ECAI 2014
T. Schaub et al. (Eds.)


doi:10.3233/978-1-61499-419-0-1217

# Real-time Adaptive Problem Detection in Poultry

**Steven P.D. Woudenberg**[1] and **Linda C. van der Gaag**[1] and **Ad Feelders**[1] and **Armin R.W. Elbers**[2]

**Abstract.** Real-time identification of unexpected values upon monitoring the production parameters of egg laying hens is quite challenging, as the collected data includes natural variability in addition to chance fluctuation. We present an adaptive method for calculating residuals that reflect the latter type of fluctuation only, and thereby provide for more accurate detection of potential problems. We report on the application of our method to real-world poultry data.

## 1 INTRODUCTION

Motivated by the importance of early detection of disease in animal production, we focus on the real-time identification of unexpected data values in the production parameters of a flock of laying hens.

At modern poultry farms, a variety of parameters are being measured over time; for laying hens for example, these parameters include the daily feed intake and egg production. The collected data represent a wealth of information from which potential problems can in essence be identified. The detection of unexpected values in the production parameters of a poultry flock is quite challenging however, as the data do not just reflect chance fluctuation but natural variability as well. While any flock's egg production is known to follow approximately the same trend over time for example, the moment of peak production can vary by more than ten weeks between flocks.

Most methods for detecting deviating values in real-time collected data build on the calculation of residuals, that is, the differences between actually measured and expected data values. In applications in which the actual data trend should maintain a specific fixed level, such residuals are readily established. For climate regulation in barns for example, temperature and humidity should be kept at constant levels, and residuals are computed by comparing actual data values to a constant. Such residuals moreover, are expected to resemble random fluctuation. Since it is generally accepted that a higher quality of deviation detection is attained with randomly distributed residuals, for applications involving natural variability the effect of differences among instances should be reduced upon residual calculation. Aimed at correcting for natural variability, related research by Mertens and his colleagues resulted in a highly tailored method for detecting trend deviations from daily egg-production and egg-weight data at poultry farms [3, 4]. Since their method relies on a startup period which can become impracticably long for other applications, and hence is not easily generalised, we developed a more general method for residual calculation for applications involving natural variability.

Our method for improved residual calculation builds on the idea of exploiting prior knowledge about an expected data trend and gradually updating this knowledge as actual data values become available. In Section 2 we briefly describe our method. Although our method is aimed at a wide range of applications, we focus in this paper specifically on the detection of unexpected values from the production data of laying hens. Section 3 reports on the use of our method on real-world data from a limited number of flocks. The paper ends with our conclusions and directions for future research in Section 4.

[1] Department of Information and Computing Sciences, Utrecht University, NL; e-mail: {S.P.D.Woudenberg, L.C.vanderGaag, A.J.Feelders}@uu.nl.
[2] Department of Epidemiology, Central Veterinary Institute, Wageningen UR, NL; e-mail: armin.elbers@wur.nl.

## 2 IMPROVED RESIDUAL CALCULATION

Our method for real-time prediction is aimed at establishing residuals including chance fluctuations only, and reduces the influence of natural variability by adapting to the trend observed in a flock at hand. The method starts with an initialization step, and iterates over a classification step and an update step for each new data value:

- *Initialization:* an initial prediction function is chosen from a prespecified function class and conveyed to a regression procedure.
- *Classification:* for each new data point, the residual is calculated as the difference between the actual and predicted value, and the new data point is classified as expected or unexpected.
- *Update:* least-squares regression is used on all available data points with the known function class, to construct a new prediction function for the next data point.

At initialization, the function to be used for predicting future data points is defined. Its function class and the initial values of its coefficients are chosen based on prior knowledge obtained from data or from experts. From the function, an appropriate number of pseudo points are drawn which are conveyed to the regression procedure to exert control over the construction of future prediction functions.

When a new data value is obtained, its residual is calculated as the difference between the actual value and the value computed from the current prediction function. A deviation detection mechanism then serves to classify the new data point as expected or unexpected, based on the calculated residual. For subsequently updating the prediction function, the regression procedure uses all actually observed data and all pseudo points. If a pseudo point had been specified for the time at which a real data value has now been obtained, the pseudo point is discarded: since pseudo points were created to represent prior knowledge about yet to be observed production levels, they become outdated over time and are superseded by real data. In addition, the remaining pseudo points are assigned decreasing weights for the regression to let real data values slowly take over the definition of the prediction function. Further details of our method are provided in [7].

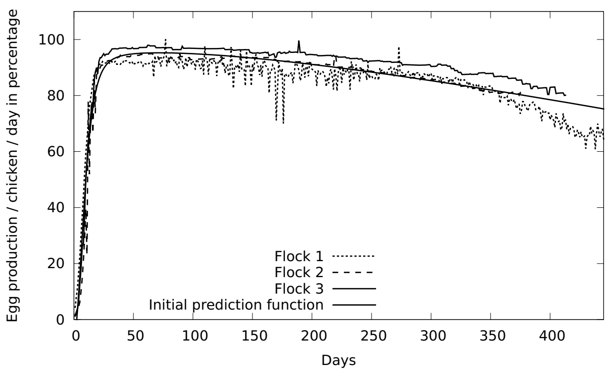
## 3 APPLICATION TO POULTRY PARAMETERS

We employed our method for real-time adaptive prediction to the daily feed-intake and egg-production data from a flock of laying hens for the purpose of identifying potential disease problems.

