Ministry for Primary Industries
Manatū Ahu Matua

Summary of MPI response to international peer review of the classification modelling methodology (CART) used to produce identification criteria for **mānuka** honey

ISBN No: 978-1-77665-582-3 (online)

May 2017

New Zealand Government

Growing and Protecting New Zealand

Disclaimer

While every effort has been made to ensure the information in this publication is accurate, the Ministry for Primary Industries does not accept any responsibility or liability for error of fact, omission, interpretation or opinion that may be present, nor for the consequences of any decisions based on this information.

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries website at http://www.mpi.govt.nz/news-and-resources/publications/

Contents

Page

1 Executive summary

The Ministry for Primary Industries (MPI) Mānuka Honey Science Programme (MHSP) consisted of a series of scientific activities leading to the development of a definition for mānuka honey.

A core component of the MHSP was the application of a classification model (CART) to develop the identification criteria and this was subjected to independent peer review by three statistical experts. In particular, reviewers were asked to comment on the suitability and robustness of the CART approach and the modelling processes that resulted in the identification criteria.

MPI evaluated the comments received by the three reviewers and summarised the comments according to the categories of questions asked i.e. overall evaluation of the approach used, factors considered in developing the identification criteria, testing the sensitivity and robustness of the identification criteria and the conclusions drawn.

Overall, the peer review of the CART methodology and interpretation of the outputs used to develop identification criteria for monofloral and multifloral mānuka honey supported the findings of the MHSP.

A number of comments were provided including questions on potential alternative ways of analysing data and possible opportunities for improvement. After detailed assessment of the comments, MPI is of the view that the points raised in these comments did not have the potential to warrant a change in the application of the CART methodology or the interpretation of the outputs.

2 Introduction

The MPI Mānuka Honey Science Programme (MHSP) consisted of a series of scientific activities leading to the development of a definition for mānuka honey and is summarised in a technical report (A summary of the manuka honey science programme: MPI Technical Paper No. 2017/28).

A core science activity was the application of a classification model (classification and regression trees – CART) to produce the identification criteria for monofloral and multifloral mānuka honey. The classification modelling approach involved building and testing CART models using marker data from different honey production years and testing the sensitivity and robustness of CART outputs under a range of scenarios.

MPI commissioned an independent peer review of the CART modelling process to provide additional confidence in the appropriateness and outputs of this part of the MHSP. In particular, the three reviewers were asked to comment on the robustness of the CART approach and the analytical processes that resulted in the identification criteria.

This document provides a summary of the reviewers' comments and MPI's response.

3 Peer review process

The peer review process involved:

- Providing a detailed technical analysis of the CART and outputs to the reviewers
- Providing a clear definition of the scope of the review (section 3.1), with reviewers requested to comment on the suitability and robustness of the CART approach, as well as the sensitivity and robustness of the various modelling processes and outputs that resulted in the identification criteria.
- Detailed evaluation by MPI of the comments provided by the peer reviewers.

3.1 SCOPE OF THE PEER REVIEW

The reviewers were asked to provide comments in the following categories;

1. **Overall analysis of the approach used**
    a. Were appropriate statistical methods used to produce the identification criteria, given the datasets?
    b. Were assumptions clearly stated and are they reasonable?
    c. Were all important data analysis aspects considered?
    d. Have limitations of the approach been identified and appropriately considered in the analyses?
2. **Factors considered in developing the identification criteria**
    a. Was the process to develop the identification criteria appropriate?
    b. Were the identification criteria proposed supported by the data and associated analyses?
3. **Testing the sensitivity and robustness of the identification criteria**
    a. Was the sensitivity of the criteria tested appropriately?

b. Do the range of scenarios explored in the bootstrapping procedures (numerical simulations) test the robustness of the identification criteria?

4. **Conclusions**
    a. Were the conclusions in the report justified by the evidence and analyses presented?
    b. Do you have any other recommendations?

At times, the reviewers provided comments on aspects of the MHSP that were additional to those relating to the CART While out of scope, these were still taken into account in the MPI response.

4 MPI evaluation of peer review comments

The MPI evaluation of comments by reviewers is presented under each of the review categories described as above. This takes the following form:

- an aggregated summary of comments received from the three peer reviewers.
- a summary of any limitations as seen by reviewers.
- an evaluation of the reviewers' comments.

A final section of the report provides the overall interpretation by MPI of the comments provided through the peer review process.

4.1 REVIEW CATEGORY 1: OVERALL ANALYSIS OF THE CART APPROACH

4.1.1 Summary of comments

All of the reviewers commented that the selection of the CART model as a methodology to develop the identification criteria was valid and appropriate. Reviewers suggested that MPI provide more technical details on the CART models used, specifically around the starting parameters. These were available in the references provided, but not explicitly stated in the report. Reviewers noted that MPI used the output from the classification trees and applied pragmatic modifications (such as rounding threshold values) to produce the final classification rules. Reviewers also noted that the main assumptions made were reasonable and clearly reported**.**

All important aspects of application of the CART were considered to have been available for peer review.

