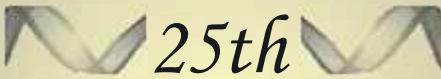


LASR

Leeds Annual Statistical Research
Workshop

2006



25th

A n n i v e r s a r y

interdisciplinary statistics & bioinformatics

4th - 6th July 2006

LASR WORKSHOP - OUR MISSION

The Leeds Annual Statistical Research Workshop (LASR Workshop) is a well established international conference. In 2006 it celebrates its 25th year.

To mark this anniversary, we have organised a particularly special event:

- We examine the role of statistics in interdisciplinary research
- We showcase new, innovative statistical methods and applications in bioinformatics
- We celebrate the history and achievements of the LASR Workshop over a quarter of a century

Year on year LASR attracts participants from Universities, industry and research bodies from all over the world. The workshops usually extend over three days with talks by internationally renowned experts. Over recent years, workshop



Image courtesy of Charles Trevelyan

themes have reflected the growing, but not exclusive, departmental interest and expertise in bioinformatics, image and shape analysis, and wavelets.

The LASR's mission is to:

- Bring together world-class scientists and statisticians
- Encourage cutting-edge interdisciplinary research
- Sustain and develop our workshops

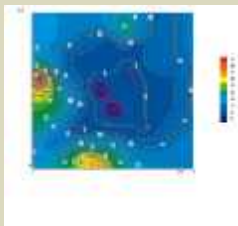
Many of the topics addressed by the LASR Workshop are connected to the theme of geometric statistics - the interpretation of information with a geometric flavour such as:

- Spatial Statistics
- Image Analysis
- Shape Analysis
- Bioinformatics

SPATIAL STATISTICS

Consider information gathered at different geographical locations, e.g. pollution levels at different monitoring stations, or rates of illness in different counties.

A common property of spatial data is "spatial continuity", which means that measurements at nearby locations will tend to be more similar than measurements at distant locations.



Estimated map of a bauxite ore from sample locations

A key problem of interest in this area involves the prediction of the process at new sites, given its values at a collection of data sites.

This type of work can be extended to spatial-temporal statistics, where the information is gathered repeatedly at different times.

Spatial-temporal modelling allows prediction not only at new locations but at different times. For example, making forecasts.

IMAGE ANALYSIS

With the development of modern computers the collection of images has become a straightforward and powerful tool for recording information. However, the interpretation of images in terms of their constituent parts requires sophisticated computational and statistical models. For example, consider an image containing a hand.

In order to make objective use of this information, the computer needs to extract the outline of the hand from the image, to recognize it as a hand, and perhaps to compare it to other hands to see how it differs.



Palm recognition for secure identification

The effects of time can be included in image analysis. We can consider how an image changes over time, such as the changing appearance of a hand or face as an individual grows.

SHAPE ANALYSIS

A definition of "shape" is the information about an object that is not affected by size, location, or orientation of the object.

For centuries, one of the major techniques in biology for separating similar animals and plants into different groups has involved the comparison of the shapes of the specimens. This field is known as morphometrics.



In modern times this task has been facilitated by the ability to record easily the positions of "landmarks" on the specimens in two or three dimensions, and new methods of statistical analysis have been developed to deal with this type of data.

Stereolithographic models of human skulls

BIOINFORMATICS

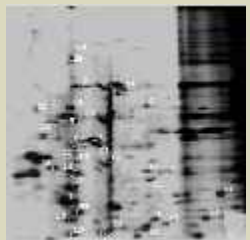
The increasing understanding of the human genome has been one of the great success stories in biology over the past 50 years. DNA in the genome contains "genes" which contain the information needed to build proteins, which in turn carry out the functions of life.

Many deep questions remain, and statistical ideas will play a key role in understanding genes and proteins. Some of the key problems in bioinformatics are:

- (i) measuring when different genes are expressed at different stages in a life cycle
- (ii) measuring the differences between genes for different species.

