# Accounting for Missing and Regenerating Pereopods in Individual Weight Estimates for Snow Crab (*Chionoecetes opilio*)

# *with Implications for Survey Biomass Estimates in the Southern Gulf of Saint Lawrence*

# By Tobie Surette

# Introduction

Snow crab (*Chionoecetes opilio*) in eastern Canada are found from the Labrador coast to the edge of the continental shelf south of Nova Scotia. Biomass indices for the assessments of these stocks commonly rely on length-weight relationships which link length-frequency samples to their mass equivalents.

However pereopod (i.e. chelipeds or walking leg) loss significantly impacts the weights of individual crab. The loss of such appendages arise through predation, handling from the fishery, moulting and intra-specific competition. In the absence of individual weight observations, these must either be ignored in the length-weight inferences or the pattern of missing appendages must be observed and used as a predictor. Though reliable and precise mass scales which can be used at sea are currently available, longstanding research surveys with long prior time series of length-frequency observations, but with no individual length-observations, can account for missing or regenerating appendages, as long as these were noted. Even in cases where mass balances are available, individual weighing may be problematic owing to crab which may be damaged during sampling (e.g. trawling) and whose weight no longer reflects its natural state.

The focus of this paper is to develop a length-weight regression model in which missing or regenerating pereopods, along with sexual maturity, can be used as predictors. It is also of interest from a biological perspective to characterize the body weight distribution of snow crab and see how it varies by size, sex and sexual maturity.

* An annual snow crab trawl survey is performed in the southern Gulf of Saint Lawrence (Figure X).
* Crab catches are sorted and measured for various characteristics, including size, carapace condition, shell hardness, presence of missing or regenerated legs, but not weighted.
* Estimates of trawlable biomass are derived using a size-weight relationship which yields a predicted weight, based on the weight of an intact crab, i.e. a crab with no missing legs.

Owing to the relatively small size of the body, loss of even a single leg has a significant effect on the overall weight of a snow crab.

Of course this varies by species.

The weight contributions of each of the crab legs will be estimated not by direct measurement (i.e. amputation) but by their relative effects with respect to the mean weight for intact crab.

While a direct experimental approach would also have been possible, it was not done on eithical grounds, and also the detachment of legs is usually accompanied by loss of hemolymph which creates a bias in the measurement.

* Having a function which can convert between a measured size of crab and the weight are important within stock assessments to link length-frequency sampling to biomass equivalents.
* Past studies have yielded such equations for specific groups of snow crab.
* These are generally allometric relations which link weight to some measure of size of the form:
* Unidimensional length or width measures of size are related to the weight, which is largely a measure of volume imply that the coefficient should be about three while the coefficient is a function of the shape and composition of the organism as well as the units being used.
* A linearized statistical version of the allometric model can be obtained by taking the logarithm on each side and including an error term :
* where the error term is usually assumed to be homogeneous .
* This model may be extended by introducing indicator variables which encode the missing or regenerating leg pattern.
* Let be a variable denoting the number of missing legs at position where indicates a missing cheliped and indicate the number missing walking legs on either side, from front to back or anterior to posterior.
* Let the number of regenerating legs be defined similarly as above.
* Including these terms the linearized allometric model becomes:
* This model is a linear model so its coefficients can be easily determined via linear regression of observed weights over the predictors, the crab size and missing leg pattern.
* Missing and regenerating legs can be seen as modifying the intercept of the linear term on the log-scale. This is equivalent to saying that missing or regenerating legs *scale* the allometric relation between weight and size, and they do so independently.
* That is they represent a constant proportion of the weight over snow crab size and these proportions of the overall weight is given by:
* The predicted median weight for this relation is given by:
* The predicted mean weight of this relation is obtained by including the contribution to the mean of the variance term:

**Fitting statistical distributions to mixed precision data.**

* **For use when instruments of varying precisions are used or**
* **Data are rounded to varying degrees.**

