# FPInnovations – SFU - UBC

**FOREST PRODUCTS STOCHASTIC MODELING GROUP: Advanced** **Manufacturing and Product Development**

**The Way Forward for 2018/19 by Project**

**Revison 2: Prepared by Jim Zidek, .**

**Financial:** the budget has eight one term Full Time RA’s in it for each year of the grant. The allocations below are provisional based on comments at the December group meeting. Four of those eight (the ones without question marks) were committed in the original plan although not all of these may be needed.

**Administration: Jim**, the PI on the grant oversees the grant in consultation with **Conroy** (as the FPInnovations contact person) and **Will** Welch, these three serving de facto as the Advisory Committee**.**

**Carolyn** is formally appointed as both the Research Coordinator (RC) and Scientific Research Associate (SRA) in accordance with the terms of the CRD Grant, although it is the Department of Statistics, UBC-V that covers her RC salary as an in-kind contribution to the research project. Her role as the RC includes organizing research group meetings and maintaining the Project’s web page <https://forestat.stat.ubc.ca/tiki-index.php>. The latter means keeping it up to date by adding links to publications, installing reports and presentation slides, adding datasets along with computer code. But the page is a wiki site, so any of us with passwords can in principle do these things ourselves. **Carolyn** is the go-to it person should you need to know ~~to know~~ about datasets and software or if you need to get access to the dataset. She will also be coordinating an experiment at FPInnovations (subject to feasibility) in the summer of 2018.

But **Carolyn** is also a very experienced data scientist who is actually employed by ASDa (the Applied Statistics and Data Science Group). In her role as the SRA, **Carolyn** is the technology transfer agent for the project. Working under a nondisclosure agreement with FPInnovations she tries out on propietary data methods and code developed under our Collaborative Research. She will be the one to check and try out your R code before it or a link to it is posted on the public part of the home page.

**Faculty, Postdocs and ASDa Staff involved in Summer of 2018:**

**Carolyn Taylor**

**Jiguo Cao**

**Jiahua Chen**

**Harry Joe**

**Jim Zidek**

**Samuel Wong**

**Seong-Hwan Jun**

**Alex Bouchard-Côté**

**UGRAs for Summer of 2018:**

**Kunjia Shi** (3rd year stats undergrad paid out) 

**GRAs for Summer of 2018:**

**Fatema Jhohura** (Stats MSc on leave to May 1)

**Xinyao Fan** (Stats MSc)

**Yanchao Luo** (Stats MSc)   
**Henry Chen** (Stats MSc @ Google returning from a year's leave)

**Chun-Hao** (Stats PhD, U of Florida)

**Projects: numbered as in the original grant proposal:**

**#1. Long term monitoring**

* Faculty: **Jiahua & Jim**
* ASDa: **Carolyn** (technology transfer)
* Full GRA Two Terms: **Boyi Hu &Yanchao Luo**

#2. **Species grouping**

* Faculty: Jim
* GRA: **Future GRA**

#3. **Property relationships**

* Faculty: **Jim & Harry Joe**
* **Carolyn** (technology transfer)
* GRA & UGRA: **Fatema Jhohura & Kunjia Shi**
* GRA**: Howard Chen**

#4. **Proof–loading rates & load levels for end–joined lumber**

* Faculty: **Will & Jim**

#5. **Duration of load**

* Faculty: **Samuel & Jim**
* UGRA: **Fatema Jhohura & Kunjia Shi**
* **UBC’s Wood Science group**

#6. **Design values**

* Faculty: **Jim & Jiahua**
* GRA**: Xinyao Fan**

#9. **Engineered wood products**

* Faculty: **Jiguo**

#11. **Enhanced grading systems**

* Faculty: **Alex, Seong & Samuel**
* GRA & UGRA: **Fatema & Kunjia**

**Technical details.**

#1. **Long term monitoring** [Conroy Lum, Jiahua Chen Pengfei Li, Yukun Liu, Boyi ~~Zhu,~~ Guangyu, Carolyn and Jim Zidek]

NOTES: Canada’s long term monitoring program was set up to provide annual assessments of the elasticity (MOE) of representative samples of lumber. The primary, secondary and tertiary units are respectively sawmills, bundles within mills and (five) pieces of lumber within bundles. Both mills and bundles were found to have a significant random efffects within the linear model for MOE, which has an approximately Gaussian distribution. This led to a rotating panel design with primary units being in the panel for an optimal period of five years each. That program is now in operation. Moreover since Gaussian theory applies, it was not too hard to develop a generalization of the two-sample test to eliminate the mill effect in testing the null hypothesis of no change.