Reviewers stated that the identification criteria were justified by the statistical analyses presented. Further attention to specific comments would add to further justification to the outputs and conclusions from the MHSP (sections 4.2.2, 4.3.2, and 4.4.2).

4.1.2 Limitations as seen by reviewers

Reviewers asked for further information on the CART models in the figures in the documentation presented for review. Details should be added around the percentage of samples which are now classified under each honey type when the CART model is applied. Threshold values on figures should be presented on the original scale of the data rather than on the log transformed scale to help the reader.

Reviewers asked why a binary classification (mānuka honey or non-mānuka honey) was not sought rather than one which classified more than two different honey types (e.g. monofloral mānuka, multifloral mānuka and non-mānuka honey). Note that the latter scenario was a key aspect of the MHSP.

Several other comments were provided on detailed technical aspects of the CART method; these can be grouped into assumptions made when building the CART (quantification of losses) and methods used to test the CART outputs (cross-validation):

- Reviewers suggested implementing a formal quantification of losses (using a loss function) to incorporate the misclassification of honey samples as other honey types when building the CART model. This would enable the misclassification of some honey types to be formally accounted for as more acceptable than others.
- One reviewer stated that while the assumptions underlying the CART model are reasonable, the use of a loss function is not targeted to the primary aim (correct classification of mānuka honey) and needs to be better defended.
- Although a modified form of cross-validation was used as part of the CART modelling approach, reviewers wanted further justification of this alternate approach in the document.

4.1.3 MPI evaluation of reviewers comments

MPI agrees with the reviewers comments regarding the presentation of the results within the figures and tables in the technical documentation. In addition, MPI notes that further clarification could be provided on methodology, particularly around CART model parameters such as the loss function.

MPI notes that the classification of samples into honey types is not as simple as a binary separation (mānuka honey or non-mānuka honey). The large variation of floral sources included within a sample classed as a "non-mānuka honey" as well as the likely presence of mānuka nectar within these samples would make it difficult to separate honey samples into only two types. In addition, the aim of the MHSP was to provide identification criteria for both monofloral and multifloral mānuka honey which a binary classification approach would not provide.

MPI did consider using a formal quantification of losses within the CART models to numerically take into account the misclassification of some honey types. This was explored as an option in initial analyses. However, with no reference standards for identifying any of the honey types, and the likely variability associated with the labelling of the honey samples by the suppliers, the quantification would be difficult to justify. Consequently, misclassifications were categorised as an ordered degree of misclassification: irrelevant, mild or severe. As the misclassification of honey types was not considered of equal importance in the approach used, a formal loss was not incorporated during building of the CART model. MPI notes that such a formal quantification of losses would need careful justification and may introduce bias into the model outputs. The quantification of losses could be examined through further sensitivity analyses, however, too many assumptions would need to be made in assigning a quantifiable loss to each misclassification.

Cross-validation was considered as a valid method to test the CART model during initial data analyses, however, a standard application of cross-validation would involve dividing the data set randomly into specific fractions, such as 10%. This would provide ten parts of (roughly) equal size. MPI considered that this was not appropriate with the current data set, as the three sources of data (2014/2015; 2015/2016; archival) had noticeable different properties. As such,

MPI determined it was not valid to treat the entirety of data as coming from a single, homogenous source. In actuality,, the approach adopted to test the CART model was a stricter test as one data set was used (e.g. data for 2014/15) to create a classification rule which was evaluated using a different data set (e.g. data for 2015/16) which had markedly different characteristics overall.

4.2 REVIEW CATEGORY 2: FACTORS CONSIDERED IN DEVELOPING THE IDENTIFICATION CRITERIA

4.2.1 Summary of comments

Reviewers commented that the process used to develop the identification criteria was appropriate and commensurate with good practice as in other similar applications.

It was suggested that some aspects of the process required further explanation, but this did not detract from the appropriateness of the process itself.

4.2.2 Limitations as seen by reviewers

Honey samples from different production years were used to build and test the CART model. Two reviewers suggested combining the honey samples across years to build and test the CART models.

One reviewer commented that while overall the model limitations are documented in the MPI report, stronger caveats could be used for interpreting results such as the influence of year to year variability on the classification of future honey samples.

4.2.3 MPI evaluation of reviewers comments

MPI recognises honey is a natural product subject to multiple influences, therefore, honey produced each year is variable even within the same honey type. However, the inclusion of archive samples in our study provides evidence that the identification criteria can be successfully retrospectively applied as far back as the 2009/2010 season.

