Inference for the physical sciences

Nick S. Jones and Thomas J. Maccarone

Supplementary data	"Audio Supplement" http://rsta.royalsocietypublishing.org/content/suppl/2013/01/03/ rsta.2012.0493.DC1.html
References	**This article cites 19 articles** http://rsta.royalsocietypublishing.org/content/371/1984/201204 93.full.html#ref-list-1
Subject collections	Articles on similar topics can be found in the following collections astrophysics (42 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Introduction

CrossMark
click for updates

Author for correspondence:
Nick S. Jones
e-mail: nick.jones@imperial.ac.uk

Inference for the physical sciences

Nick S. Jones[1] and Thomas J. Maccarone[2]

[1]Department of Mathematics, Imperial College, London SW7 2AZ, UK
[2]School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK

There is a disconnect between developments in modern data analysis and some parts of the physical sciences in which they could find ready use. This introduction, and this issue, provides resources to help experimental researchers access modern data analysis tools and exposure for analysts to extant challenges in physical science. We include a table of resources connecting statistical and physical disciplines and point to appropriate books, journals, videos and articles. We conclude by highlighting the relevance of each of the articles in the associated issue.

1. Introduction

We began discussing the plan for this meeting and issue some years ago when we realized that one of us had some interesting data challenges in need of a new analysis approach, and the other had developed some signal-processing machinery, and was looking to expand the range of its applications. This volume is about increasing access to modern data analysis by physical scientists and giving entry pathways for analysts into physical topics.

Statistical analysis of datasets remains a hurdle that can prevent optimal science from being done. One prominent recent example concerned the investigation into the Climate Research Unit at the University of East Anglia. The group had been accused of research fraud by climate change deniers. An independent review led by Lord Oxburgh concluded that the scientists had behaved in an entirely ethical manner, but did level the criticism: 'we cannot help remarking that it is very surprising that research in an area that depends so heavily on statistical methods has not been carried out in close collaboration with professional statisticians', with part of the motivation for the criticism being that the group had not always used the most modern or the

most appropriate statistical methods for the problems they were trying to solve (although we stress that the report also noted that there was no clear evidence that different results would have been obtained with better methods) [1].

In recent years, collaborations with statisticians (whether located in departments of statistics or else-where) have become common for researchers dealing with large datasets in fields such as biology or the social sciences, where the mathematical content of undergraduate degree programmes can be modest. Physical scientists generally have the underlying mathematical background to apply the most sophisticated data analysis techniques themselves, but often struggle with learning where to look for better approaches than the ones to which they have become accustomed.

Most physical scientists are familiar with techniques at the level of *Numerical recipes* [2] (which goes well beyond just covering data analysis) and, of course, many far exceed it. *Numerical recipes* was first written in 1986, and while it has been updated since then, the goal of its authors was not to write a book that would explain all the modern best practice, but to write a book that would be accessible, and would bring practice among scientists to a level only a decade or so behind the state of the art.

'Our 1980s goal for NR, overtly discussed by the authors, was to provide a means for bringing average practice into the 1960s - or 1970s in a few areas (FFT)', wrote Press & Teukolsky [3], roughly a decade after the book had first been published. They had previously argued, in a manner devoid of hyperbole, that 'average practice' by physicists and astronomers in the academic departments they knew best—which were and are among the elite American institutions—was generally what applied mathematicians were doing in the 1920s. Speaking broadly, apart from the use of the fast Fourier transform, the undergraduate courses on numerical methods in a typical British physics department are only slightly more advanced now than what Press & Teukolsky [3] describe—there simply is not room in the curriculum to cover everything we would like to teach.

