

Statistics are working



The World Food Programme in Nepal

IMAGes continues with illustrations of some of the many statistical developments and applications in Aotearoa.

Modelling earthquakes

An automated statistical model of the stresses in the earth's crust may eventually be incorporated into Geonet, the national array of seismometers monitoring earthquake activity in Aotearoa.

With his colleagues in Earth Sciences at Victoria University, Dr Richard Arnold has developed new and more robust methods of estimating the properties of the ruptures that cause earthquakes.

"An earthquake occurs when two blocks slide against each other in a particular plane, sometimes with an area of only a few square metres. We need to understand the

three-dimensional orientation of the fault plane and the movement in that plane to characterise the tectonic stresses that drive the earthquake.

These stresses are anisotropic – the crust is compressed more strongly in one direction than another.

"A single earthquake gives only a very limited view of what's going on in the crust, but multiple earthquakes accumulate statistical evidence to characterise the stresses: Smaller earthquakes on lesser faults help us understand the stresses the large faults are experiencing. Changes in stress can indicate impending seismic events, including volcanic eruptions." He is automating the model so that it can monitor stress changes routinely across the country.



Survey design

The reliability of population estimates in ecology is always a concern for statisticians. Standard techniques using randomly-placed transect lines in forests or oceans to measure population numbers were known to produce estimates with a high degree of uncertainty, says Associate Professor Rachel Fewster, of the University of Auckland.

Systematic survey designs, where transects start from a certain point and evenly cover the area, were known to be more reliable, how much more was unknown so they were assigned the same poor reliability as random lines. "With systematic designs, the first line determines everything, so we can't use standard statistical theory," she said. Fewster's variance estimation for systematic surveys enabled the improvement to be estimated, and showed from repeated simulations that the systematic estimates can be much more reliable than random estimates.

The result "made a big difference to a Canadian survey of threatened dolphins in fiords", and to surveys of hyenas in the Serengeti. When correctly estimated, the surveys were up to twice as reliable as previously thought. The method was included in the software package Programme Distance, distributed by the University of St Andrews in Scotland.

Targeting aid

Countries including Nepal, Cambodia, Timor-Leste, Bangladesh, the Philippines and Bhutan have benefited from New Zealand statistical expertise in estimating poverty for efficient aid allocation, in work funded by the UN World Food Programme.

Professor Stephen Haslett of Massey University in Palmerston North has worked with teams using generalised linear models combined with the country's own census data to produce detailed small-area maps of poverty levels. "If you just look at national sample surveys, they're not accurate enough. What we're doing is affecting the allocation of \$100m of aid a year in Nepal alone; the money may not be spent anywhere near as well without it. A lot of sample survey results get used to form policy."

He has also designed new water and sanitation surveys of small and medium enterprises in countries such as Azerbaijan and Vietnam for the World Bank, and national surveys monitoring entrepreneurship in Tonga and Uganda.

Finding gene copies

Dr Mik Black at the University of Otago is involved in next generation gene sequencing, "which is driving the Thousand

Genome Project, an international effort to sequence fully the genomes of 2,500 individuals throughout the world. The amount of data is phenomenal – we're all upskilling ourselves on how to handle it."

One aspect of the project is looking at particular gene copy numbers. "We have two copies of each gene, one from mum and one from dad, but we can also have more copies of any gene from ancient variations. Sometimes this has become an increased risk for particular diseases. The analysis is not particularly hard but the volume of data is huge, so finding the multiple copies is difficult."

