

Mathematics for Industry 28

Robert S. Anderssen

Philip Broadbridge

Yasuhide Fukumoto

Kenji Kajiwara

Matthew Simpson

Ian Turner *Editors*

Agriculture as a Metaphor for Creativity in All Human Endeavors

Mathematics for Industry

Volume 28

Aims & Scope

The meaning of “Mathematics for Industry” (sometimes abbreviated as MI or Mfi) is different from that of “Mathematics in Industry” (or of “Industrial Mathematics”). The latter is restrictive: it tends to be identified with the actual mathematics that specifically arises in the daily management and operation of manufacturing. The former, however, denotes a new research field in mathematics that may serve as a foundation for creating future technologies. This concept was born from the integration and reorganization of pure and applied mathematics in the present day into a fluid and versatile form capable of stimulating awareness of the importance of mathematics in industry, as well as responding to the needs of industrial technologies. The history of this integration and reorganization indicates that this basic idea will someday find increasing utility. Mathematics can be a key technology in modern society.

The series aims to promote this trend by (1) providing comprehensive content on applications of mathematics, especially to industry technologies via various types of scientific research, (2) introducing basic, useful, necessary and crucial knowledge for several applications through concrete subjects, and (3) introducing new research results and developments for applications of mathematics in the real world. These points may provide the basis for opening a new mathematics oriented technological world and even new research fields of mathematics.

More information about this series at <http://www.springer.com/series/13254>

Editor-in-Chief

Masato Wakayama (Kyushu University, Japan)

Scientific Board Members

Robert S. Anderssen (Commonwealth Scientific and Industrial Research Organisation, Australia)

Heinz H. Bauschke (The University of British Columbia, Canada)

Philip Broadbridge (La Trobe University, Australia)

Jin Cheng (Fudan University, China)

Monique Chyba (University of Hawaii at Mānoa, USA)

Georges-Henri Cottet (Joseph Fourier University, France)

José Alberto Cuminato (University of São Paulo, Brazil)

Shin-ichiro Ei (Hokkaido University, Japan)

Yasuhide Fukumoto (Kyushu University, Japan)

Jonathan R.M. Hosking (IBM T.J. Watson Research Center, USA)

Alejandro Jofré (University of Chile, Chile)

Kerry Landman (The University of Melbourne, Australia)

Robert McKibbin (Massey University, New Zealand)

Andrea Parmeggiani (University of Montpellier 2, France)

Jill Pipher (Brown University, USA)

Konrad Polthier (Free University of Berlin, Germany)

Osamu Saeki (Kyushu University, Japan)

Wil Schilders (Eindhoven University of Technology, The Netherlands)

Zuowei Shen (National University of Singapore, Singapore)

Kim-Chuan Toh (National University of Singapore, Singapore)

Evgeny Verbitskiy (Leiden University, The Netherlands)

Nakahiro Yoshida (The University of Tokyo, Japan)

Robert S. Anderssen · Philip Broadbridge
Yasuhide Fukumoto · Kenji Kajiwara
Matthew Simpson · Ian Turner
Editors

Agriculture as a Metaphor for Creativity in All Human Endeavors

 Springer

Editors

Robert S. Anderssen
Mathematics, Information and Statistics
CSIRO
Acton, ACT
Australia

Kenji Kajiwara
Kyushu University
Fukuoka
Japan

Philip Broadbridge
La Trobe University
Melbourne, VIC
Australia

Matthew Simpson
Queensland University of Technology
Brisbane, QLD
Australia

Yasuhide Fukumoto
Kyushu University
Fukuoka
Japan

Ian Turner
Queensland University of Technology
Brisbane, QLD
Australia

ISSN 2198-350X

ISSN 2198-3518 (electronic)

Mathematics for Industry

ISBN 978-981-10-7810-1

ISBN 978-981-10-7811-8 (eBook)

<https://doi.org/10.1007/978-981-10-7811-8>

Library of Congress Control Number: 2017963860

© Springer Nature Singapore Pte Ltd. 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This book is the proceedings of the conference “Forum Math-for-Industry 2016 (FMfI2016)” held at Queensland University of Technology (QUT), for November 21–23, 2016, for which the unifying theme was “Agriculture as a metaphor for creativity in all human endeavors” and collects together selected papers presented there.

