Nadler 2007; Hubbard et al. 2007; Mateo and Leung 2008) and a recent surge has come about in relation to the so-called "citizen" or "crowdsourced" weather stations (Beele et al. 2022; Meier et al. 2017; Napoly et al. 2018). The common approaches are heuristic, statistical, or regression based. Heuristic approaches include thresholds, duplicate checks, and step-change tests. Statistical approaches compare samples to the mean and standard deviations of cyclic data and previous extrema. Regression approaches compare samples to neighbouring station interpolations, previous forecasts, and daily observations. Unlike the proposed approach, none of these approaches consider sub-daily dynamics of the variables in a time-series sense. The more recent approaches tend to take advantage of the high spatial concentration of "crowdsourced" weather stations in urban environments. High concentrations are however not likely in an agricultural setting.

The DLM (also called the linear dynamic system, linear Gaussian state-space model, or Kalman filter) is known for its ability for state estimation amongst noise (Murphy 2012). It is also well equipped to perform outlier detection in time-series (e.g. (Ting, Theodorou, and Schaal 2007; Zhang and Liu 2009)). Furthermore, Kalman filters are well-known amongst the meteorological community (usually as Ensemble Kalman filters for NWP) (Roh et al. 2013; Evensen 2009). However, little work has been done in developing it for quality control of weather sensors.

Methodology

A DLM is employed to make predictions of a given weather variable over time. It fuses data from a TPWS and from a Numerical Weather Prediction (NWP) product, and also models the dynamics of the data over time. The Kalman filter is used to perform inference in the DLM by updating the weather variable predictions at each time given past data. A weather variable prediction is the form of a Gaussian distribution. If a given test observation is an outlier relative to the predicted Gaussian, it is classified as a suspect sample.

The DLM is automatically calibrated using an initial set of data with length T. This calibration is based on a noise estimate of this initial dataset.

Dynamic Linear Model

The DLM is a linear-Gaussian state-space model (West and Harrison 2006; Murphy 2012). It comprises a latent variable $\boldsymbol{h}_t \in \mathbb{R}^{H\times 1}$ and an observed variable $\boldsymbol{v}_t \in \mathbb{R}^{V\times 1}$. In this study, $\boldsymbol{h}_t$ represents modelled time-series components the particular weather variable of interest and $\boldsymbol{v}_t$ represents hourly TPWS observations and NWP data. The state-space equations are

$$\boldsymbol{h}_t = \boldsymbol{A}\boldsymbol{h}_{t-1} + \boldsymbol{\eta}_h \tag{1}$$

$$\boldsymbol{v}_t = \boldsymbol{B}\boldsymbol{h}_t + \boldsymbol{\eta}_v \tag{2}$$

where $\boldsymbol{\eta}_h$ and $\boldsymbol{\eta}_v$ are zero mean Gaussian distributions with covariances $\boldsymbol{\Sigma}_h \in \mathbb{R}^{H\times H}$ and $\boldsymbol{\Sigma}_v \in \mathbb{R}^{V\times V}$ respectively. The matrix $\boldsymbol{A} \in \mathbb{R}^{H\times H}$ is the state transition matrix and $\boldsymbol{B} \in \mathbb{R}^{V\times H}$ is the measurement matrix.

The data is assumed to comprise an offset and a diurnal periodic component with a period of $\tau = 24$ hours. The DLM models this by combining a local-level and a free-form seasonal model (West and Harrison 2006) to produce the state transition matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & -1 & \cdots & -1 & -1 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{3}$$

Given τ and the structure of $\boldsymbol{A}$, the latent dimension is $H = 24$.

The DLM naturally performs data fusion when the observed variable comprises multiple data sources measuring the same latent variable. With the TPWS and NWP data $V = 2$, the observed variable is $\boldsymbol{v}_t = [x_t \quad y_t]^\top$, and

$$\boldsymbol{B} = \begin{bmatrix} 1 & 1 & 0\cdots & 0 \\ 1 & 1 & 0\cdots & 0 \end{bmatrix} \tag{4}$$

$\boldsymbol{\Sigma}_v$ represents the covariance of the data noise and determines how sensitive the model is to a particular data source. With an estimate of the data noise for the TPWS and NWP data given by $\boldsymbol{\epsilon} = [\epsilon_x \quad \epsilon_y]^\top$, the covariance matrix is

$$\boldsymbol{\Sigma}_v = \begin{bmatrix} \epsilon_x^2 & 0 \\ 0 & \epsilon_y^2 \end{bmatrix} \tag{5}$$

We also consider a model that excludes the NWP data. For this, $V = 1$ and the second dimension is removed in $\boldsymbol{v}_t$, $\boldsymbol{B}$, and $\boldsymbol{\Sigma}_v$.