The shape of a protein is one of the key properties which determines how a protein works. The advances in geometric statistics and particularly shape analysis from the LASR workshops enables us to explore new ways of modelling protein structure.

A key challenge in adapting shape analysis to protein bioinformatics is the lack of natural labelling on the proteins.



Electrophoretic gel with annotated landmarks

BIOINFORMATICS - THE LEEDS CONNECTION

Astbury considered that there might be no radical difference between fibrous and crystalline proteins as one of his great contributions. Overall, Astbury *'influenced everybody's thinking about large biological molecules'* and *'was the father of all those who since then interpreted other types of fibrous structure...and who can recognize types of twist from the pattern on blurs on rather obscure fields.'*

However, there were two principal discoveries missed by Astbury that, some say, could have been helped by seeking collaboration.

The Alpha-Helix For Proteins

Why He Missed It

Like others in his field, he was looking for rational helical models. He did not realise/know that the peptide bond was planar and rigid. His models were too flexible whereas those of Linus Pauling had a much more limited conformational flexibility making it easier to construct the correct model.



W.T. Astbury
1898 - 1961
Professor of Biomolecular Structure
University of Leeds

Why He Almost Made It

Astbury repeatedly stated in public that a helix was the most likely structure for linear polymers.

DNA As The Genetic Material & The Double-Helical Structure

Why He Missed It

Astbury clung to the idea that the genetic material must be protein rather than nucleic acid. He was not looking for a molecule which could carry the genetic message and replicate, nor did he realise that the structure of DNA fibres is dependent on humidity, consequently his diffraction photographs were mixtures of the A and B forms, thus difficult to interpret.

Why He Almost Made It

His *'pile of pennies'* model for DNA was OK as far as it went. He clearly had the concept of a linear message in his mind when he wrote of *'the long scroll on which is written the pattern of life'*.

Perhaps the most important lesson we can learn from Astbury's experience is the importance of both independent *and* collaborative research. The University of Leeds strongly encourages interdisciplinary research. The LASR team within the School of Mathematics is committed to inter-disciplinary research in bioinformatics with biology; biophysics; molecular medicine; universities from around the world and industry.



Very early X-ray picture (1930s) of collagen fibres (frog's toe tendon) from Astbury's laboratory.

Such X-ray images were instrumental in DNA discoveries by Watson and Crick.

Key Speakers 2006

| | |
|-------------------------------------|---|
| Brian Athey | University of Michigan |
| Julian Besag, FRS | University of Washington |
| Clive Bowman | GlaxoSmithKline |
| Fred Bookstein | Universities of Vienna, Seattle |
| Sir David Cox, FRS | Nuffield College, Oxford |
| Walter Gilks | University of Leeds |
| Chris Glaseby | Biomathematics & Statistics Scotland |
| Julia Goodfellow | Biotechnology and Biological Sciences Research Council |
| Thomas Hamelryck | University of Copenhagen |
| John Kent | University of Leeds |
| Alf Linney | University College of London |
| Kanti Mardia | University of Leeds |
| Janet Thornton, FRS, CBE | European Bioinformatics Institute |
| David Westhead | University of Leeds |

Conference Opened by

Professor Michael Arthur, Vice-Chancellor, University of Leeds

Celebration Dinner at the Royal Armouries

Attended by

The Right Worshipful the Lord Mayor of Leeds, Cllr Mohammed Iqbal

After Dinner Speaker

Julia Goodfellow, Chief Executive

Biotechnology and Biological Sciences Research Council (BBSRC)

LASR 2006 Sponsored by



GlaxoSmithKline

Professor Kanti V. Mardia, Senior Research Professor
Department of Statistics
School of Mathematics
University of Leeds, Leeds LS2 9JT
Email: k.v.mardia@leeds.ac.uk
Tel : +44(0)113 343 5100
Fax: +44(0)113 343 5090

www.maths.leeds.ac.uk/lasr2006