One of our goals for this issue was, in essence, to provide a resource that would help bring practicing physical scientists forward a few more decades. Note that the source lectures themselves are available online [4], with slides and resources available at http://www2.imperial. ac.uk/~njones/sigprocandinference.htm. Working with a shorter volume than *Numerical recipes*, and trying to pack the material into a 2 day meeting, of course, that we could not hope to take an approach either as pedagogical or as comprehensive as was taken by the authors of *Numerical recipes*. What we can do, and what the authors of the articles here have (heroically) done, is to collect basic descriptions of a range of techniques with promise for application in physics, astronomy, geophysics and biophysics, but which are not necessarily being applied very widely in those areas. At the same time, we wished to bring data analysts up to speed on a range of problems on which they haven't been working, but to which they might bring key new insights and approaches.

We had contributors from departments of statistics, mathematics, physics, computer science, engineering, geophysics, earth science, oceanography and biology, and contributors from both industry and academia. The vast majority of the participants knew only a small fraction of the other participants. As the meeting progressed, we were delighted to see new collaborations starting on the spot, involving people who had never heard of one another before.

2. Engaging with new tools in data analysis

This issue is specifically about attempting to increase the flux of new method into the physical sciences. As noted, in the UK it is, unfortunately, often the case that undergraduates leave their physics degrees with only a couple of lectures in statistics: enough error analysis to do their experiments. Combine this with the UK's emphasis on research PhDs, and a gappy statistical education is possible.

While running the risk of being obvious, we hazard a few suggestions for such a scientist (beyond reading this issue and listening to the lectures [4] http://www2.imperial.ac.uk/~njones/sigprocandinference.htm). Machine, or statistical, learning is now a refined and highly applicable discipline; so, if investigating colleagues to serve as potential collaborators, it is worth considering those in engineering and computer science departments, as well as those in statistics. Machine/statistical learning and data mining have particularly embraced online videos; so there is now a good back-catalogue of expository lectures on videolectures.net (including numerous machine learning summer schools and components of the major conferences). In terms of good books to start with, *All of statistics* [5] is a good one-smallish-volume introduction to statistics that has a modern orientation. We believe that Bayesian approaches can be more appealing/natural to physical scientists and so suggest exploring the Bayesian inference/machine learning literature. There have been a number of active physicists in this area, and it can be stylistically more penetrable than some of the more conventional statistical literature [6–9]. As you'll see from table 1, a number of these books have free pdfs to help you make a suitable choice. Mackay [8] is notably accomplished in connecting diverse literatures and is a very natural read for physicists; however, for a straight read about inference, we suggest Bishop [6], Barber [7] and Hastie *et al.* [15]. Sivia & Skilling [9] is written with physicists in mind and should also be considered. Our issue has a bias towards topics in more general inference than time-series analysis (but note [34–38]), but we provide in table 1 some books, a number of which are free to browse. Brockwell & Davis [12] and Shumway & Stoffer [13] are standard first sources for time-series analysis, Vetterli *et al.* [10] an established (and free) read from a signal-processing perspective, and readers might enjoy a modern introduction in Mallat [11]. We make extensive journal suggestions in table 1, but we will specifically pick out the *IEEE Signal-processing magazine*, which has a good level of introductory articles that connect the reader with the state of the art. Table 2 provides a brief glossary to partly help translate terminology.

3. Missing topics in this issue

This issue must, by construction, be incomplete in character. There are thus a number of very lively areas that are not included, of which we now give a summary that is also very incomplete. A major missing area is the study of sparsity/compressed sensing—this is very fast moving, but a magazine introduction can be found in Candès [39], and the work by Mallat [11] is aligned with this literature. A core problem of modern data analysis is that though we have large amounts of data we have relatively few distinct instances of that data (see comments in Bishop [40]): we know many features about each object (p features), but have relatively few objects (n objects) so $p \gg n$. This challenge of trying to reason about a few data points in a high-dimensional space is discussed in a recent issue of *Phil. Trans. R. Soc. A*, and the associated programme at the Newton Institute has online talks [41,42]. In this material, connections to the sparsity programme are made explicit.