Setting the values of $\boldsymbol{\Sigma}_h$ generally lower than $\boldsymbol{\Sigma}_v$ assumes that the data is more uncertain than the model's dynamics contained in (1) [1].We thus set

$$\boldsymbol{\Sigma}_h = 0.8\boldsymbol{I}\min(\epsilon_x^2, \epsilon_y^2) \tag{6}$$

where $\boldsymbol{I}$ is the $H \times H$ identity matrix.

Noise Estimation

The data noise $\boldsymbol{\epsilon}$ used for the DLM model is estimated by extracting it from the data using a high pass filter. A 3rd order Butterworth high-pass filter is used, which provides a roll off of -18 dB/decade (Oppenheim and Schafer 2010). A cut-off frequency of $5/\tau$ is used to remove low frequency components such as the diurnal cycle. The noise of each data stream is then represented by the standard deviation of the high-pass-filtered data stream and stored in $\boldsymbol{\epsilon} = [\epsilon_x \quad \epsilon_y]^\top$.

In our experiments, we found that the approach can occasionally underestimate the true noise, resulting in over confident predictions. A lower threshold is thus applied to the noise estimate. The lower threshold is 0.7 for the TPWS data and 1.5 times the TPWS noise for the NWP data. A higher NWP data noise encourages the model to rely more on the TPWS data.

Kalman Filtering

The Kalman filter performs inference on the DLM by computing the posterior filtered distribution $p(\boldsymbol{h}_t|\boldsymbol{v}_{t-T:t}) = \mathcal{N}(\boldsymbol{f}_t, \boldsymbol{F}_t)$, where $\boldsymbol{f}_t$ is the filtered mean, $\boldsymbol{F}_t$ is the filtered covariance, and $\mathcal{N}$ is the Gaussian. The Kalman filter operates in a two-step prediction-update process defined by the following equations:
Prediction step:

$$\boldsymbol{\mu}_h = \boldsymbol{A}\boldsymbol{f}_{t-1} \tag{7}$$

$$\boldsymbol{\Sigma}_{hh} = \boldsymbol{A}\boldsymbol{F}_{t-1}\boldsymbol{A}^\top + \boldsymbol{\Sigma}_h \tag{8}$$

$$\boldsymbol{\mu}_v = \boldsymbol{B}\boldsymbol{\mu}_h \tag{9}$$

$$\boldsymbol{\Sigma}_{vv} = \boldsymbol{B}\boldsymbol{\Sigma}_{hh}\boldsymbol{B}^\top + \boldsymbol{\Sigma}_v \tag{10}$$

Update step:

$$\boldsymbol{K} = \boldsymbol{\Sigma}_{hh}\boldsymbol{B}^\top\boldsymbol{\Sigma}_{vv}^{-1} \tag{11}$$

$$\boldsymbol{f}_t = \boldsymbol{\mu}_h + \boldsymbol{K}(\boldsymbol{v}_t - \boldsymbol{\mu}_v) \tag{12}$$

$$\boldsymbol{F}_t = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{B})\boldsymbol{\Sigma}_{hh} \tag{13}$$

The computed $\boldsymbol{f}_t$ and $\boldsymbol{F}_t$ can be transformed back to the observation space to provide an estimate of the variable of interest in the form $\mathcal{N}(\mu_t, \sigma_t^2)$ using

$$\mu_t = \boldsymbol{C}\boldsymbol{f}_t \tag{14}$$

$$\sigma_t^2 = \boldsymbol{C}\boldsymbol{F}_t\boldsymbol{C}^\top + \epsilon_x \tag{15}$$

where $\boldsymbol{C} = [1 \quad 1 \quad 0 \quad \ldots \quad 0]$, which is identical to the first row of $\boldsymbol{B}$.

[1] E.g. see the function of the Kalman gain (11) in equations (12) and (13)

Table 1: Dataset properties. ($\mathcal{U}$ denotes a uniform distribution).

Dataset	No. of stations	Period	Test Samples
Temperature	99	2017 - 2018	daily minimum
Humidity	100	2018 - 2019	9am and 3pm
Wind gust	100	2016 - 2020	daily maximum

At time $t = 1$, the Kalman filter update equation requires $\boldsymbol{f}_0$ and $\boldsymbol{F}_0$. We set $\boldsymbol{F}_0 = 5\boldsymbol{\Sigma}_h$ and $\boldsymbol{f}_0$ is set using a sequence of T calibration data samples. Assuming T is divisible by τ, the ith element of $\boldsymbol{f}_0$ is set as

$$f_0[i] = \begin{cases} \frac{1}{T}\sum_{t=1}^{T} x_t & i = 0 \\ \frac{\tau}{T}\sum_{t=1}^{T/\tau} x_{\tau(t-1)+i} & i = 1, 2, \ldots H-1 \end{cases} \tag{16}$$

