(SS18) Progress and future direction in automated palynology

Date: August 27
Place: Room 5236 (oral)
Organizers: Katherine Holt & Keith Bennett
Contact email address: k.holt@massey.ac.nz
Purpose: Nearly all branches of palynology require the palynologist to spend countless hours at the microscope recognizing, identifying, counting, measuring and describing pollen. The needs and prospects for automation of palynology were first heralded in the 1960's, but after 50 years of continuing technological advancement, virtually all pollen counting is still done manually.
However, in more recent years there have been an increasing number of groups working towards automation of various aspects of palynology and the field is starting to see some genuine progress. In this session we will draw together those groups who are actively involved in developing methods to partially or completely automate any aspect of the pollen-counting process. The primary purpose is for those involved in automated palynology to share ideas and advancements between groups and to expediate the development of a robust, reliable automated pollen counting and classification system. Of equal importance, will be presentation of the state of play of automated palynology to all branches of the palynology community, and to demonstrate that automated palynology is achievable and has the potential to deliver significant benefits in terms of time savings, more accurate counts, repeatable counts, larger counts, and high taxonomic resolution.
All branches of palynology stand to benefit from partial or complete automation of the palynology process, so the proposed session will be of interest to all those who attend the conference.

Oral Presentation

Aug. 27 [AM1] Room: 5236

Chair: Katherine Holt

9:00-9:20 **Principles, potential and methodology for automated pollen counting** SS18-O01 (27)

K.D. Bennett

9:20-9:40 **Development of automatic airborne pollen-sorting counter using pollen auto-fluorescence** SS18-O02 (338)

Kotaro Mitsumoto, Katsumi Yabusaki, Hideki Aoyagi

9:40-10:00 **A bias-optimized layered machine-learning framework for high-diversity pollen samples** SS18-O03 (420)

Surangi W. Punyasena, David K. Tcheng

10:00-10:20 **Testing the analysis of airborne pollen samples with an automated trainable system** SS18-O04 (277)

Elisabeth Levac

Aug. 27 [AM2] Room: 5236

Chair: Keith Bennett

10:50-11:10 **Accelerating pollen phenomics with the Classifynder** SS18-O05 (263)

Ryan Lagerstrom, Yulia Arzhaeva, Leanne Bischof, Simon Haberle, John La Salle, David Lovell, Andrew Young

11:10-11:30 **Applying an automated palynology system to counting Quaternary fossil pollen: Can it be done?** SS18-O06 (193)

Katherine Holt, John Flenley, Robert M. Hodgson, Colin Plaw, Ken Mercer, Kevin Butler

SS18-O01 (27)
Principles, potential and methodology for automated pollen counting

K.D. Bennett

Queen's University Belfast, Northern Ireland, UK, k.d.bennett@qub.ac.uk

Counting pollen manually is intensive of time, and thus costly, but also highly skilled and carried out to a high standard by many analysts around the world. Automating the process has long been an aspiration, offering the potential to free the time of analysts, but limited by lack of sufficient computer power. A system now available (Classifynder: http://www.classifynder.com/) combines mechanical engineering (to drive the microscope stage) with software (image analysis) enables automated pollen counting. The requirements for using such a system have implications for the way that pollen analysts work. The ultimate goal must be that pollen counts obtained by an automated system are acceptable in publications, in the same way that manual pollen counts are accepted. Achieving that requires testing to show that automated counts are valid and can be used with confidence without manual checking. It may be the case that an automated count does not return the same result as a human count because of differences in the way that pollen grains are recognised by the eye and an image analysis system, but both are approximations to the true (possibly unknown and unknowable, because of breakage, partial concealment) count on a slide. Freely available image libraries are needed so that image analysis software at all installations use the same basis for identifications. Differences may exist in what can be identified with image analysis relative to the eye, with each having strengths with certain types of morphology (the eye with counting furrows and pores; image analysis with discerning fine differences in surface patterning). Image analysis is expected to be very strong with long counts, made without breaks, and completely consistently. It is likely to be years before automated systems come into routine use. Early applications may include very high counts or very large numbers of levels for a small number of pollen types, where this is relevant to an investigation. Instead of counts of 1000 grains at 100 levels for 100 pollen types in a Holocene sequence, it will be possible to count 10000 grains per level, or 1000 grains at 1000 levels, for the ten most important types. The analyst may be able to move to levels of quality in counts that have hitherto been impossible. The possibility is ready for exploitation.

Keywords: palynology, automated pollen counting.