Physical scientists and some mathematicians can seem to have a notion of a mathematical model that diverges from what similarly mathematical scientists in statistics, engineering and computer science call a model. One might oppose the rich models which include microscopic detail (or have principled methods for eliminating it) made by physical scientists to comparatively thin, flexible, statistical models where part of their construction is their suitability for diverse data types and the robust characterization of their statistical properties. Despite this apparent difference, most models can be used to simulate data, and these results can be compared with observations. An area that is suited to physical scientists wanting to reason statistically about rich models is approximate Bayesian computing (ABC) [43]: this attempts to substitute the construction of a likelihood with simulations from that model. We provide no explicit discussion of ABC in this issue, but attempts to fit richer models do appear in Raue *et al.* [44], Stathopoulos & Girolami [45], Sambridge *et al.* [46] and Cornish [47].

Finally, this volume does little to engage with the challenge of giant datasets and the fast algorithms, that they require. Data mining is particularly concerned with efficient algorithms, and we direct the reader to relevant parts of table 1. We note, of course, that alternative processing

Table 1. This table is designed to help researchers take the first steps in engaging with each others' disciplines. What we include is subjective, flawed and definitely incomplete, but we hope it has some value. We emphasize the core journals and meetings associated with each discipline and highlight free resources (data mining and machine learning conferences appear on videolectures.net, and free pdfs and resources are highlighted with asterisks). IEEE, Institute of Electrical and Electronics Engineers; DMKD, Data mining and Knowledge Discovery; TKDE, Transactions on Knowledge and Data Engineering; ICASSP, International Conference on Acoustics, Speech, and Signal Processing; NIPS, Neural Information Processing Systems; ICML, International Conference on Machine Learning; AIStat, Artificial Intelligence and Statistics; KDD, Knowledge Discovery and Data Mining; ICDM, International Conference on Data Mining; ASA, American Statistical Association; UCI, University of California, Irvine; IET, Institution of Engineering and Technology; IUGG, International Union of Geology and Geophysics; ICSB, International Conference on Systems Biology; IDL, Interactive Data Language.

discipline	example journals	major meetings/interface meetings	example introductory texts/interface texts	canonical resources
signal processing and time-series analysis	IEEE Signal Processing Magazine	IEEE ICASSP	[10]*	R packages*
	IEEE Transactions on Signal Processing	IEEE Statistical Signal Processing Workshop	[11]	MATLAB Signal Processing toolbox
	IEEE Transactions on Information Theory		[12]	
	Journal of Time Series Analysis		[13]	
			[14]	
machine learning	Journal of Machine Learning Research	NIPS*	[15]*	mlcomp.org*
	Machine Learning	ICML*	[6]	UCI Machine Learning Repository*
	Advances in Neural Information Processing	AIStat*	[7]*	
			[8]*	
data mining	DMKD	KDD*	[16]	Weka [18]*
	TKDE	ICDM	[17]	
statistics	Journal of the American Statistical Society	Joint Statistical Meeting, ASA	[5]	R packages*
	Annals of Statistics	Royal Statistical Society Conference	[19]	MATLAB statistics toolbox
	Annals of Applied Statistics		[20]	
	Journal of Royal Statistical Society A,B,C			
assorted physics	Reviews of Modern Physics	American Physical Society Meetings	[9]	
	Physical Review A,B,C,D,E, Letters, Nature Physics	Cosmostats		
		MaxEnt meetings		

(Continued.)

4

Table 1. (*Continued.*)

discipline	example journals	major meetings/ interface meetings	example introductory texts/ interface texts	canonical resources
astronomy and astrophysics	Monthly Notices Of The Royal Astronomical Society	Statistical Challenges in Modern Astronomy	[21]	IDL
	The Astrophysical Journal	Astronomical Data Analysis Software and Systems	[22]	Astronomy Users Library [24]*
	Astronomy and Astrophysics	International Astronomical Union General Assembly	[23]	StatCodes [25]*
		American Astronomical Society		
atmospheric physics	Journal of Atmospheric and Solar-Terrestrial Physics	American Meteorological Society Annual Meeting	[26]	
	Journal of Geophysical Research – Atmospheres		[27]	
	Bulletin of the American Meteorological Society			
geophysics	Journal of Geophysical Research	American Geophysical Union Fall Meeting	[28]	
	Geophysical Journal International	European Geosciences Union General Assembly	[29]	
	Geophysical Research Letters	IUGG Conference on Mathematical Geophysics	[30]	
	Geophysics			
	Physics of the Earth and Planetary Interiors			
	Solid Earth			
biological physics	Biophysical Journal	Biophysical Society	[31]	
	Physical biology	ICSB	[32]	
	Molecular Systems Biology		[33]	
	IET Systems Biology			