The agricultural process of planting a seed, fertilizing, growing, and harvesting has a clear parallel with the application of mathematics to a practical problem. The seed becomes the question being asked, the fertilization is the conceptualization of the mathematical framework within which to seek the answer, the growing is the solution process, and the harvesting is the articulation and implementation of the answer.

In agriculture, the breeding of the seed to plant involves genetics; the germination of the plant involves moisture alone; the growth involves the interaction between the biology and environment with soil, water, and weather the key drivers; the survival depends on its ability to respond to viral and fungal infections and stress challenges; and the flowering and setting of the seed for the next generation depend on the occurrence of environmental cues.

For understanding the processes and mechanisms involved with each of these steps, mathematical modeling is central. This is reflected in the emergence of new mathematically focused agriculture endeavors such as “precision agriculture,” “smart agriculture analytics,” and “digital agriculture.”

The success of agriculture practice relies fundamentally on its interconnections with and dependence on biology and the environment. Both play fundamental roles including the adaption of biology to cope with environmental challenges of biotic and abiotic stresses and global warming. FMfI2016 explored the contribution of mathematics within the framework of the interaction of agriculture with biology and the environment.

The contents of this volume report on productive and successful interaction between industry and mathematicians, as well as on the cross-fertilization and collaboration that occurred. The book contains excellent examples of the roles of

mathematics in innovation and, thereby, the importance and relevance of the concept Mathematics_FOR_Industry.

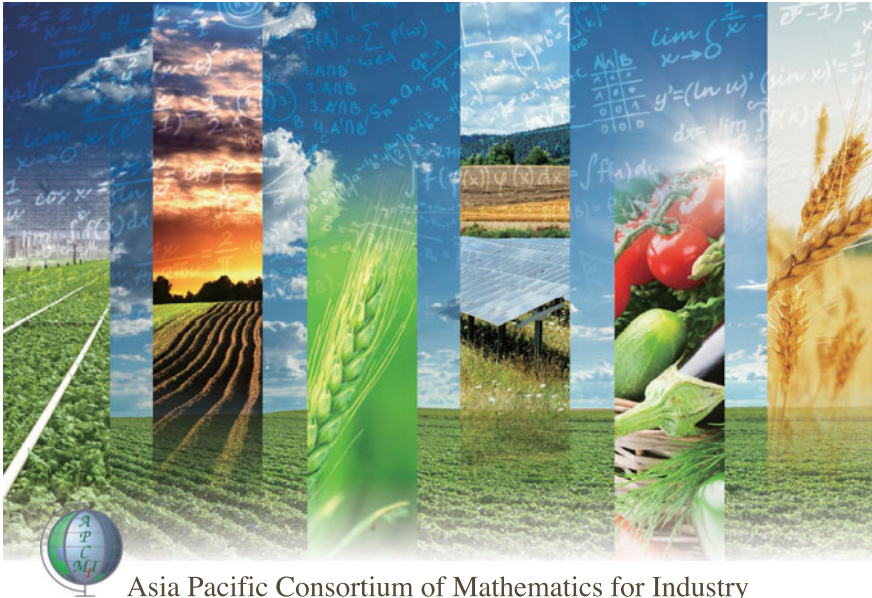
We would like to thank the participants of the forum and the members of the Scientific Board of the Forum, especially Troy Farrell, Matthew Simpson, and Ian Turner of QUT. Without their cooperation and support, we would never have experienced the great excitement and success of the forum. Moreover, we would like to express our deep appreciation for the great help of the conference secretaries during the preparation and organization of the forum, and Chiemi Furutani for the proceedings.