SS18-O02 (338)
Development of automatic airborne pollen-sorting counter using pollen auto-fluorescence

Kotaro Mitsumoto[1], Katsumi Yabusaki[1], Hideki Aoyagi[2]

[1] *Kowa Company, Ltd, Japan, k-mitumt@kowa.co.jp*
[2] *Faculty of Life and Environmental Sciences, University of Tsukuba, Japan*

The information about the coming airborne pollen concentration is helpful for preventing pollinosis. We have tried to make pollen counting methods, instruments and integrated systems that can provide accurate and real-time forecast of pollen dispersion in local area. Historically pollen counting has been performed under microscopic observation. In these several years, some groups have developed automatic pollen counting instruments on the basis of the particle counting techniques. These instruments give us immediate information of airborne pollen. However, several improvements such as miscounting dust particles as targeted pollens or non ability of pollen sorting have been pointed out. We developed a new method to classify pollen species by monitoring their auto-fluorescence images excited by UV light. The pollen size and the ratio of blue to red pollen auto-fluorescence spectra (B/R ratio) were calculated by image processing. The pollen size and B/R ratios varied among the species. Besides the scatter-plot of pollen size versus B/R ratio showed that pollen species could be classified by using both parameters. The size of pollen and B/R ratio were confirmed by means of particle flow image analysis and the fluorescence spectra, respectively. These results suggested that a flow system could be applied to classifying and counting airborne pollen grains. We developed a novel flow particle counter that automatically classifies airborne pollen grains. This counter was designed for detecting both scattered light and characteristic fluorescence excited by UV of airborne particle simultaneously in the flow cell. We measured airborne pollens focusing on the following two species, *Cryptomeria japonica* and *Chamaecyparis obtuse*, since they coexist in the springtime and cause serious pollinosis in Japan. The Durham sampling method was used as a conventional microscopic counting, and our novel method for comparing their performances. The daily pollen concentration from the counter was roughly correlated with the results of the Durham sampling method at all study sites. Furthermore, we assembled the improved pollen counter, which equipped with blue laser. The daily pollen concentration of the improved counter was highly correlated with the results of the Durham sampling method (r = 0.925 for *C. japonica*, r = 0.604 for *C. obtusa*). Our results indicate that the new pollen counter has a strong potential for counting and identifying airborne pollen grains in real time.

Keywords: pollinosis, image analysis, flow cytometry, pollen fluorophotometry, *Cryptomeria japonica.*

SS18-O03 (420)

A bias-optimized layered machine-learning framework for high-diversity pollen samples

Surangi W. Punyasena[1], David K. Tcheng[2]

[1] *Department of Plant Biology, University of Illinois at Urbana-Champaign, 505 South Goodwin Avenue, Urbana, Illinois, 61801, USA, punyasena@life.illinois.edu*
[2] *Illinois Informatics Institute, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, 1205 West Clark Street, Urbana, Illinois, 61801, USA*

Contained within the fossil pollen and spore record is one of the most comprehensive histories of terrestrial vegetation and its response to long-term environmental change. However, the inefficiencies inherent in the collection and analysis of palynological data have meant that this record remains one of the few areas of scientific inquiry where automated data acquisition and image analysis have made little headway. Pollen counts are taken much in the same way they were a century ago: by a highly trained expert with a transmitted light microscope. Manual counting has meant that large-scale systems analyses based on fossil pollen data are largely restricted to specific geographic areas where there has been historically enough expertise or enough interest to coordinate data collection efforts – and with a few exceptions, this largely means that such analyses have been restricted to North America, Europe, and other temperate, lower species diversity systems. Development of large spatial and long temporal pollen datasets in the tropics, because of the high taxonomic diversity within palynological samples and the limited number of experts, has lagged behind temperate communities. In order to address the problem of data collection and data

availability for tropical systems, we are developing a radical alternative to traditional palynological counts – an automated counting system capable of dealing with high-diversity pollen samples. Based on layered machine learning algorithms, this high-throughput system has been developed and is being tested using high-resolution images of seventeen years of modern Neotropical pollen rain samples. There are >110 recognized morphospecies in these samples. The system is designed to handle the large, uncompressed image data that we generate (~5 TB/day). Our results are an early demonstration that automation is a feasible and desirable alternative to standard palynological counting practices, particularly in the case of high diversity samples, and should advance the quantity and quality of tropical palynological research, and promote the macrosystems-level analyses of palynological data needed to understand community dynamics and change over long temporal scales.

Keywords: palynology, machine-learning, automation, tropics.

SS18-O04 (277)

Testing the analysis of airborne pollen samples with an automated trainable system

Elisabeth Levac

Bishop's University, Canada, elevac@ubishops.ca

Seasonal allergies are on the rise and symptoms range from a simple irritation of the nasal passage, to asthma and chronic obstructive pulmonary distress. This is why more and more countries are implementing pollen forecasting services to help allergy sufferers reduce their exposure to allergens. Pollen forecasts in North America are usually based on pollen data from previous years, and are not always accurate, or predict high levels of certain pollen types before pollination has actually started. Therefore, reliable pollen forecasting requires in situ monitoring. Pollen monitoring is labor intensive, with the analyst typically spending 3-5 hours at the microscope to analyze each daily sample. There are risks of errors due to misidentification (e.g. with folded pollen grains) or underestimates when pollen concentrations in the samples are too high. Most pollen types are identified at the genus level, while others are identified at the family level. A new automated system called the *Classifynder* can now locate the pollen grains on the slides, capture and store images for each, and classify them. Since the images are stored, it is possible to reclassify the images later and improve the ability of the software to properly identify the pollen grain. Such a system not only significantly decreases the amount of time spent at the microscope, it can also allow discriminating morphologically similar pollen species which is not possible with the human. One prototype of the Classifynder is now being tested at Bishop's University with airborne pollen samples collected with a Burkard pollen samples installed on the roof of Sherbrooke’s Museum of Nature and Science. While silicone oil is the mounting medium giving the best images with this system, it tends to drip under very warm conditions (in sunny weather). We are presently testing other mounting media to go around this problem. Petroleum jelly does not give good results, but glycerin jelly gives acceptable results. A database of the most common local airborne pollen types is being built. The local pollen season in southern Quebec is just starting at the time this abstract is written (April 2012) and further results will be presented at the conference.