Table 2. A partial glossary that helps connect physical, statistical and computer science/engineering terminology. It is a modified version of a similar table in Wasserman [5].

estimation/learning	approximating unknown quantities
covariates/features/measurement coordinates	the distinct dimensions in which a data point, X, lies e.g. a data point might have three space and a time coordinates
data/training set	a set of data points, $\{X1, X2, X3, . . .\}$ each of which is associated with a label $\{Y1, Y2, Y3, . . .\}$. The data points are sometimes called the independent variable (e.g. measurements at a particular space–time coordinate) and the label the dependent variable (e.g. a temperature or a binary value at that coordinate).
classification/supervised learning	given data points, each of which is associated with a symbolic label, predict the labels of future data points
regression	given data points, each of which is associated with a number (or set of numbers), predict the labels of future data points
clustering/unsupervised learning	exposing structure (sometimes groups) in sets of unlabelled data points

architectures can also offer speedups—e.g. graphical processing units (note their recent use as a tool in statistics [48]).

4. Advice on how to read this issue

(a) General interest for Bayesian inference

This issue was assembled with the (experimental) physical scientist in mind. We suggest that statistical ideas can be more palatable to physical scientists if couched in a Bayesian context, partly because some of the assumptions that are bundled up in frequentist estimators are made apparent when one is forced to choose priors (for an engaging video lecture about the vices and virtues of Bayesianism, see Jordan [49]). We thus begin with three articles from a Bayesian statistical learning perspective. In our opinion, all of them are excellent reading for most physical scientists, independent of their background (we note that two of these three authors have PhDs in physics). Each presents a mix of introductory content and current tools. Bishop [40] first introduces Bayesian machine learning and explains model construction in the context of graphical models. The author then explains how there are automatic tools for converting a given model into a practical tool for inference. Then, Ghahramani [50] gives another introduction to Bayesian inference and then considers Bayesian non-parametric inference in detail. These methods will likely be qualitatively different from those that a typical physicist is exposed to while being flexible and intuitive. Roberts *et al.* [34] considers in detail one class of Bayesian non-parametrics: Gaussian processes. This paper gives both a readable introduction to the topic and discusses details of their use in the context of time-series data.

(b) Specific interest

The rest of the issue now depends on the background of the reader. The next six papers consider statistical tools for particular problems and are discussed in more detail below [35,36,44,45,51,52]. The remainder, but one, of the papers consider (statistical) challenges provoked by particular physical settings in: Astronomy (time-series analysis) [38], geosciences [46], gravitational

waves [47], systems biology [53], biological physics (time series) [37] and atmospheric physics [54]. The last paper of the issue [55] addresses the challenging and exciting problem of inferring an appropriate, interpretable, dynamical model (for symbolic time-series data), given relatively weak constraints on the structure of that model.

(c) Generic methods for specific challenges

We begin with a paper [44] that considers the challenge of parameter identifiability: how one can identify the character of our posterior uncertainty over parameters (given priors, data and model structure) and use this to motivate the collection of more experimental data. The next two papers are not explicitly Bayesian [51,52] but consider unsupervised approaches to data. The first [51] gives an update on the independent components analysis method (a generalization of principal components analysis) that, though now a canonical tool for exposing structure in our data in numerous areas, is perhaps unfamiliar to many physical scientists. The second [52] outlines the utility of simple methods to approximate the Kolmogorov complexity. Using standard compression tools turns out to be a very successful means of quantifying the similarity of two data objects: opening the prospect of finding natural organizations of large sets of data with very little domain specific knowledge. We next have two papers that consider oscillatory signals. The first [36] considers statistical challenges in rotary component analysis (a method of relevance to the rotating flows one observes at sea or in the atmosphere). The second [35] specifically engages with the multi-variate setting, with applications ranging from climate data to medical physics. Our final paper on generic methods [45] considers a new class of Markov chain Monte Carlo algorithms that find fruitful application to inference challenges in chemistry and systems biology.