**The weight data were recorded using different balances which had different precisions. When the observed precision is known, it may be handled by integration of the error distribution over the range of values implies by the precision. This allows for proper weighing of the data observations. This poses a problem for smaller weight values, since the allometric relation is linearized, which homogenizes the variance on the log scale, but the relative observation error increases for small crab. For instance a 20g crab measured at a 1g precision has a 5% range of probable values associated with it whereas a 100g crab would have 1%. Unfortunately, this precision was not recorded during measuring.**

* **The precision of the weight observations plays an important role in the relative weighting that each observation should have in the regression. Thus proper treatment of the precision has an effect on the inferred parameters.**
* **For example, a 20g and 100g crab measured using a 1g precision balance will have a 5% versus 1% relative error associated with it. Smaller crab will thus have an inordinate effect on the regression as their errors increase with decreasing size. Ignoring these differences in relative error is not an option, as the statistical leverage for small observations, i.e. those far from the covariate mean, means that will have all the more impact on the estimated allometric coefficients.**
* **We can address this issue by discretization of the error: calculating the probability mass associated with weight values by integrating the error distribution over the range of values implied by the precision of the instrument. For instance, a 20g weight observation made on a 1g precision balance may be interpreted as meaning that it lies in the range of 19.5g to 20.5g. An one would integrate the error disitrution over these values to obtain the probability mass associated with the observation.**

**where is typically some zero-mean error distribution parameterized by its variance.**

**However, in this case, the measurements for 2012 and 2013 were made using two different scales, one with 1g precision and the other with 5g precision, but the balance used was not noted.   
This means that the observation error distributions are mixed. The proportion of observations made with each balance varied with crab size.**

**So what solution could be used for this?**

# Expression of the missing, regenerated and maturity terms as part of the typical log-linear model approach leads to a practical problem. The allometric model is able to make weight predictions for each of the component parts of the crab: the inner disc, each lag and associated maturity changes. However, the sum of the resulting weights is not equal to the prediction for the whole crab:

* + The problem scaling contributions corresponding to each of the log-linear terms in the model corresponds to multiplicative effects when converted to the regular scale.
  + This implies, among other things, that the reduction in weight associated with a single leg loss does not exactly double when two legs are lost, as the multiplicative coefficients are squared rather than doubled.
  + To remedy this, the model was expressed as a series of additive allometric models, each corresponding to one part of the crab. This model formulation has the additive weight property which was desired.
  + For any specific, the model can be reduced to a simple allometric model.

# The alpha terms need to be modified on account of the crab not being equal to the sum of its parts.

# The new parameterization has the property that the individual contributions of each crab part, i.e. the disc, the legs and maturity parts, can be individually accounted for.

# Data

* Two separate observational data sets, from two different surveys covering the same area were analyzed.
* The first stems from the September multi-species survey for 2012 to 2017, which targets mainly groundfish using a Western IIA bottom trawl, but which also catches snow crab as an incidental catch. Using a length-stratified sampling scheme, the carapace width, missing leg, sex and maturity of each snow crab were noted.
* The first is from the snow crab annual survey data which measured XXX crab in XXXX.
* Male (and females?) were measured for all sizes(?)
* Shell conditions, missing leg and regenerating leg patterns for each organism.
* Also included were data from the September Multi-species survey which also recorded the missing leg (regenerating?) pattern and the weight of a number of snow crab over a large size range of sizes and for both sexes. However, the carapace condition was not noted. This may pose problems as epibiontic coverage, notably for females, may sometimes be sever leading to an inflation of the weights of affected samples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Survey** | **Year** | **Sex** | **Immature** | **Mature** |
| September groundfish | 2012-2017 | Male |  |  |
| Female |  |  |
| Snow crab survey |  | Male |  |  |
| Female |  |  |

