**Meeting with Giri 12th November 2018**

The use of catch per unit effort (CPUE) data for fisheries stock assessments assumes that there is a direct relationship between CPUE and relative abundance (i.e. stock biomass). However, a variety of factors influence CPUE: in abalone fisheries this may include the experience and skill of the diver, where the diver is fishing, seasonal variation in abalone availability, and potentially others. Standardisation of CPUE data uses statistical models to remove variation in CPUE from factors that do not relate to stock biomass, thereby increasing the likelihood that the annual pattern is more likely to represent relative abundance.

A variety of modelling techniques are suitable for standardising CPUE data, which are widely discussed in scientific literature. One of the simplest approaches is using simple linear models (LMs), often termed multiple regression. This technique is suitable when CPUE data are normally distributed, or can be logarithmically transformed into a normally distributed form. However, abalone CPUE effort is typically positively skewed whereby the majority of CPUE is distributed at a level dictated by the productivity of the reef, SMU or zone in question, and the remainder contains atypically high values where divers have had very high catch rates. Statistical testing of the distribution of abalone data at the zone, SMU, and SMU by year level indicates that no Victorian abalone CPUE data meet the assumptions of LMs and therefore it is important that models that are able to account for the non-normal residual error structure are used.

The simplest modelling technique to account for the abovementioned distributions are generalised linear models (GLMs), in which the response variable (i.e. CPUE) can take any distribution from the exponential family. As a result, these models are able to account for the non-normal residual error structure apparent in these data and are widely used in fisheries throughout the world. VFA have historically used GLMs for standardising both abalone CPUE and fishery independent survey estimates and these are viewed as being a statistically sound method. However, through further analyses of these data, an improved method has been identified (Giri and Gorfine, 2017). This technique uses generalised linear mixed models (GLMM), which are able to account for more than one source of random variability and the variables that are likely to affect CPUE (other than changes in relative abundance) are able to be either fixed or random effects (detailed below). Further, GLMMs are able to accurately model unbalanced data sets, which is frequently the case in abalone fisheries where there is high variability in the amount of fishing by individual divers both spatially and temporally.

One example of where it is beneficial to incorporate random factors in CPUE standardisation of abalone fisheries is the effect of the diver. The efficiency of individual divers can vary for a variety of reason but a major factor is likely to be their past experience in a particular area. A GLMM is able to model the effect a particular diver within each reef code, SMU, or any other spatial variable to account for this variation and remove its effect from the data so that relative abundance can best be estimated. A GLM cannot account for this variation and models treats the effect of a diver as consistent across all spatial regions. It must be noted that it is possible to include the interaction between diver and reef code, SMU or other spatial component into a GLM, which can capture some of this variation, provided, however, that the data are balanced (i.e. similar amount of dives undertaken in each spatial unit), which is rarely the case. Further, a GLMM is able to incorporate interactions with the quota year and other variables, whereas incorporating an interaction with year in a GLM renders the model unusable for standardisation of CPUE data as the standardised values of year no longer singularly represent the effect of year alone. A GLMM is able to do so without affecting the model and an example of when it may be necessary to do so is the effect of diver and year where it is likely that a divers skill increases as they gain experience and, should the diver remain in the fishery for a very long time, it is foreseeable that their efficiency may begin to decline again with age. Another example pertinent to the Western Zone is the interaction between year and SMU: AVG differentially affected SMUs with some more heavily impacted than others and one SMU (Lady Julia Percy Island) not experiencing the disease. As a result, a variety of closures, structured fishing and legal minimum length changes occurred, all of which affect catch rate through time in a way that is not necessarily attributable to relative abundance. A GLMM can account for the interaction between year and SMU to attempt to remove some of this variation, whereas it is not trivial to do so using a GLM.

In summary, statisticians and fisheries scientists at VFA, have thoroughly investigated the available CPUE data and tested the assumptions required by LM, GLM and GLMM. They have fit all of the abovementioned models and the GLMM with a logarithmic link function and gamma error distribution provides the highest prediction accuracy and the lowest standard error of prediction. The gamma error distribution provided the lowest deviance among a variety of potential candidate distributions appropriate for positively skewed data and was therefore deemed the most appropriate for the available data. Random variables were systematically added to the model under the guidance of fisheries scientists familiar with Victorian abalone fisheries and their effect was assessed by Chi-square change in deviance tests. The lower the deviance, the better the model fits the available data, and the Chi-square test was used to assess the addition of each random variable until they no longer resulted in a significant decrease in deviance, which prevents over-parameterisation of the model. A thorough description of these methods is available in Giri and Gorfine (2017).

The abovementioned technique has been applied in the current stock assessments using GenStat Software and the GLMM model that best describes abalone CPUE data was:

* Fixed effects – QuotaYear, SMU, QuotaYear:SMU
* Random effects – Month, Diver, ReefCode, Month:QuotaYear, Diver:QuotaYear, ReefCode:QuotaYear

The random variables explained 41.26% of variation in CPUE with over half of this (22.51%) being the effect of diver. The remaining variation comprises variation by the fixed effects and any variation that could not be explained by the model.