Fukuoka, Japan
April 2017

Yasuhide Fukumoto
On behalf of the Organizing Committee of
the Forum Math-for-Industry 2016 and the
Editorial Committee of the Proceedings



	FMI2008	FMI2009	FMI2010	FMI2011	FMI2012	FMI2013	FMI2014	FMI2015	FMI2016
Location	Tokyo	Fukuoka	Fukuoka	Honolulu	Fukuoka	Fukuoka	Fukuoka	Fukuoka	Brisbane
Date	Sep.16-17	Nov.9-13	Oct.21-23	Oct.24-28	Oct.22-26	Nov.4-8	Oct.27-31	Oct.26-30	Nov.21-23
Theme	The 1st Forum: Consortium Math For Industry	Casimir Force, Casimir Operators and the Riemann Hypothesis	Information Security, Visualization, and Inverse Problems, on the basis of Optimization Techniques	TSUNAMI - Mathematical Modelling- Using Mathematics for Natural Disaster Prediction, Recovery and Provision for the Future	Information Recovery and Discovery	The Impact of Applications on Mathematics	Applications + Practical Conceptualization + fruitful Innovation	The Role and Importance of Mathematics in Innovation	Agriculture as a metaphor for creativity in all human endeavors
Visuals									



Asia Pacific Consortium of Mathematics for Industry
Forum "Math-for-Industry" 2016
 Agriculture as a metaphor for creativity in all human endeavors
 November 21-23, 2016

Science and Engineering Centre, Gardens Point Campus, Queensland University of Technology, Brisbane, Australia

INVITED SPEAKERS

Alona Ben-Tal	(Massey University, New Zealand)
Peter Caley	(CSIRO, Australia)
Adelle Coster	(The University of New South Wales, Australia)
Bronwyn Harch	(Queensland University of Technology, Australia)
Kai Hirose	(Kyushu University, Japan)
Tony Jakeman	(Australian National University, Australia)
Julien Lerat	(Bureau of Meteorology, Australia)
Osamu Maruyama	(Kyushu University, Japan)
James McCaw	(The University of Melbourne, Australia)
Kerrie Mengersen	(Queensland University of Technology, Australia)
Mikio Murata	(Tokyo University of Agriculture and Technology, Japan)
Toshiyuki Nakagaki	(Hokkaido University, Japan)
Mark Nelson	(University of Wollongong, Australia)
Takashi Okayasu	(Kyushu University, Japan)
Tristan Perez	(Queensland University of Technology, Australia)
Julia Piantadosi	(University of South Australia, Australia)
Mick Roberts	(Massey University, New Zealand)
Akiko Satake	(Kyushu University, Japan)
Jason Sharples	(The University of New South Wales, Australia)
Jo Simpson	(Fonterra, New Zealand)
Takeshi Tsuji	(Kyushu University, Japan)
Hayato Waki	(Kyushu University, Japan)
Peter Waterhouse	(Queensland University of Technology, Australia)
Yoshihiro Yamanishi	(Kyushu University, Japan)
Shintaro Akamine	(Kyushu University, Japan)
Elliot Carr	(Queensland University of Technology, Australia)
Anton Gulley	(The University of Auckland, New Zealand)

ORGANIZING COMMITTEE

Bob Andersen	(CSIRO, Australia)
Philip Broadbridge	(La Trobe University, Australia)
Trey Farrell	(Queensland University of Technology, Australia) Chair
Yasuhide Fukumoto	(IMI, Kyushu University, Japan)
Alexandra Hogan	(Australian National University, Australia)
Kenji Kajiwara	(IMI, Kyushu University, Japan)
Robert McKibbin	(Massey University, New Zealand)
Matthew Simpson	(Queensland University of Technology, Australia)
Ian Turner	(Queensland University of Technology, Australia)

Supported by

Asia Pacific Consortium of Mathematics for Industry
 QUT School of Mathematical Sciences, Science and Engineering Faculty
 Institute of Mathematics for Industry, Kyushu University