However after it was put into operation, interest turned to the bending strength of lumber (MOR), which can have in some cases a decidedly non-Gaussian distribution, depending on the species from which the lumber is obtained. The twoparameter Weibull provides a reasonable approximation. But little seems to known about assessing trend in this case. **Jim** has developed under contract and under pressure of time an exact nonparametric test of the null hypothesis. But it will not be very powerful and better alternatives are needed. As well, **Carolyn** developed again under contract, an ad hoc way of bootstrapping rotating samples for MOR trend analysis.

The following issues need to be addressed

* Trend assessments for MOR rotating panel designs (longitudinal clusters)
  + How to bootstrap longitudinal (rotating panel) MOR data
  + How to assess trends in MOR quantiles
  + How to deal with the matroids (arrays) – one core cell vs

in rotation, one alternative cell in each successive years

to minimize resource costs, cells representing different

sizes of lumber

* How to construct a generalization of the two sample t test for the two parameter Weibull-that is one where the big primary unit effect is eliminated.
* Development of a multivariate control charting method (UTS, MOR etc).

After the Canadian LTM program was set up, the US decided they needed one, but with the difference that just one cross sectional sample is be drawn independently every fifth year. **Jiahua & co-investigators** have developed an alternative to the Kruskal-Wallis test for comparing samples. However this technology needs to be transferred through **Carolyn** with her testing on propietary data under her nondisclosure agreement, the DRM method. Specifically:

* implementation of the DRM method for clustered data and independent samples (needed as alternative to Kruskal Wallis that has been proposed).
  + Find code
  + Apply it to successive LTM data samples to compare the DRM approach to Kruskal-Wallis
  + Write technical note (TN) for **FPInnovations**

#2. **Species grouping** [Yumi Kondo, Carolyn Taylor, Constance van Eeden, Davor Cubranic & Jim Zidek]

Lumber is sometimes marketed in clusters called species groups, each group being represented by a sample of strength-tested lumber but together, represented by a single set of design values DVs. How to assign these values has become controversial. In particular, the published method for finding the subset of the species called the “controlling subgroup” on which the calculation of the DVs is based has some unanticipated and undesirable properties.

Two methods have been developed and published by the Forest Products Stochastic Modelling Group (FPSMG), one taking a classical nonparametric approach and the other Bayesian in nature has two different versions. The first has been published, the second submitted in a paper currently under review. Software has also been developed and is available for applying the method.

The recent reemergence of the issue has sparked interest in the theory. So there are two projects of emergence.

* **Carolyn** needs to be involved in more actively promoting the new methods as part of the technology transfer
* The Bayesian methods makes an independence assumption among model parameters that could well be relaxed at considerable computational cost, but seems worth exploring since it would enable strength to be borrowed (**Future GRA**).

#3. **Property relationships** [Davor Cubranik, Henry Chen, Lang Wu, Tara Cai, Harry Joe, Shenyi Pan and Jim Zidek]

Quite a number of projects have come under this heading and it overlaps with the duration of effects as well as proof loading projects. The problem was explored by **Tara Cai** in her UBC PhD thesis where she built on an the approach called “breaking the same board twice”. The idea was to proof load lumber species on one strength property e.g. ulimate tensile strength (UTS) and break the survivors on another e.g. modulus of rupture (MOR). Under multivariate normal theory it is possible to tease an estimate of the correlation between UTS and MOR, that in turn makes it possible to predict one from knowledge of the other. This can cut down the cost of monitoring for it means testing for only one of the strengths. Her thesis reports a method for estimating and adjusting for the degree of damage to MOR due to the proof loading on UTS.