MPI did build CART models with the entire dataset (from both 2014/215 and 2015/16 years) when exploring which type of CART models to use in the analyses. However, this scenario would only work if the whole dataset was relatively homogenous. Given the fact that there were differences between years we did not consider it valid in general to combine them. The data from different years were combined to explore North Island and South Island differences, but this was to ensure identification criteria could successfully be applied to samples sourced from across New Zealand.

4.3 REVIEW CATEGORY 3: TESTING THE SENSITIVITY AND ROBUSTNESS OF THE IDENTIFICATION CRITERIA

4.3.1 Summary of comments

The sensitivity assessment adopted to test the criteria was found to be comprehensive and well described by the reviewers. Although the CART approach as chosen by MPI was judged suitable, reviewers recommended that MPI also consider other statistical models.

Reviewers commented that the range of scenarios explored to test the robustness of the identification criteria was sufficient given the available data and choice of analysis.

4.3.2 Limitations as seen by reviewers

Reviewers stated that the classification success of the CART models is relatively low in comparison to other applications of a classification model. However, reviewers accepted the original classification of the honey samples may not have been correct; thereby contributing to a low success. One reviewer suggested a formal quantification of the misidentification would be useful.

One reviewer commented that the summary statistics are based on the misclassification rates but a more complete discussion of the results would relate to the nature of the misclassifications. For example the sensitivity and specificities for the comparisons of primary interest would be more informative than the aggregate misclassification rates.

Reviewers noted that the variability between the two honey production years indicates identification criteria produced from the CART model may not have the same success rate in classifying honey samples from future seasons. However, they recognised it would be difficult to achieve the same classification success without accounting for environmental influences in the model i.e. rainfall and temperature data.

4.3.3 MPI evaluation of comments

MPI did consider the suitability of a range of alternative statistical models for classification of the honey types. Given that the aim was to produce an output that was transparent, we selected CART over other models, such as boosted regression trees (BRTs) and Random Forests, despite the supposedly better performance of the latter methods. Direct comparison with these would be interesting but would be an academic exercise only.

MPI notes that the classification rates are low *in comparison* with other classification models applied in biological systems. However, with regards to honey the classification rates are higher than expected because of the different approaches by suppliers to determine honey type. It is true there are areas of the classification which could receive further attention, however, one of the key aspects of the classification is that no overseas honeys were classified as New Zealand mānuka honey and only one non-mānuka honey sample was classified as a monofloral mānuka honey. In terms of product authenticity, these are important results in terms of implementing the outcomes of the work.

Prior to the MHSP, there were no clear criteria to separate multifloral from monofloral mānuka honey. It also must be noted that many honey samples of other floral sources will contain some level of mānuka pollen/nectar. The problem is therefore in the heterogeneity of the product as honey is a natural substance resulting from a complex interaction of bees and plants, rather than being a limitation of the analysis. Ideally, the level of misidentification of honey samples in the training dataset could be quantified, but this is not practical or possible given the variety of approaches used by suppliers to identify the floral source of a honey sample. This would need to be the aim of entirely independent research project as the misidentification of honey samples would be dependent on supplier, honey type and region.

MPI agrees with the reviewer commenting that the nature of the misclassifications is important, but disagrees that this was not discussed or considered. The severities of misclassification are defined in the methodology section of the technical document and they were summarised and compared in the results section. Thus the sensitivity and specificity of the different severities of misclassification were assessed as part of the process.

MPI recognises honey is a natural product subject to multiple influences, therefore, honey produced each year is variable even within the same honey type. It is important to recognise that identification criteria should be monitored over time for the potential impact of influences such as climate change, as well as bee and plant disease events. It is possible that nectar

production and associated chemicals may change if there is marked change in environmental conditions. Similarly disease events that affect bees or the source plant may affect the levels of markers in the honey produced. However, the inclusion of archive samples in our study provides evidence that the identification criteria can be successfully retrospectively applied as far back as the 2009/2010 season.

Including environmental variables in the CART model is theoretically possible and would be an interesting academic exercise. In practice, it would be extremely difficult. This would involve accessing or collecting relevant environmental data at the relevant scale of the apiary site and at the correct time and location for each honey type produced. The same environmental data would need to be available to producers who would use the identification criteria derived from the CART results. This would be very difficult to achieve. Incorporating environmental data may also conclude that authentic mānuka honey could only be classified, and therefore produced, under certain environmental conditions. This would not achieve the aim of the MHSP.

5 Overall assessment of comments by MPI

Overall the independent peer review supported the application of the CART modelling approach and the interpretation of outputs to develop identification criteria for mānuka honey. The peer reviewers did not identify any major causes of concern.

A number of comments were provided on potential limitations and opportunities for improvement. After detailed assessment of the comments, MPI is of the view that the points raised in these comments did not have the potential to warrant a change in the use of the CART or the interpretation of the outputs.