FMfI2016 Organizing Office

QUT School of Mathematical Sciences, Science and Engineering Faculty
 Gardens Point Campus fmf2016@imi.kyushu-u.ac.jp
<http://apcmfi.org/fmf2016/index.html> Tel: +61-7-3138-2308

Forum "Math-for-Industry" 2016 Agriculture as a metaphor for creativity in all human endeavors Science and Engineering Centre, Garden's Point Campus, Queensland University of Technology, Brisbane, Australia				
Time	20-Nov	21-Nov	22-Nov	23-Nov
	SUNDAY	MONDAY	TUESDAY	WEDNESDAY
9:00-9:30		Tristan Perez Queensland University of Technology, Australia	Mark Nelson University of Wollongong, Australia	Alona Ben-Tal Massey University, New Zealand
9:30-10:00		Adelle Coster The University of New South Wales, Australia	Julien Lerat Bureau of Meteorology, Australia	Hayato Waki Kyushu University, Japan
10:00-10:30		Bronwyn Harch Queensland University of Technology, Australia	Takashi Okayasu Kyushu University, Japan	Takeshi Tsuji Kyushu University, Japan
10:30-11:00		Morning Tea	Morning Tea	Morning Tea
11:30-12:00		James McCaw The University of Melbourne, Australia	Jason Sharples The University of New South Wales, Australia	POSTER PRESENTATION
12:00-12:30		Mick Roberts Massey University, New Zealand	Julia Piantadosi University of South Australia, Australia	
12:30-13:00		Mikio Murata Tokyo University of Agriculture and Technology, Japan	Peter Caley CSIRO, Australia	
13:00-13:30		Lunch	Lunch	Lunch
13:30-14:00				Yoshihiro Yamanishi Kyushu University, Japan
14:00-14:30		Kerrie Mengersen Queensland University of Technology, Australia	EXCURSION	Peter Waterhouse Queensland University of Technology, Australia
14:30-15:00		Kei Hirose Kyushu University, Japan		Akiko Satake Kyushu University, Japan
15:00-15:30		Anthony Jakeman Australian National University, Australia		Afternoon Tea
15:30-16:00		Afternoon Tea		Joanne Simpson Fonterra, New Zealand
16:00-16:30		Shintaro Akamine Kyushu University, Japan		Osamu Maruyama Kyushu University, Japan
16:30-17:00		Elliot Carr Queensland University of Technology, Australia		Toshiyuki Nakagaki Hokkaido University, Japan
17:00-17:30		Anton Gulley The University of Auckland, New Zealand		
17:30-18:00	APCMfi Meeting	IMI-IAB Meeting		
18:00-18:30				
18:30-19:00	FMI Executive Meeting			WORKSHOP DINNER
19:00-19:30				
19:30-20:00				
20:00-				

Biology	Agriculture	Environment	Early Career Researchers	Other Topics
---------	-------------	-------------	--------------------------	--------------

Contents

The Shape of Things to Come—Using Geometric and Morphometric Analyses to Identify Archaeological Starch Grains	1
Adelle C. F. Coster and Judith H. Field	
New Mathematical Models of Antimalarial Drug Action to Improve Drug Dosing Regimens	7
James M. McCaw, Pengxing Cao, Sophie Zaloumis and Julie A. Simpson	
Biodiversity and the Ecology of Emerging Infectious Diseases	13
M. G. Roberts and J. A. P. Heesterbeek	
Reaction–Diffusion Equations and Cellular Automata	21
Mikio Murata	
Geometry of Timelike Minimal Surfaces and Null Curves	37
Shintaro Akamine	
Accounting for Modelling Errors in Parameter Estimation Problems: The Bayesian Approximation Error Approach	47
Ruanui Nicholson, Anton Gulley, Jari Kaipio and Jennifer Eccles	
Reducing Sludge Formation in the Activated Sludge Process	53
M. I. Nelson	
Incorporating Prior Knowledge in the Calibration of Hydrological Models for Water Resources Forecasting	61
Julien Lerat	
Maintaining Reliable Agriculture Productivity and Goyder’s Line of Reliable Rainfall	73
Julia Piantadosi and Robert S. Anderssen	