For simplicity Tara’s approach ignores covariates for that specimen notably stiffness (MOE) a nondestructively measured covariate that is a strongly related to both UTS. The modification was left for future work, which Tara was unable to do.

That led to the next project:

* Develop within a multivariate normal theory framework, a generalization of Tara’s work and in particular, an extension within that context of a model for estimating the effects on quantiles of the damage caused by proof loading (**Henry**). A draft paper has been written but before it is submitted needs to incorporate **Henry**’s work that is to be completed in the summer of 2018.

**Tara’**s thesis also foresees application of her approach to that of finding the relationship between the strengths of different sizes of lumber. For example, lumber specimens 12’ in length can be proof loaded on MOR. A random length of 8’ cut from the survivors could then be tested on UTS. ``Damage’’ could now formally reflect the fact that the 12’ lengths are cut from different types of logs than 8’ lengths in industrial production and this effect could be estimated using **Tara**’s quantile function method. However it would include actual damage if the proof loading level is set too high, so the experiment might involve both low and high proof loading levels to separate the two types of damage.

Subsequent to the completion of Tara’s thesis **Harry Joe & Shenyi Pan** developed an alternative approach using coppula’s. This approach has the advantage that is can handle non-Gaussian marginal distribution and the authors in their draft paper show how, with one margin being a Weibull distribution, the other a normal. This leads to a new project.

* Implement the **Harry-Joe** approach on propietary data (**Carolyn**)

At the same time **Harry Joe & Shenyi Pan** revisit the problem of estimating the damage model and found what appears to be an anomaly in the much used pplot() diagnostic function at least with respect to how it compute points in the all-important tails regions for the two distributions. This will lead to adjustments in the estimated damage function found by Tara’s method.

* The effect on the damage estimates seen by **Tara** by replacing qqplot() with quantile functions needs to be investigated. In particular a new parametric model for damage needs to be found (**Fatema &** **Kinjia**)

The important dataset developed by **Dave Barrett** and **Ricardo Foschi** was provided to the Group by **Erol Karacabeyli.**  A different method for calculating damage was applied to the **Barrett-Foschi** dataset according to **Erol.** There a constant load was applied to lumber specimens over an extended period. Those that did not fail were then broken in a ramp load test. Their strengths were then compared with controls that were matched on their elasticity (MOE) and the difference was ascribed to damage. This leads to a new project.

* Compare matched sample approach (**Barrett-Foschi** data) with those deriving from those of **Tara**, **Henry** & **Harry** for damage model estimation. That project will involve the background reading as well as leading the codes, finding and exploring the relevant data (**Kinjia**).

#4. **Proof–loading rates & load levels for end–joined lumber** [Ciprian Pirvu, Tianji Shi, Will Welch and Jim Zidek]

Need to revise papers developed under Ciprian’s watch (**Ciprian – Conroy to contact**)

#5. **Duration of load** [Erol Karacabeyli, Conroy Lum, Ciprian Pirvu, Henry Chen, Tara Cai, Nancy Heckman, Samuel Wong, Chun-Hao Yang and Jim Zidek]

* Identify specific ways in which work to date can inform policy making
  + Need to get data on new forest products to apply work already done.

(**Jim to approach UBC Wood Science Folks for data**)

* Write a review paper for the users. (**Samuel and Jim**)
* The test dataset provided by Erol also contains a problem that remains to be tackled. See #3 for details.

#6. **Design values** [Seagle (Yang) Liu, Matias, Ruben Zamar and Natalia Nolde]

A paper has just appeared that incorporates some fo the work Seagle did in his MSc thesis, which in particular looks at fitting the two parameter Weibull distribtution by artificially right censoring about 90% of the data before using the likelihood method. As well Seagle explored use of the bootstrap and he saw that the left tail of strength distributions are mixtures of two or more Weibulls corresponding presumably to different failure modes. Finally he published code for doing some of this work.

A number of projects come out of this work (**Xinyao**!).