(d) Tools provoked by specific disciplinary challenges

Our other key goal for the meeting was to present some problems from the physical sciences for which signal processing and inference presents an ongoing challenge, with presenters giving the state of the art in data analysis among practising scientists in those fields and reviewing outstanding problems.

We chose astrophysics and geophysics as two areas for special consideration. Astrostatistics is a rapidly growing field of research. Its datasets can be quite large, and they can also present challenges distinct from those in almost all other areas of the natural sciences, with Poisson variations, which in the small number statistics regime are non-Gaussian, often representing a larger fraction of the total error budget than in other areas. In many cases, astronomers study transient events, so that an experiment cannot be repeated in order to improve the signal to noise, so extracting as much information as possible from the signal that nature presents is of utmost importance. Astronomical time series from ground-based observations are often limited by sampling patterns interrupted by cloud cover. The earth sciences present similar challenges with regard to key information often coming in short bursts of information (e.g. earthquakes), and the time and spatial sampling of events often being limited by the location of detectors—working underwater, for example, can be far more costly than working on the ground.

In astrophysics, the representative problems we chose for the meeting were ones of time-series analysis motivated, in part, by testing whether Einstein's general theory of relativity properly describes gravity. One of these topics is the detection of gravitational waves, which are expected to produce small, short-lived, not-quite-periodic fluctuations, against what may be a background with its own (sometimes) nonlinear variability [47]. The challenge is twofold: first the signals must be detected; next one must attempt optimal extraction of the parameters of the astrophysical event that produced the gravitational waves. The article outlines the state of the art of methodology for finding and characterizing gravitational waves, the basics of the instrumentation used, the challenges it presents and the physics of gravitational radiation itself. The other astrophysical topic we considered is time-series analysis of photons from celestial objects [38].

In Vaughan [38], motivations for doing time-series analysis are presented, focusing on examples involving accretion discs around black holes. The key challenges involve understanding nonlinear and non-stationary time series, dealing with sparse or irregular sampling of time series and attempting to estimate the transfer function between two moderately well-correlated time series.

In geophysics, the meeting included presentations on the state of art in using seismic signals to make inferences about the Earth's atmosphere and its interior. In one case [46], the use of earthquakes to make inferences about the Earth's interior was discussed. This considers the best methodology to determine the plausible range of models that could describe the Earth's interior (given the spatial sampling of seismometers, and the spatial distribution of the earthquakes themselves). A key aspect of the approach is to allow the level of complication of the models used to describe the data to be determined by the data themselves. Hedlin & Walker [54] considers the use of seismic signals to constrain the properties of the Earth's atmosphere. In this case, 'infrasound'—low-frequency sound waves, inaudible to humans, which are produced by transient events in the atmosphere, such as sprites—is used. These sound waves couple to the Earth's surface and can be detected by seismometers, and networks of seismometers can thus be used to trace out the properties of the Earth's atmosphere.

We finally address biological physics applications of signal processing and inference. We first consider methods to probe biological networks [53]: in this review, the operation of regulatory networks is explained and the inference of biochemical reactions is placed in their network context. Little & Jones [37] discuss the noisy steppy signals that are produced in the measurements of some cellular and molecular systems: example experimental methods and data types are outlined and tools for probing them are described.