Vertex Representation of Convex Hulls for Fast Evaluation of Boundary Constraints in Model-Based Calibration for Automotive Engines 85
Florin Nae and Hayato Waki

Mathematical Modeling of Rock Pore Geometry and Mineralization: Applications of Persistent Homology and Random Walk 95
Takeshi Tsuji, Fei Jiang, Anna Suzuki and Tomoyuki Shirai

Statistical Machine Learning for Agriculture and Human Health Care Based on Biomedical Big Data 111
Yoshihiro Yamanishi, Yasuo Tabei and Masaaki Kotera

Compactly Supported Solutions of Reaction–Diffusion Models of Biological Spread 125
Maureen P. Edwards, Bronwyn H. Bradshaw-Hajek,
María Jesús Muñoz-Lopez, Peter M. Waterhouse and Robert S. Anderssen

Two Challenging Difficulties of Protein Complex Prediction 139
Osamu Maruyama

An Overview of Methods to Identify and Manage Uncertainty for Modelling Problems in the Water–Environment–Agriculture Cross-Sector 147
A. J. Jakeman and J. D. Jakeman

Index 173

The Shape of Things to Come—Using Geometric and Morphometric Analyses to Identify Archaeological Starch Grains

Adelle C. F. Coster and Judith H. Field

Abstract Starch grains are tell-tale characteristics of plants that can remain long after the decomposition of the rest of the material. The understanding of historical plant use, for sustenance and plant-based medicines, as well as agricultural practices is enhanced by the identification of residual starch remains. Classifications, however, have previously relied on expert identification using largely subjective features. This can be enormously time consuming and subject to bias. A method has been developed to construct robust classifiers for starch grains of unknown origin based on their geometrical and morphometric features. It was established to allow insight into plant food use from archaeological remains but could be used in many different contexts.

Keywords Mathematics-for-Industry · Starch grains · Identification
Geometric analysis · Morphometric analysis

1 Introduction

Starch grains can be preserved for millennia on grinding tools and surfaces. They are insoluble granules of carbohydrates that build up in plants. They have an initial growth point, or hilum, and then layers of material build up around this point. As they are simply carbohydrate, it is not possible to tell from their chemical composition the species from which they came. However, due to the different cellular structures of different plants, the starch grains have shapes that are characteristic of the particular species and organelle from which they are derived. The features of the starch grains from known plant species can form a reference library. The within-species

A. C. F. Coster (✉)

School of Mathematics and Statistics, UNSW, Sydney, Australia

e-mail: A.Coster@unsw.edu.au

J. H. Field

School of Biological, Earth and Environmental Sciences, UNSW, Sydney, Australia

e-mail: judith.field@unsw.edu.au

© Springer Nature Singapore Pte Ltd. 2018

R. S. Anderssen et al. (eds.), *Agriculture as a Metaphor for Creativity*

in *All Human Endeavors*, Mathematics for Industry 28,

https://doi.org/10.1007/978-981-10-7811-8_1

features and variability can then be utilised to both describe the species and also as a discriminator to classify unknown samples, recovered from artefacts, soils or other materials.

2 Experimental Methods

The starch grains can be recovered from the artefact, soil or other material, suspended in a mixture of glycerol and water and imaged using a brightfield, differential interference contrast microscope. In the current approach, we extract the two-dimensional maximum-projection-area grain shape (identified as the in-focus edge of the grain), or region of interest, ROI, from the light micrograph. Additionally, the hilum point is identified. In our method, a hybrid approach to edge detection is employed which combines automatic outline analysis with some expert intervention to finalise the outline and hilum positioning. This is because purely automated edge detection is difficult across a large variety of images—the grains have different depths of fields, and the assessment of the in-focus edge can be unclear when the morphology causes different shadowings [1]. An accurate edge is important as the ROI is used to obtain the discriminative features to classify the grains. Morphological dilation and erosion, common image analysis techniques for edge detection and object separation, degrade the features of the ROI.