Evaluation

A test sample x_t can be evaluated according to the predicted distribution $\mathcal{N}(\mu_t, \sigma_t^2)$ using to the two-sided p-value

$$p_{\text{two-sided}}(x_t|\mu_t, \sigma_t^2) = 1 - 2\left|0.5 - \Phi\left(\frac{x_t - \mu_t}{\sigma_t}\right)\right| \tag{17}$$

where Φ is the cumulative distribution function. The test sample x_t is suspect if

$$p_{\text{two-sided}}(x_t|\mu_t, \sigma_t^2) < 0.1 \tag{18}$$

Datasets and Testing

In this study data from BoM weather stations are used as TPWS data and the BoM's ACCESS-R (Bureau of Meteorology 2010) provides the NWP data source. We consider daily minimum temperatures, daily maximum wind gusts, and relative humidity observations at 9am and 3pm. Data are gathered for ∼100 stations across Australia, each with at least two years of data. See Table 1. The ACCESS-R data are only available for temperature, and thus the DLM for wind gusts and relative humidity considers the TPWS data streams only.

To test the approach, synthetic datasets are created where data samples are randomly perturbed. A total of 2.7% of minimum temperatures perturbed with uniform distributions $\mathcal{U}(-6°C, -2°C)$ or $\mathcal{U}(2°C, 6°C)$. Similarly, a total of 10% of maximum wind gusts perturbed with $\mathcal{U}(5m/s, 14.6m/s)$.

Directly perturbing relative humidity can violate the relative humidity range of [0%, 100%]. Thus, the dew point temperature is perturbed with $\mathcal{U}(-10°C, -4°C)$ or $\mathcal{U}(4°C, 10°C)$ and translated to relative humidity according to (Bureau of Meteorology 2009):

$$\text{RH} = \frac{100\exp\left(1.8096 + \frac{17.2694\text{T}_d}{237.3+\text{T}_d}\right)}{\exp\left(1.8096 + \frac{17.2694\text{T}_a}{237.3+\text{T}_a}\right)} \tag{19}$$

Here T_d is the dew point temperature and T_a is the dry bulb temperature. Similarly, the DLM is also applied to dew point temperature and the results are converted to relative humidity via (19).

Table 2: Average performance results.

Dataset	Hit rate (recall)	False positive rate	Accuracy
Temperature	80.0%	10.3%	88.9%
Humidity	91.4%	7.78%	92.2%
Wind Gust	84.0%	8.51%	91.0%

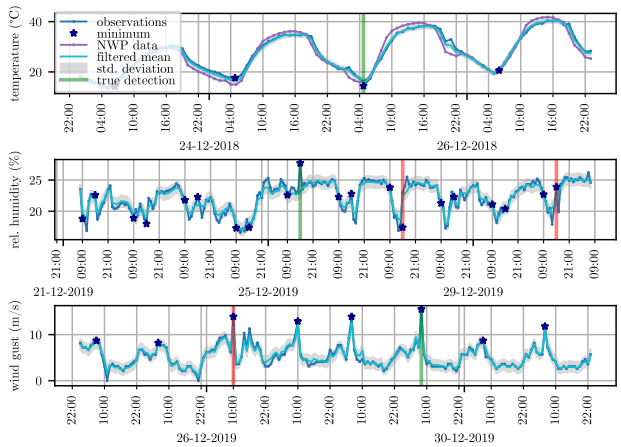

Figure 1: Plots with examples of true and false detections.

During testing, the DLM is applied to the perturbed hourly data-streams for each TPWS and each weather variable. The perturbed test samples are compared with their corresponding DLM predictions and classified as valid or suspect according to (18). The hit rate (number of true positives over the total number of positive cases) and false positive rate (number of false positives over the total number of negative cases) are computed from the classifications.

Results

The average hit rate, false positive rate, and accuracy for the various datasets are presented in Table 2. The high hit rates indicate that the model is able to detect the majority of the perturbed samples. The low false positive rates indicate a low probability of incorrectly classifying a valid sample as a suspect.

A plot of a true detection for temperature is illustrated in the top panel of Fig. 1. The perturbed value is not significantly lower than the predicted minimum, however it is still detected. This particular TPWS provides a relatively smooth diurnal cycle and this has detected through the noise estimation. As such, the DLM is confident in the data and produces a low variance in its predictions. Small perturbations are thus able to be detected.