References

1. Report of the International Panel set up by the University of East Anglia to examine the research of the Climatic Research Unit. See http://www.uea.ac.uk/mac/comm/media/press/crustatements/sap.
2. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 2007 *Numerical recipes: the art of scientific computing*, 3rd edn. Cambridge, UK: Cambridge University Press.
3. Press WH, Teukolsky SA. 1997 'Numerical recipes: does this paradigm have a future?' *Comp. Phys.* **11**, 416–424.
4. Signal processing and inference for the physical sciences. See http://royalsociety.org/events/2012/signal-processing-inference/.
5. Wasserman L. 2005 *All of statistics: a concise course in statistical inference*. New York, NY: Springer.
6. Bishop CM. 2007 *Pattern recognition and machine learning*. New York, NY: Springer.
7. Barber D. 2012 *Bayesian reasoning and machine learning*. Cambridge, UK: Cambridge University Press.
8. Mackay D. 2003 *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.
9. Sivia DS, Skilling J. 2006 *Data analysis: a Bayesian tutorial*. Oxford, UK: Oxford University Press.
10. Vetterli M, Kovacevic J, Goyal VK. 2012 Foundations of signal processing and Fourier and wavelet signal processing. See http://www.fourierandwavelets.org/index.php.
11. Mallat S. 2009 *A wavelet tour of signal processing: the sparse way*. Amsterdam, The Netherlands: Elsevier.
12. Brockwell PJ, Davis RA. 2010 *Introduction to time series and forecasting*. New York, NY: Springer.
13. Shumway RH, Stoffer DS. 2010 *Time series analysis and its applications*. New York, NY: Springer.
14. Proakis JG, Manolakis DK. 2006 *Digital signal processing: principles, algorithms, and applications*. Upper Saddle River, NJ: Prentice Hall.
15. Hastie T, Tibshirani R, Friedman J. 2009 *The elements of statistical learning: data mining, inference, and prediction*. New York, NY: Springer.