* Sexual maturity was determined via morphometric observation of measurement in the case of males.
* In 2012 and 2013, crab were weighed using different scales, one with a 1 gram precision and the other with a 5 gram precision. In 2012 and 2013, the 1 g scale was used for crab weighing less than 80g. In 2012 the 5 g was used in larger proportions of observations with increasing crab size. The pattern was similar for 2013, but to a lesser degree (50 % case for larger sizes).
* Comapre the CH versus CW pattern from the two survey SVP
* Review of literature which state the length-weight relations of snow crab (Alaska and ENS and Newfoundland as well)
* Crab which were otherwise damaged: cracked shells, missing carapaces, missing abdomens, or otherwise suspected in having water loss were removed from the analysis. There were 592 such cases.
* Crab with undetermined maturity stage were assigned a maturity stage via a snow crab data binomial GAM regression on carapace width by year and sex. Probabilities lower than 0.2 were assigned as immatures and greater than 0.8 as matures. Mid probs were removed from the analysis.
* Males smaller than 40mm were assumed to be immatures.
* Females smaller than 40mm were assumed to be immatures.
* There

# Data issues and Modelling:

### Rounding:

* Length measurements were rounded to the nearest millimeter. This was not considered in the following analyses.
* It is possible to treat the error-in-variables problem using a Bayesian analysis.
* The weight measurements were rounded to the nearest gram.
* For smaller sized crab, the relative measurement error becomes important.
* To account for this, the underlying continuous error distribution of the weight observations was discretized to integer values by integrating the probability density of the underlying error distribution over a 1 g interval, half a gram from either size of the observed weight, yielding a probability mass, which was used to evaluate the likelihood function.
* Observed weights of 1g or less were treated as censored values lying below the value w = 1.5g.
* This discretization allows for proper weighing of the observations, with smaller crab weight observations having more uncertainty associated with them than larger values.

### Measurement errors:

* Given the large number of observations, coupled with the difficulty of measuring at sea, there were a small but significant proportion of observations which could be considered as outliers.
* Given the large number of factors which can influence crab weight, identifying outliers by some ad hoc threshold values is problematic, as there are in fact a large number of them, one for each specific case.
* Thus a robust regression was opted, one which used a mixture distribution or two Gaussians, each centered on the regression line.
* An extra Gaussian component, one with a larger variance term is included to model outliers about the regressed mean. The variance terms of each component as well as the relative proportions of each were estimated along with the regression coefficients.
* Figure X shows an example of the continuous and discretized versions of this error model.

## RV survey data and processing

* The southern Gulf multispecies survey, using a trawl designed designed to catch mainly groundfish, also catches snow crab.
* Though catchability is fairly low, the area swept by the trawl is much greater and so sizeable catches of snow occur.
* Since 2012, both length and weight measurements were taken , stratified by millimeter carapace width. Chela heights were also measured with the aim of determining male morphometric maturity.
* Female maturity was determined by direct observation of the abdominal segments.
* The missing leg pattern was also observed but no distinction was between naturally occurring missing legs and those induced by human activity .
* Leg loss during sampling (i.e. trawling, extraction from the trawl or manipulation during sampling)

|  |
| --- |
| Table X : Summary of weight-length regression coefficients obtained from Hébert et al. 2002. Units are weights in grams and carapace widths in millimeters. |
|  |
| |  |  |  |  | | --- | --- | --- | --- | | **Carapace category** | **Maturity** |  |  | | New | Adolescent |  | 3.453 | | Intermediate | Adolescent |  | 2.899 | | Old | Adolescent |  | 3.065 | | New | Adult |  | 3.315 | | Intermediate | Adult |  | 3.039 | | Old | Adult |  | 3.098 | |

* Currently, commercial biomass estimates, from which quota levels are set, are estimated from length-frequency data assuming that commercial-classed crab are intact and thus have no regenerating or missing legs or chelipeds.
* This study analyzes a set of observational data to determine the length-weight relationship for crab which accounts for any specific pattern of missing or regenerating legs.
* We then use this function to account determine the difference in biomass estimates which account for these factors.
* We must test for possible interactions between the missing or regenerating covariates and the carapace width.
* There is the median estimate and the mean (bias corrected) predictions.

