***Introduction***

The ability to accurately estimate tree biomass is essential for scientific, economic and policy purposes. Concerns about global climate change have led policy-makers to focus efforts on maximizing the carbon stocks and sequestration abilities of forests (Domke et al. 2012; Eggleston et al. 2006; Van Breugel et al. 2011). Thereby, effective policy design requires that forest carbon stocks can be accurately and cheaply estimated. Furthermore, forest carbon-offset trading on emerging carbon markets ties estimates of carbon stocks to monetary transactions, amplifying the implications of misestimation (Kerchner et al. 2015; Newell et al. 2013). Since forest carbon cannot be directly measured, carbon stocks are most often estimated using allometric models (Jenkins et al. 2003; Van Breugel et al. 2011). Allometric models, first described by Huxley and Tessier in 1936, relate low-dimensional, easily-obtained measurements such as height and diameter, to difficult and destructive measurements such as individual tree biomass (Huxley and Tessier 1936). Difficulty in obtaining accurate data, and a lack of consensus on methodology lead to significant uncertainty and inaccuracy in both biomass and carbon estimates, necessitating the exploration of this problem using novel statistical techniques (Sileshi 2014; Jenkins et al. 2003; Van Breugel et al. 2011).

One major source of uncertainty in allometric models lies in the assumptions made about basic model structure (Sileshi 2014; Van Breugel et al. 2011). Most commonly, allometric scaling relationships adopt the form of a power-function, owing to the apparent linearity of allometric relationships plotted on a log-log scale (Picard et al. 2014; Sileshi 2014). These model forms are referred to as “simple” allometric models (Huxley and Tessier 1936). “Complex” allometric models, on the other hand, specify relationships that are non-linear on a double logarithmic scale (Huxley and Tessier 1936; Picard 2015). A third method of allometric modelling invokes a priori assumptions about the physiological relationships between tree components (West et al. 1999; MacFarlane 2015). This paper focuses on the first two types of models, whose formulation are driven by statistical analysis rather than theoretical deduction. The volume of literature touting each of these methods has created an overwhelming predicament for managers, economists and policy-makers attempting to choose between and apply these models.

The use of a Generalized Additive Model (GAM) allows us to relax the assumption of linearity adopted by “simple” allometric models, without specifying an alternative, non-linear model form as is done with “complex” allometric models. This is made possible by the non-parametric nature of GAMs, which relate the response variable (tree biomass) to the predictor variables (diameter and height) via unknown smoothing functions (Hastie and Tibshirani 1987). This flexibility allows the relationship between the predictor and response variables to break the assumptions that accompany “simple” and “complex” model specifications and potentially lead to a better fit to the data (Hastie and Tibshirani 1987). Given the economic and social importance of properly estimating carbon stocks, such improvements to the accuracy of allometric models is critical.

Most models incorporate heterogeneity in allometric relationships among species or groups of similar species (Jenkins et al 2013; Picard et al 2015). However, few account for variation in allometric scaling relationships among variables other than species and diameter. Evidence suggests that incorporating additional sources of heterogeneity into allometric models could improve their accuracy (Weiskittel et al. 2015; Fatemi et al. 2011). For example, studies detailing differential growth patterns among trees of different ages suggest that allometric relationships may vary according to tree age (Bond 2000; Fatemi et al. 2011).

Another significant barrier to the development of accurate allometric models is the difficulty and expense involved in directly measuring tree biomass (Van Breugel et al. 2011; Sileshi 2014; Jenkins et al. 2003). In order to measure the biomass of a single tree, it must be cut-down, dissected, transported to a lab, dried and then weighed (Whittaker and Woodwell 1968; Fatemi et al. 2011). The destructivity and inconvenience of this process prohibits the collection of large datasets, and this implicit restriction on sample-size limits the degree of certainty in models derived from single-study datasets. This study addresses this difficulty in data collection by synthesizing multiple, comparable datasets collected in the White Mountains of New Hampshire over the course of several decades.

Through the formulation of a Generalized Additive Model for allometric relationships in the northern hardwood forests of New Hampshire, this study seeks to answer several questions: (1) Which factors are important for allometric relationships when assumptions of linearity on a log-log scale are relaxed? (2) Is there a difference in allometric scaling relationships between trees in young and old stands? Answering these questions will help improve the accuracy of this particular allometric model, guide the development of future models for a more diverse set of regions and species, and aid managers, economists and policymakers attempting to apply allometric scaling relationships.