16. Witten IH, Frank E, Hall MA. 2011 *Data mining*. Burlington, MA: Morgan Kaufmann.
17. Hand D, Mannila H, Smyth P. 2001 *Principles of data mining*. Cambridge, MA: MIT Press.
18. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann, P, Witten IH. 2009 The WEKA data mining software: an update. See http://www.cs.waikato.ac.nz/ml/weka/.
19. Davison AC. 2008 *Statistical models*. Cambridge, UK: Cambridge University Press.
20. Ross SM. 2009 *Introduction to probability and statistics for engineers and scientists*. New York, NY: Academic Press Inc.
21. Karttunen H, Kröger P, Oja H, Poutanen M, Donner KJ. 2007 *Fundamental astronomy*. New York, NY: Springer.
22. Feigelson ED, Babu GJ. 2012 *Modern statistical methods for astronomy*. Cambridge, UK: Cambridge University Press.
23. Wall JV, Jenkins CR. 2012 *Practical statistics for astronomers*. Cambridge, UK: Cambridge University Press.
24. IDL. Astronomy Users library. See http://idlastro.gsfc.nasa.gov/.
25. StatCodes: on-line statistical software for astronomy and related fields. See http://www2.astro.psu.edu/statcodes/.
26. Houghton J. 2002 *The physics of atmospheres*. Cambridge, UK: Cambridge University Press.
27. Warner TT. 2010 *Numerical weather and climate prediction*. Cambridge, UK: Cambridge University Press.
28. Stacey FD, Davis PM. 2008 *Physics of the Earth*. Cambridge, UK: Cambridge University Press.
29. Gubbins D. 2004 *Time series analysis and inverse theory for geophysicists*. Cambridge, UK: Cambridge University Press.
30. Parker RL. 1994 *Geophysical inverse theory*. Princeton, NJ: Princeton University Press.
31. Phillips R, Kondev J, Theriot J. 2009 *Physical biology of the cell*. New York, NY: Garland Science.
32. Nelson P. 2007 *Biological physics*. New York, NY: Worth.
33. Stumpf M, Balding DJ, Girolami M. 2011 *Handbook of statistical systems biology*. Chichester, UK: Wiley.
34. Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S. 2013 Gaussian processes for time-series modelling. *Phil. Trans. R. Soc. A* **371**, 20110550. (doi:10.1098/rsta.2011.0550)
35. Olhede SC. 2013 Modulated oscillations in many dimensions. *Phil. Trans. R. Soc. A* **371**, 20110551. (doi:10.1098/rsta.2011.0551)
36. Walden AT. 2013 Rotary components, random ellipses and polarization: a statistical perspective. *Phil. Trans. R. Soc. A* **371**, 20110554. (doi:10.1098/rsta.2011.0554)
37. Little MA, Jones NS. 2013 Signal processing for molecular and cellular biological physics: an emerging field. *Phil. Trans. R. Soc. A* **371**, 20110546. (doi:10.1098/rsta.2011.0546)
38. Vaughan S. 2013 Random time series in astronomy. *Phil. Trans. R. Soc. A* **371**, 20110549. (doi:10.1098/rsta.2011.0549)
39. EJ Candès MB. 2008 Wakin, an introduction to compressive sampling. *IEEE Sig. Proc. Mag.* **25**, 21–30. (doi:10.1109/MSP.2007.914731)
40. Bishop CM. 2013 Model-based machine learning. *Phil. Trans. R. Soc. A* **371**, 20120222. (doi:10.1098/rsta.2012.0222)
41. Statistical challenges of high-dimensional data. See http://rsta.royalsocietypublishing.org/site/issues/statistical_challenges.xhtml.
42. Statistical theory and methods for complex, high-dimensional data. See http://www.newton.ac.uk/programmes/SCH/index.html.
43. Csillery K, Blum MGB, Gaggiotti OE, Francois O. 2010 Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418. (doi:10.1016/j.tree.2010.04.001)
44. Raue A, Kreutz C, Theis FJ, Timmer J. 2013 Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Phil. Trans. R. Soc. A* **371**, 20110544. (doi:10.1098/rsta.2011.0544)
45. Stathopoulos V, Girolami MA. 2013 Markov chain Monte Carlo inference for Markov jump processes via the linear noise approximation. *Phil. Trans. R. Soc. A* **371**, 20110541. (doi:10.1098/rsta.2011.0541)
46. Sambridge M, Bodin T, Gallagher K, Tkalčić H. 2013 Transdimensional inference in the geosciences. *Phil. Trans. R. Soc. A* **371**, 20110547. (doi:10.1098/rsta.2011.0547)
47. Cornish NJ. 2013 Gravitational wave astronomy: needle in a haystack. *Phil. Trans. R. Soc. A* **371**, 20110540. (doi:10.1098/rsta.2011.0540)
48. GPU. Stochastic simulation for statistical data analysis. See http://www.oxford-man.ox.ac.uk/gpuss/.

49. Jordan MI. Bayesian or frequentist, which are you? See http://videolectures.net/mlss09uk_jordan_bfway.

50. Ghahramani Z. 2013 Bayesian non-parametrics and the probabilistic approach to modelling. *Phil. Trans. R. Soc. A* **371**, 20110553. (doi:10.1098/rsta.2011.0553)

51. Hyvärinen A. 2013 Independent component analysis: recent advances. *Phil. Trans. R. Soc. A* **371**, 20110534. (doi:10.1098/rsta.2011.534)

52. Vitányi PMB. 2013 Similarity and denoising. *Phil. Trans. R. Soc. A* **371**, 20120091. (doi:10.1098/rsta.2012.0091)

53. Thattai M. 2013 Using topology to tame the complex biochemistry of genetic networks. *Phil. Trans. R. Soc. A* **371**, 20110548. (doi:10.1098/rsta.2011.0548)

54. Hedlin MAH, Walker KT. 2013 A study of infrasonic anisotropy and multipathing in the atmosphere using seismic networks. *Phil. Trans. R. Soc. A* **371**, 20110542. (doi:10.1098/rsta.2011.0542)

55. Chattopadhyay I, Lipson H. 2013 Abductive learning of quantized stochastic processes with probabilistic finite automata. *Phil. Trans. R. Soc. A* **371**, 20110543. (doi:10.1098/rsta.2011.0543)