* Review ASTM 5457 (**Jiahua & Jim**)
* Read Seagle’s MSc thesis and see what he does. Learn what is in his CRAN package and document it sufficiently well that **Carolyn** can assess it on proprietary data and transfer it to the community of users, perhaps with a Shiny interface. Finally see if he includes a way of fitting a mixture of Weibulls to the left hand tail again with a view to transferring the technology.
* Visit the code Kumi published for the species grouping problem and read the paper she first--authored, all with a view to using her non-parametric strength distribution estimation method not for grouping species but for modelling the lower tail of the distribution.
* Finally develop the extreme value approach for modelling the peak-under-threshold after studying what Seagle did in his theory and apply it to the various datasets we have.
* The extreme value approach can be enhanced by adding a Bayesian layer to the model. This will involve reviewing the literature on Bayesian versions of the peak-under-threshold approach.

#7. **Data fusion** [Zhiyong Chen, Nelson (Hao) Chen and Will Welch]

* Apply theory to more sophisticated computer models (**Zhiyong & Will)**

#9. **Engineered wood products** [Kevin Groves, Martin Feng, Tianyu Guan, Peijun (Perry) Sang, Jiguo Cao]

* Finish two manuscripts on identifying the wood type (sound, rot and bark) and measuring their content by analyzing Vix-NIR spectroscopy data (**Jiguo and Tianyu**).
* Need to develop effective method where NIR spectroscopy is used for the determination of phenolic glue curing of wood panels. **Jiguo** needs to meet with **Martin and Kevin**.
* Document computer code (**Jiguo and Tianyu**).

#11. **Enhanced grading systems** [Zarin Piroux, Bruce Lehman, Seong Jun, Alex Bouchard-Côté, Samuel Wong, and Jim Zidek]

Modern high tech mills scan lumber as it is produced and apply the grading rules developed for human graders to separate the lumber into grades, each of which has its own published design values, i.e. stipulated strengths for engineering applications. Work to date has been the principle strength reducing characteristics, the knots in lumber left where branches used to grow. The scanners in mills can see the knots on all four edges but matching them is tricky as the branches can grow through the tree in weird and crooked ways.

**Seong Jun**’s recent PhD thesis written presents a computer modelling approach that does the matching quite accurately. But the dataset Seong gathered for his analysis is quite small and needs to be expanded. That leads to the two projects:

* Carry out an experiment at the FPInnovations lab in the summer of 2018 under the supervision of **Conroy, Zarin & Bruce**. **Jim** to assemble a team of those employed in the summer of 2018 with salary supplements.
* Analysis of the enlarged database. (**Seong** will train **Fatema &** **Kinjia on use of his code.)** The relevant parts of **Seong’s** PhD thesis will need to read. (**Fatema &** **Kinjia).**

The next stage will involve the transfer of technology including a code vignette.

* Develop professional quality image processing system for knot detection (**Seong & Carolyn**).

The published strength prediction model published by Samuel & Jim involves a Poisson distributed random number of knots of varying sizes. The model runs on a multiprocessor cluster and is computer intensive. The next project will see the latent knots in the model replaced by Seong-knots.

* Develop a strength prediction model under the supervision **of Samuel, Seong Alex & Jim** (**Chun-Hao & Future GRA**) based on the **Wong-Zidek** model.

A major project that has not yet being tackled has to do with the configuration patterns in the knots on a piece of knots. It may well be that a bunch of small knots in a tight cluster could have the same impact on strength as a single large knot. That leads to the next project

* Develop a spatial model to reflect interknot spatial correlations depending on the knots features, e.g. proximity to other knots, location on the board and so on. This will be a difficult problem since it will have to take account of the angle of the fibres around the knots, and the interactions between the knots. This may need to rely on a suitably modified of model developed by **Bruce Lehman**. It is expected that **Seong, Samuel, Alex & Jim** will be involved (**Future** **GRA).**

#12. **Duration of load effects in bio-based products** [Erol Karacabeyli, Conroy Lum, Ciprian Pirvu, Vincent Zhai, Tara Cai, Nancy Heckman, Samuel Wong, and Jim Zidek]

* Build on work done in #5 but look at a new range of forest products.

#13. **Non-destructive testing** [Conroy Lum, Dustin Johnson, Will Welch and Nelson Chen]

* Extend current theory to allow for replicate measurements on each method
* Develop new ASTM standard.