2.1 *Geometric and Morphometric Features*

The ROI and hilum location are used to calculate the geometric and morphometric features of the grain. These include the area, perimeter and centre of mass of each grain. Hilum offset measures encode the position of the centre of mass and then compared to the hilum position. The maximum length through the hilum of the grain has previously been shown to be a good discriminant for some plant taxa [2]. This maximum length line running through the hilum also provides a reference angle from which other features can be observed. Other characteristics such as circularity and other shape matching measures and curvature metrics can also be calculated.

In our case, the digitised ROI has very closely spaced edge points. For completeness, however, we approximate the periphery radius as a piecewise linear function of the angle about the hilum position relative to the maximum length line.

The starch grains in our studies have no convexity issues, so it is possible to expand the periphery radius as a radial Fourier series. Thus, we can generate a model from a truncated sum that can approximate the grain shape. In practice, we have found that the perimeter features can be represented by as few as five terms (see [6]). The radial harmonic components of the Fourier decomposition are characteristic of the grain shape and can be used to discriminate between different species.

However, if convexity is an issue, other decompositions can be employed such as wavelet shape (e.g. [3]), multi-scale fractal dimension or curvature scale-space analysis (e.g. [4]).

3 Starch Grain Classification

In order to discriminate between species, however, it is important to have a comprehensive, well-curated reference library with which to compare grains of unknown origin. The reference set for the classifier also needs to be appropriate to the geographic region [5]. The reference grains need to be sourced from the appropriate plant parts—e.g. seeds, fruiting bodies and tubers. It is also good practice to use multiple samples for each species, to, for instance, account for variation in environmental conditions and their possible impact on the size of the grains. The reference set for the classifier should also be, for our archaeological purposes, of important economic plants from the appropriate location, altitude and climate from the people occupying the site at the period of interest. Evidence of use by the people using the land is also important.

Samples of starch grains from the reference species were analysed to calculate the predictor variable values for each grain in the population. Within species, the shapes will have some variance and may also vary with orientation. We have found that the within-species variation appears to be captured by approximately 100 grains in the 80 species we have analysed so far, [6, 7], ensuring a statistically significant result and allowing the decomposition of the species into sub-grouping as required.

The morphometric measures of the two-dimensional projections are used to determine the classifiers that were best able to discriminate between the grains. Classifiers were considered for various choices of the predictor variables, the classifier type, the training set (the species to be considered) and the output classes—the species or their sub-groupings, and in latter case, the method of resulting species prediction.

Series of classifiers were built taking different combinations of the predictor variable, which included the maximum length, area, perimeter, circularity, hilum offset measures and the Fourier components. Other measures such as the shape matching variables were investigated but found to be of lesser discriminative value than those listed above.

Possible combinations of predictor variables can be explored by calculating a MANOVA of sets of predictor variables for the reference species (or sub-species). Separation of the species by the MANOVA is an indicator that it may be feasible to discriminate between the species in a classifier. It may be possible, however, to still positively identify the presence of a subset of the training set, even if others are indistinguishable.

In these investigations, the classifiers were broadly discriminant, nearest neighbour and decision tree; however, other algorithms including neural networks and support vector machines could also be used. In deciding which algorithms to use,

there is a trade-off in performance and the number of design parameters that need to be explored.

The method to assess the classifiers depends on the type of output desired. For instance, if a distribution of a particular species is to be estimated, then the classifier needs to be designed to best classify all the unknown grains. Designing a classifier to obtain a very high confidence, true-positive identification for some of the unknown grains may be to the detriment of certain classifications for others.

It is the latter option which was taken in assessing the archaeologically provenanced starch grains. Whilst the distribution of species was of interest, it was deemed to be of more importance to know that there was evidence for a particular species being present in the samples—this would, for instance, indicate that a particular plant was consumed if the grain was found on a grinding stone or in tooth calculus.