Keywords: palynology, pollen monitoring, microscope, Canada, mounting medium.

SS18-O05 (263)

Accelerating pollen phenomics with the Classifynder

Ryan Lagerstrom[1], Yulia Arzhaeva[1], Leanne Bischof[1], Simon Haberle[2], John La Salle[3], David Lovell[4], Andrew Young[5]

[1] *CSIRO Mathematics, Informatics and Statistics, Locked Bag 17, North Ryde NSW 1670, Australia, Leanne.Bischof@csiro.au*
[2] *School of Culture, History and Language, H C Coombs Bldg 9, The Australian National University, Canberra, ACT 0200, Australia*
[3] *CSIRO Ecosystem Sciences, Black Mountain Laboratories, Clunies Ross Stree, Black Mountain ACT 2601, Australia*
[4] *CSIRO ICT Centre & CMIS, GPO Box 664, Canberra, ACT 2601, Australia*
[5] *CSIRO Plant Industry, Black Mountain Laboratories, Clunies Ross Street, Black Mountain ACT 2601, Australia*

With support from the Atlas of Living Australia, and in collaboration with Massey University, researchers from CSIRO and ANU have been investigating how Massey University's Pollen Classifynder can accelerate our understanding of pollen and its role in nature. Given the importance, abundance and diversity of pollen, it is vital to exploit assistive technologies like the Classifynder to enable acquisition and analysis of pollen samples. It is also vital that we understand the strengths and limitations of automated systems so that they can be used (and improved) to complement the strengths and limitations of human analysts to the greatest extent possible. This presentation will review some of our initial experiences with the Classifynder system and our exploration of alternative classifier models to enhance both accuracy and interpretability. We emphasise that overall classification performance depends on several-sometimes interacting-factors, including the problem domain; the sample of the population of interest; the representation of that sample; and the classifier model(s) employed. Our initial experiments in the pollen analysis problem domain have been based on samples from ANU's pollen reference collection (2890 grains, 15 species) and images bundled with the Classifynder system (400 grains, 4 species). These samples have been represented using the Classifynder feature set. In addition to Classifynder's native neural network classifier, we have evaluated linear discriminant and decision tree classifiers on these data with encouraging results. Our hope is that these findings will help enhance the performance of future releases of the Classifynder and other systems for accelerating the acquisition and analysis of pollen samples.

SS18-O06 (193)

Applying an automated palynology system to counting Quaternary fossil pollen: Can it be done?

Katherine Holt[1], John Flenley[1], Robert M. Hodgson[2], Colin Plaw[2], Ken Mercer[2], Kevin Butler[3]

[1] *Institute of Natural Resources, Massey University, PB 11-222 Palmerston North, New Zealand, k.holt@massey.ac.nz*
[2] *School of Engineering and Advanced Technology, Massey University, PB 11-222 Palmerston North, New Zealand*
[3] *School of People, Environment and Planning, Massey University, PB 11-222 Palmerston North New Zealand*

We have previously described a complete pollen location and classification system called the ‘Classifynder’ (Holt et al., 2011). This system employs robotics and image processing to locate and image pollen on slides, coupled with a neural network-based classifier to identify the pollen in the captured images. Previous tests of this system have involved comparing counts by both humans and

the Classifynder of slides containing fresh pollen of a limited number of taxa. The results of these tests were particularly positive, with machine counts very similar to human counts, and the Classifynder proving to be more consistent than the human palynologists. In this paper, we have attempted to test the accuracy of the Classifynder system in locating and classifying Quaternary fossil pollen, and assess the limits posed by the presence of broken and deformed pollen and debris. It is the presence and often considerable abundance of these types of objects which present some of the biggest challenges to automated palynology. Slides of pollen extracted from a New Zealand lake sediment sequence using standard pollen extraction procedures were counted repeatedly by several experienced palynologists and by the Classifynder. Raw counts from the Classifynder were scrutinized by a human palynologist, with any incorrectly classified images reassigned to their correct taxon, and debris images deleted. The degree of similarity/difference between the human and user-adjusted Classifynder counts is used as the basis for assessing the accuracy of the Classifynder system at counting fossil pollen. Results demonstrate that the accuracy of the neural network-based classifier can be quite variable, caused partly by misclassification of deformed or broken grains. However final Classifynder counts of the fossil samples matched very closely with the human counts, usually within one standard deviation of the mean. This indicates that although the system is misclassifying some pollen grains, it is still recognizing all pollen objects on the slide, and with checking by a human palynologist, can produce counts of fossil pollen to the same degree of accuracy as a human palynologist, but with much less time and effort required from the palynologist.