
**Journal of agricultural, biological, and environmental statistics, ISSN 1085-7117**
**Volume 20, number 2 (june 2015)**



**Abstract**
In genetic association studies, for each underlying genetic model, there is an optimal test. Usually, the true model is unknown, but knowledge from previous genome-wide association studies for the disease under investigation is available and provides information about the underlying model. We study how to incorporate this prior information about plausible genetic models to achieve better efficiency robustness in genetic association studies. Two procedures are proposed and studied. The first one reduces the set of genetic models using the prior information. The second one derives posterior probabilities for the plausible genetic models using trend tests. Then the trend test based on the posterior mean of the genetic model or a weighted trend test over various genetic models can be employed, similar in spirit to the efficiency robustness approach. In the proposed procedures, the strong Hardy–Weinberg disequilibrium in cases is studied. Simulations are conducted to compare the proposed methods with existing ones. The usefulness of the results for the analysis of data collected in replication studies is investigated, the proposed methods are shown to have better overall performance than existing one, and the methods are applied to analyze the real data.



**Abstract**
The desire to group observations generated from multivariate time series is common in many applications with the goal to distinguish not only between differences in the means of individual variables but also changes in their covariances and in the temporal dependence of observations. In this analysis, we compare ten model-based clustering methods in terms of their ability to identify such features under four scenarios in which data are simulated with varying levels of variable and temporal dependence. To consider these methods in a realistic environment, we focus our analysis on wind data, where observations are often strongly correlated in time, and the dependence of variables is known to vary across different regional weather patterns. In particular, we assess each method's performance when applied to wind data simulated under a realistic two-regime Markov-switching vector autoregressive (VAR) model with a diurnally varying mean. A Gaussian mixture model and a basic Markov-switching model outperform the other methods considered in terms of misclassification rates and number of clusters identified. These two methods and an additional Markov-switching VAR model are then applied to one year of averaged hourly wind data from twenty meteorological stations, and we find that the methods can identify very different features in the data. Supplementary materials accompanying this paper appear on-line.

**On the Residual Plot in a Mixture Model** P. 218-228

Chang Cui, Chang Xuan Mao, Jinhua Zhong...

**Abstract**

The residual plot is a graphical diagnostic that can tell whether a mixture model is necessary. We establish the approximate joint normality of the residual function nonparametrically. We propose an approach to constructing simultaneous asymptotic confidence bands. Our investigation is extended to the empirical log-ratio plot to obtain better visual effect sometimes. The type I error rate is under control to some degree in the proposed procedures with confidence bands based on nonparametric estimation. The proposed procedures are adopted in proteomic and ecological applications. It is found that the probabilities being detected on gels vary over proteins, and that there is no evidence for beetles to be removed with different probabilities.

**Improving the Efficiency and Precision of Tree Counts in Pine Plantations Using Airborne LiDAR Data and Flexible-Radius Plots: Model-Based and Design-Based Approaches** P. 229-257

G. J. Melville, A. H. Welsh, C. Stone

**Abstract**

This paper explores and develops design-based and model-based methods which are suited to sampling strategies developed for LiDAR-assisted plantation inventories. Much of the model-based theory is either recent or adapted from other areas of sampling. The design-based theory extends and adapts previous work to the present situation. The methodology is developed around the increasing utility and precision of LiDAR as a sampling tool for operational forest inventory. Flexible-radius plots, as a means of optimizing the sampling effort, are examined from a sampling perspective. Mixed models are also employed to model the residual variance using specified correlation structures and this includes predictors which utilize local trend such as those employed in kriging. In the design-based setting, model-assisted estimators are used, including regression and ratio estimators. A plot-based survey of a young, single-aged stand located within a *Pinus radiata* plantation in the northern tablelands of New South Wales is used to illustrate the theory. Model covariates are obtained from airborne laser scanning (LiDAR) data.

**Concordance: A Measure of Similarity Between Matrices of Time Series with Applications in Dendroclimatology** P. 258-278

Maryann R. Pirie, Christopher M. Triggs...

**Abstract**

A fundamental assumption in dendroclimatology is that the common signal produced by multiple trees of the same species, growing under similar environmental conditions within the same climate region, relates to changes in the climate within that region in the same way. However, there are concerns that the climate response of young kauri trees may differ to older kauri. As a result, the inclusion of radii from young kauri may weaken the climate signal of the composite chronology. To address this concern, a subset of the data containing tree rings formed when the trees were young was compared to those formed when the trees were old. These subsets contained time series of correlated tree rings aligned by year with start and end years differing for each series. Existing techniques for comparing subsets of time series lack reliability for ragged arrays of dependent non-stationary time series. The concordance method was developed to overcome this. Concordance is a non-parametric method based on bootstrapping that is used to test the hypothesis that two subsets of time series are similar in terms of mean, variance or both. Simulations show that concordance is effective for detecting difference in both the level and scale of two submatrices containing non-stationary and dependent time series. When applied to tree-ring data, the concordance method was able to detect evidence against the subset of young tree rings having the same mean, variance or both than older, more established trees.

**Glacier Terminus Estimation from Landsat Image Intensity Profiles** P. 279-298

Joseph Usset, Arnab Maity, Ana-Maria Staicu...

**Abstract**

Mountain glacier retreat is an important problem related to temperature increase caused by global climate change. The retreat of mountain glaciers has been studied from the ground, but there exists a need for automated methods to catalog glacial change with a wider scope. A viable approach is to extract intensity profiles from Landsat images along the glacial flowline and follow the terminus location over time. We propose a new robust and accurate statistical algorithm to estimate the movement of glacial termini over time from these extracted image intensity profiles. The method we propose first uses regression splines to smooth the image intensity profiles. For each profile, the glacial terminus location is assumed to lie near a point of high negative change in the smoothed profiles. An approximate path of termini locations over time is obtained by an algorithm that seeks to minimize the cumulative first derivative value across the profiles. Spline smoothing is applied to this pilot path for estimation of long-term terminus movement. The predictions from the method are evaluated on simulated data and compared to available ground measurements for the Nigardsbreen, Gorner, Rhone, and Franz Josef glaciers.