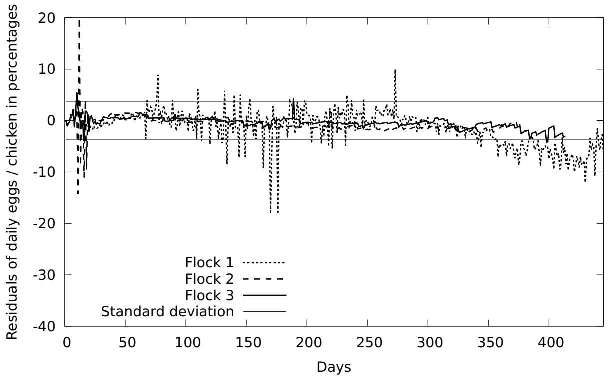
### 3.1 Knowledge Acquisition

For the initialization step of our method, we had to obtain prior knowledge about the data trends expected for the production parameters under study. More specifically, the function classes of these trends and their initial coefficients had to be established. The function class for the trend in egg production was available from the literature [2] and was established to be $f(t) = \frac{100}{1+a \cdot r^{\sqrt{t}}} - (b + c \cdot \sqrt{t} + d \cdot t)$ for time $t$ in general. The function class for the feed-intake trend was elicited from experts and was found to be linear. To arrive at the initial prediction functions, the coefficients involved were estimated from historical data; we will return to this issue in Section 3.2.

From open interviews with poultry experts, we learned that a commonly accepted range of expected values per production parameter does not exist, as even the random fluctuations in the collected data are dependent of the flock and of the farm's management. For example, a farm on which eggs are collected at approximately the same time each day will reveal less fluctuation in the daily egg production, than a farm on which the collecting of eggs is scheduled less strictly. For evaluating the performance of our method therefore, no golden standard for classifying actual data values was available, and we decided to elicit classifications for all data points from experts. To support the elicitation, we plotted the feed-intake and egg-production data from a selection of flocks over time. In order to mimic reality, we put the plots before an expert, keeping future data values hidden and uncovering them in a stepwise manner; one of us took minutes of the expert's remarks upon surveying the data trend. We repeated this approach with several experts, who thereby provided us with a silver standard for the classification of the collected data values from some real-world flocks of laying hens.




(a) Actual egg production data




(b) Egg production residuals

**Figure 1**: Plots of the actual data values of egg production (a), and of the residuals computed with our adaptive prediction method (b).

### 3.2 Results

For studying the performance of our adaptive prediction method, we had available complete production data from ten healthy flocks of laying hens. We used the data from seven of these ten datasets to obtain the coefficients for the initial prediction functions modelling the general trends of feed intake and egg production. To this end, we computed the trends from the selected datasets and averaged their coefficients to find the ones for the initial prediction functions. For the remaining three flocks, classifications of all collected data values were acquired from experts, as described in the previous section.

Figure 1(a) plots the egg-production data from the three flocks for evaluation, along with the initial prediction function. Figure 1(b) shows the residuals calculated for the three flocks by means of our prediction method. By comparing the two plots, it is readily seen that the prediction function initially adapts fairly quickly to the actual data. During the run the function further adapts to the data trends in the flocks, yet without closely following large or sudden deviations. For the other production parameters, similar results were obtained. From these findings, we cautiously conclude that our method is able to effectively reduce the effects from natural variability and to calculate residuals highlighting random fluctuations.

With the calculated residuals, we used various well-known mechanisms for deviation detection [1, 5, 6]. Even with a naive implementation of a simple Shewhart control chart with the negative standard deviation as the lowest accepted value, our method resulted in perfect classification of the egg-production data from a single flock. For the feed intake, a sensitivity of $0.88$ and a specificity of $1.0$ were found for all three flocks. With more sophisticated implementations of deviation detection, we expect to find even better results.

## 4 CONCLUSIONS AND FUTURE WORK

We presented a new method for real-time adaptive calculation of residuals for applications in which collected data include the effects of natural variability in addition to random fluctuation, and reported results from using our method on the daily production data from some flocks of egg laying hens. These results suggest that our method is capable of calculating residuals in which the effect of natural variability is effectively reduced, and hence is likely to provide for better deviation detection. Our future research plans are aimed at enhancing our method with mechanisms for multi-dimensional deviation detection to arrive at even higher performance on real-world data.

**Acknowledgement.** We thank all poultry experts who assisted us.

## REFERENCES

[1] G.E.P. Box, G. Jenkins, *Time Series Analysis, Forecasting and Control.* Holden-Day Incorporated, 1990.
[2] C. Lokhorst, Mathematical curves for the description of input and output variables of the daily production process in aviary housing systems for laying hens. *Poultry Science*, **75**: 838–848, 1996.
[3] K. Mertens *et al.*, Data-based design of an intelligent control chart for the daily monitoring of the average egg weight. *Computers and Electronics in Agriculture*, **68**: 222–232, 2008.
[4] K. Mertens *et al.*, An intelligent control chart for monitoring of autocorrelated egg production process data based on a synergistic control strategy. *Computers and Electronics in Agriculture*, **69**: 100–111, 2009.
[5] D.C. Montgomery, *Statistical Quality Control.* Wiley Hoboken, 2009.
[6] E.S. Page, Continuous inspection schemes. *Biometrika*, 1954.
[7] S.P.D. Woudenberg, L.C. van der Gaag, A. Feelders, A.R.W. Elbers, Real time adaptive residual calculation for detecting trend deviations in systems with natural variability, submitted for publication.