The accuracy of the different classifiers constructed was assessed both by re-substitution and cross-validation. Re-substitution assesses the accuracy of a classifier that has been constructed using all the data. Each (known) grain in the training set is identified using the classifier and the rate of true and false positive classifications determined. Given the natural within-species variation and possible overlaps between species, it is unlikely that 100% accuracy will be achieved, even when re-substituting the grains used in the classifier construction. Re-substitution accuracy, however, does not necessarily correspond to the accuracy of the classifier when presented with a grain that was not used in its construction. This accuracy can be assessed by cross-validation, whereby the training set is randomly partitioned into two subsets. One is used to construct the classifier, and the another is withheld, and then used to test the classifier performance. Note that the species of origin for all the grains is known. The process of partitioning, training and testing is repeated multiple times to validate the cross-fold error in classifying the withheld subset. As the process withholds part of the training data each iteration, the classifiers constructed are (a) not identical to that constructed with all the training data and (b) may mean that the cross-validation error may be an overestimate. In general, this means that the accuracy of the cross-validation classifiers is less than that obtained using re-substitution. Both measures are, however, useful in indicating the classifier performance and can be used in concert to determine the which of the suite of classifiers is of most utility for a given problem.

In developing the methodology for starch grain identification, we further developed the idea of the cross-fold validation for this system. Given we are interested in true-positive, confident predictions we consider a reverse cross-fold validation error. Rather than the usual cross-validation, where a number of grains of a known species are classified and the number of mis-classifications measured, here we look at classification results from all the grains tested. For a given species classification, we then determine how many were not actually that species. The classifier can perform differently for different species. We developed, [7], a measure to encode the dependence of the true-positive rate on the classifier score for the different output species called the positive prediction value, *PPV*. The *PPV* can then be used to choose the classifier and determine the confidence of the results. If the prediction score corresponds to, say, $PPV = 0.9$ for that species, then 90% of the predictions

with scores above this level were correctly classified. Lesser scores, corresponding to lower *PPV* values, mean that the unknown could possibly be the predicted species, albeit with less confidence.

Once the different classifiers are constructed for the training set, using different algorithms and combinations of predictor variables, and assessment can be made of their performance, discarding those with low re-substitution success and high cross-fold validation errors. Further analysis of the performance for the individual species within the training set via the *PPV* and the individual species cross-validation errors then allows us to choose the “best” classifier for the problem at hand.

As part of our studies, once the classifier was optimised and the unknown grains classified, the unknowns were furthermore re-analysed by an expert microscopist, skilled in starch analysis. The quantitative system outlined above does not take categorical or subjective attributes into account when performing the classifications. Subjective features of starch grains include the presence of lamellae and pitting of the grain surface. Taking the cross-fold validation confidence and *PPV* for the predicted species into account, each prediction was deemed to be validated, probable or a false positive. We found a high correspondence with the predicted values, except in some species of similar geometry where subjective features such as lamellae were prominent in some species and not others.

4 Discussion

The methodology has been used to create starch grain classifiers for a number of important archaeological and ethnographic studies. We have identified that the inhabitants of North Queensland rainforests undertook complex processing of some starchy nuts, which were otherwise toxic [7]. Trading patterns and plant use in the highlands of Papua New Guinea have been identified from residue remains on grinding tools. The foraging and consumption of seeds have been investigated from ground stone artefacts from Woomera in South Australia [8].

Whilst the development of the collection of reference grains has been a time-consuming venture, as we are maintaining user input into the collection of the individual ROIs, this library then becomes accessible to multiple studies. If the plant species is present or is an appropriate inclusion in the training set for a given study, these can be deployed immediately. As an evolving collection, this represents a valuable curated resource. It then remains to analyse the newly sources unknown grains—once they have been found in the field!

The methodology is currently being employed to identify starch grains recovered from a variety of archaeological contexts in Australia, Papua New Guinea, and the USA. It is not, however, limited to archaeological samples, and could, for instance, be used to study the provenance of produce such as honey if starch grains in the sample can be identified.