Not all TPWS produce smooth diurnal cycles. In general, the approach can fail when there is (1) a significant difference between the TPWS and the NWP data, (2) a lack of a diurnal cycle, or (3) a severe non-linear change in the dynamics. Examples of this are illustrated in Fig. 2. In the top panel the NWP data differs from the TPWS observations and there is also a poor diurnal cycle in the data. The result is

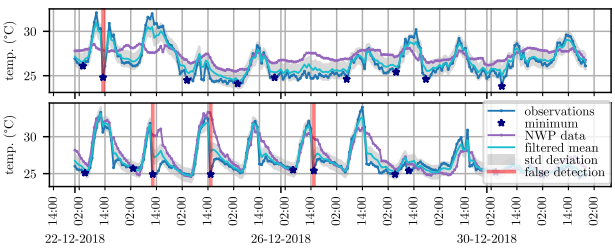

Figure 2: Examples of conditions that cause false or missed detections.

that a high noise level is estimated and the model predictions are associated with high variance. The model is thus desensitized to noise, reducing its ability to detect the perturbations. In the bottom panel of Fig. 2, the NWP underestimates the minimum temperatures and the TPWS data contain non-linear changes in the dynamics, resulting in false positive detections.

Plots of true and false detections for relative humidity and wind gust are presented in the middle and bottom panels of Fig. 1 respectively. These variables generally do not have as strong diurnal cycles as temperature and the data has a higher noise levels. The DLM however generally performs better on these datasets. This is mostly due to the model not using the NWP data. Using NWP data assists with identifying errors such as biases, but it can also have a negative affect when its predictions are not accurate. This usually occurs when the NWP is a low resolution prediction and is not designed to predict highly localised weather affects. Selecting an accurate and high resolution NWP model is crucial.

Conclusion

We present an approach to detect unusual readings from Third-Party Weather Station data and test it on daily minimum temperature, maximum wind gust, and relative humidity. We show that the model is robust in that the same model parameters function over a range of weather stations and weather variables.

Our approach involves Kalman filtering and an initial model calibration via noise estimation. The approach is simple to implement and maintain, making it well suited to production environments. Finally, given that the approach is a time-series based approach, we believe that it will complement the common statistical and heuristic approaches typically used for weather data quality control.

Acknowledgement

We thank the BoM for their assistance in providing the data (including the perturbed data) and for the many discussions on quality control.

References

Beele, E.; Reyniers, M.; Aerts, R.; and Somers, B. 2022. Quality control and correction method for air temperature data from a citizen science weather station network in Leuven, Belgium. *Earth System Science Data*, 14(10): 4681–4717.

Bureau of Meteorology. 2009. calculation of relative humidity. Technical report, BoM.

Bureau of Meteorology. 2010. Operational implementation of the ACCESS Numerical Weather Prediction systems. *NMOC Operations Bulletin*, 83: 12.

Evensen, G. 2009. *Data Assimilation: The Ensemble Kalman Filter*. Earth and Environmental Science. Springer Berlin Heidelberg. ISBN 9783642037115.

Hubbard, K. G.; Guttman, N. B.; You, J.; and Chen, Z. 2007. An Improved QC Process for Temperature in the Daily Cooperative Weather Observations. *Journal of Atmospheric and Oceanic Technology*, 24(2): 206 – 213.

Leung, C. K.-S.; Mateo, M. A. F.; and Nadler, A. J. 2007. An EffectiveMulti-Layer Model for Controlling the Quality of Data. In *11th International Database Engineering and Applications Symposium (IDEAS 2007)*, 28–36.

Mateo, M. A. F.; and Leung, C. K.-S. 2008. Design and Development of a Prototype System for Detecting Abnormal Weather Observations. In *Proceedings of the 2008 C3S2E Conference*, C3S2E '08, 45–59. New York, NY, USA: Association for Computing Machinery. ISBN 9781605581019.

Meier, F.; Fenner, D.; Grassmann, T.; Otto, M.; and Scherer, D. 2017. Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate*, 19: 170–191.

Murphy, K. 2012. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning series. MIT Press. ISBN 9780262018029.

Napoly, A.; Grassmann, T.; Meier, F.; and Fenner, D. 2018. Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data. *Frontiers in Earth Science*, 6.

Oppenheim, A.; and Schafer, R. 2010. *Discrete-time Signal Processing*. Prentice-Hall signal processing series. Pearson. ISBN 9780132067096.

Roh, S.; Genton, M. G.; Jun, M.; Szunyogh, I.; and Hoteit, I. 2013. Observation Quality Control with a Robust Ensemble Kalman Filter. *Monthly Weather Review*, 141(12): 4414 – 4428.

Ting, J.-A.; Theodorou, E.; and Schaal, S. 2007. Learning an Outlier-Robust Kalman Filter. In Kok, J. N.; Koronacki, J.; Mantaras, R. L. d.; Matwin, S.; Mladenic, D.; and Skowron, A., eds., *Machine Learning: ECML 2007*, 748–756. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-74958-5.

West, M.; and Harrison, J. 2006. *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics. Springer New York. ISBN 9780387227771.

Zhang, F.; and Liu, Y. 2009. Study on Quality Control Method of Using the Kalman Filtering. In *2009 International Conference on E-Business and Information System Security*, 1–5.