# Results:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Body part** | **% by weight** | **95 mm** | **100 mm** | **110 mm** | **120 mm** | **130 mm** |
| Discus |  |  |  |  |  |  |
| Cheliped |  |  |  |  |  |  |
| Walking leg I |  |  |  |  |  |  |
| Walking leg II |  |  |  |  |  |  |
| Walking leg III |  |  |  |  |  |  |
| Walking leg IV |  |  |  |  |  |  |
| Total | 100 % | X g | X g | X g | X g | X g |

* Discuss cases where epibiontic ocverage biases the weight measuremnts

## Commercial Biomass Example:

# References

Hébert, M., Benhalima, K., Miron, G., Moriyasu, M. 2002. Moulting and Growth of Male Snow Crab, Chionoecetes opilio (O. Fabricius, 1788) (Decapoda, Majidae), in the Southern Gulf of St. Lawrence. Crustaceana 75(5): 671-702.



**Data**

|  |
| --- |
| **Things to check:** |
| * RV samples include legs missing by manipulation as well as those occurring naturally. Insure yourself that this was done in a consistent manner. |
| * Hemolymph loss may add variability to the weight values. |

**SCS Protocol Description:**

**A snow crab survey is performed annually in the southern gulf of Saint Lawrence from July 10th to the end of September.**

**Snow crab are separated from other species in the catch and the following data are recorded: the carapace width and chela height using a digital Vernier caliper to the nearest 0.1 mm; the carapace condition which ranks (from 1 to 5) the relative age since the last moult, with 1&2 representing crab which moulted in the survey year; the sex of the crab. Sexual maturity of males was assessed morphometrically via the relative size of the chela with respect to the carapace width, whereas the width of the abdomen was used in the female case, determined visually. Any missing or regenerating pereopods were noted. Pereopods which were lost due to trawling or handling were identified separately.**

**Each crab was weighed to a precision of 0.1 g. Any damage, including cracked carapace or pereopods or parts thereof, along with missing abdomens, were noted. Marel Marine balances, suitable for use at sea, were used.**

**Crab were not dried off prior to measuring so adhered water on the carapace may be an issue, though the time between catch sorting would allow some drying to have taken place.**

**Snow crab smaller than 40 mm were automatically assumed to be immature.**

**Model:**

Crab size and weight is assumed to follow an allometric relationship:

or its linearized form , where is measured in grams and is the carapace width in millimeters.

* Given that is a volumetric measure and is a linear size measure the parameter, corresponding to the dimensional ratio between the two measures, will generally be close to 3. is a scale parameter.
* We may extend this model to include predictor variables in a number of ways.
* We restrict ourselves to versions of the model which are additive rather than multiplicative as we want the component body parts to sum to the whole.
* We will add covariates to the scale coefficient, though the coefficient can be similarly generalized if we way to consider interaction between predictors with size .
* We consider to represent the scaling parameter for an intact immature crab, i.e. no missing pereopods.
* We decompose this term into the different contributions from each body part:

where refers to the contribution from the crab body only, is the scale parameter for observation , is the number of missing pereopods (0, 1 or 2) at position on either side of the crab, is the scaling contribution for a missing pereopod position , is the number of regenerating pereopods (0, 1 or 2) at position , is the scaling contribution for a regenerating pereopod position , is the maturity of the crab and is the scaling contribution for a mature crab.