The cultivation of wheat has been integral to human civilisation for millennia. In modern times, different cultivars of wheat produce ‘soft’ and ‘hard’ flours, suitable for

different food preparations. It is planned to characterise the wheat grains producing the different flour qualities and correlate the starch grain features with the kernel strength and hardness profiles [9, 10].

The approach taken here could also be used beyond starch, at, for instance, larger physical scales. Grain shape is a key factor affecting the mechanical properties of granular materials. It has long been of interest in sedimentology (see for instance [11–14]). An accurate classifier of grain shape could be used to quickly and accurately provide information on the contents of seed samples such as wheat, oats, rye and barley. This could be coupled with grain handling operations to recover and sort mixed grains or grade samples. Some work has been done in this area [15] and could be extended to use the approaches of the method developed for starch identification.

References

1. J. Wilson, K. Hardy, R. Allen, L. Copeland, R. Wrangham, M. Collins, Automated classification of starch granules using supervised pattern recognition of morphological properties. *J. Archaeol. Sci.* **37**, 594–604 (2010)
2. B. Lance, J. Field, Intra-taxonomic variability in starch reference collections and the implications for ancient starch studies. *Unpublished Manuscript on file in School of BEES* (UNSW Australia, 2006), pp. 21
3. H. Drolon, F. Druaux, A. Faure, Particles shape analysis and classification using the wavelet transform. *Pattern Recogn. Lett.* **21**, 473–482 (2000)
4. João B. Florindo, André R. Backes, and Odemir M. Bruno, Leaves shape classification using curvature and fractal dimension image and signal processing, ed. by A. Elmoataz et al. in *Proceedings of the 4th International Conference, ICISP 2010*, Québec, Canada, June 30–July 2, 2010. LNCS 6134 (2010), pp. 456462
5. J. Field, Reference collections for starch studies, in *Ancient Starch Analysis*, ed. by R. Torrence, H. Barton (Left Coast Press, California, 2006), pp. 95–113
6. A.C.F. Coster, J.H. Field, What starch grain is that?—A geometric morphometric approach to determining plant species origin. *J. Archaeol. Sci.* **58**, 9–25 (2015)
7. J.H. Field, L. Kealhofer, R. Cosgrove, A.C.F. Coster, Human-Environment Dynamics during the Holocene in the Australian Wet Tropics of NE Queensland: a starch and phytolith study. *J. Anthropol. Archaeol.* **44**, 216–234 (2016)
8. S. Luu, B. Stephenson, A.C.F. Coster, P.T.T. Nguyen, J. Field, Residue analysis of ground stone artefacts from Woomera, South Australia: a pilot study. A Report Prepared for Mr Tim Owen, GML Heritage and the Kokatha Aboriginal Community, Woomera, SA, 30 Sept 2015
9. R.S. Anderssen, R. Haraszi, Characterizing and exploiting the rheology of wheat hardness. *Eur. Food Res. Technol.* **229**, 159–174 (2009)
10. R. Haraszi, A. Juhasz, M. Sissons, M. Rakszegi, L. Tamas, R.S. Anderssen, Rheological hardness index for assessing hardness of hexaploids and durums. *Cereal Chem.* **90**, 430–438 (2013)
11. W.C. Krumbein, F.J. Pettijohn, *Manual of Sedimentary Petrography* (Appleton-Century-Crofts, New York, 1938)
12. J.C. Griffiths, *Scientific Method in Analysis of Sediments* (McGraw-Hill Inc., New York, 1967)
13. G. Lees, A new method for determining the angularity of particles. *Sedimentology* **3**, 2–21 (1964)
14. R. Ehrlich, B. Weinberg, An exact method for characterization of grain shape. *J. Sediment. Petrol.* **40**, 205–212 (1970)
15. J. Paliwal, N.S. Visen, D.S. Jayas, N.D.G. White, Comparison of a neural network and a non-parametric classifier for grain kernel identification. *Biosyst. Eng.* **85**(4), 405–413 (2003)