**Error distribution:**

* The Gaussian error distribution used in the length-weight regression was extended in response to two major issues.
* The first was the presence of a relatively small but significant proportion of outliers in the data, which make the tails of the error around the mean heavier (increased kurtosis) than expected for a Gaussian distribution.
* The second issue was the precision of the weight observations which was sufficiently low in many cases (generally for smaller sized crab) that the increase in observation error was too large to ignore.
* In addition, three difference weight balances were used with three different precision levels.
* Different combinations of these were used for different years, and lower precision balances were usually used for larger sized crab.
* Thus a generalized Gaussian distribution was used which allowed for extra kurtosis was used.
* This continuous distribution was then discretized to account for the rounded nature of the weight observations, which depended on the balances being used.
* Outliers in the data are due on the one hand to observation errors and on the other to natural factors which can affect the weight of a given crab.
* Among the known factors are the growth of epibiontic organisms, such as barnacles or bryzoans which, in addition to the weight they contribute themselves, also increase the surface area of the crab, which increases the quantity of adsorbed water content.
* Also, the condition of the carapace and its moult stage can also affect the degree of carapace calcification can also affect the weight.
* The degree of muscle development, especially after moulting.
* Unnoted carapace damage can lead to underestimation of weight, as there may be loss of hemolymph.
* There may also be technical issues such as untared balances or uncalibrated balances. However, for the data presented here the balances were supposed to be calibrated daily.
* Rough sea conditions can also lead to fluctuations in weight readings, requiring the experimenter to mentally average out the values.
* Marrel scales, which compensate for movement at sea, were used for all measurements.
* Those used in the RV survey had precisions of 5g, 1g and 0.1g, under regular conditions.
* In the model, we interpret the weight balance precisions in the following formal manner:
* Let the weight observation be made with a balance with precision .
* Then we assume that the true weight lies within the interval .
* Is it possible to obtain a posterior update given a prior uniform distribution over this interval?
* The final consideration is that samples from a given year were made with multiple balances with different precisions, but it is generally not known which balance was used for some of the observations.
* More precise balances were used for smaller crab, but not consistently so, with the changeover from more to less precise balances occurring over a range of increasing crab sizes.
* For the RV survey, the 2012 and 2013 data were measured with a mixture of 1 g and 5 g precision balances, while 2014 and 2015 with a 1 g precision balance only and 2016 and 2017 with a mixture of 0.1 g and 1 g precision balances.
* The relative proportion of each balance can be estimated as a function of crab size by considering the proportions of weight observations which are a multiple of the balances’ precisions.
* While the data could have been rounded down to lower precision, it was felt that valuable information would have been lost about the length-weight relation for smaller crab.
* We opted rather to include these sources of uncertainty within the model structure.

**Formal model description:**

* Let us now formally develop the error structure of the model.
* Let represent the mean of the allometric model for a given crab size .
* Let be a continuous valued random variable which is modelled as a mixture of two zero-mean Gaussian distributions and with and relative proportion .
* The probability density function for is given by:
* where is the probability density function for a standard Gaussian distribution.
* We refer to as a kurtotic Gaussian random variable, as it has heavier tails than a Gaussian variable for .
* Note that this distribution collapses to the Gaussian case for or , with one or the other variance parameters becoming redundant.
* It also collapses when and then becomes redundant.
* The final step to define the likelihood is to discretize is to discretize the continuous distribution above to account for the precision of the weight balances that were used.
* The probability mass function for a discrete weight observation is given by:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

**Determining the balance precisions:**

The question is now how to determine the precision of a balance given a

* Given a set of weight observations , measured at two precisions and where .
* We also assume that is some multiple of .
* We know the precisions of those observations which are not a multiple of , as those values could only have been generated by a balance of precision .
* Formally, .
* The remaining observations of the form , for some positive integer need some attention.
* One option is to simply assume that they are all of precision, which leads to some loss of information.
* Another option, which we adopt here, is to model the proportion of precisions as a function of crab size, since there proportion of ….
* The log-likelihood for this model is given by:

where is the complete parameter vector for the model, which includes the allometric coeffcients, the kurtotic Gaussian error model as well as the precision proportions model parameters.

|  |
| --- |
|  |
| **Figure X**: Example of a kurtotic Gaussian probability distribution (solid black line) and its discretization for precision . Dashed black lines show the two components making up the kurtotic mixture with an 85% signal (sharper central component) and 15% noise (diffuse low-level component) composition. |