



7º CONGRESO FORESTAL ESPAÑOL

**Gestión del monte: servicios
ambientales y bioeconomía**

26 - 30 junio 2017 | Plasencia
Cáceres, Extremadura

7CFE01-096

Edita: Sociedad Española de Ciencias Forestales
Plasencia. Cáceres, Extremadura. 26-30 junio 2017
ISBN 978-84-941695-2-6

© Sociedad Española de Ciencias Forestales

Calibración De Regresiones Ponderadas Geográficamente Para La Estimación De La Relación Altura-Diámetro

QUIRÓS SEGOVIA, M.¹

¹ Facultad de Forestales y Tecnología de la Madera. Universidad Mendel de Brno. República Checa.

Resumen

En este documento se realizó un estudio detallado de la calibración y aplicación de las Regresiones Ponderadas Geográficamente en la modelación de relaciones altura-diámetro. Las Regresiones Ponderadas Geográficamente son una metodología de análisis espacial capaz de analizar procesos heterogéneos, como los existentes en el medio natural. En general, se observó que una calibración mediante las distancias de Minkowsky proporciona mejores ajustes que las calibraciones básicas. Las Regresiones Ponderadas Geográficamente tienen claras ventajas sobre otros métodos tradicionales como las regresiones lineales o los modelos de efectos mixtos. Particularmente importante es la capacidad de predicción de coeficientes en áreas no muestreadas, de esta manera estas regresiones no necesitan mediciones adicionales ni calibraciones.

Palabras clave

Relación altura-diámetro, Modelación Forestal, Inventario de Recursos Forestales

1. Introducción

La altura del árbol es una variable muy importante en la gestión forestal. Por ejemplo, la altura total junto con el diámetro normal permite estimar el volumen del árbol, en masas regulares la altura dominante junto con la edad se utilizan para estimar el índice de sitio y la distribución de alturas en masas forestales es muy importante para describir la estructura vertical. Sin embargo, la medición de alturas suele ser cara, por ello se utilizan modelos altura-diámetro. La construcción de modelos altura-diámetro ha sido tradicionalmente realizada usando métodos lineales, modelos de efectos mixtos o modelos aditivos generalizados. Sin embargo, hay otras metodologías que no son tan utilizadas pero pueden ser perfectamente válidas. En este documento ha sido analizada la metodología de las regresiones ponderadas geográficamente (RPG). Las RPG son regresiones que analizan datos espacialmente distribuidos de acuerdo con una ventana móvil de diferentes anchos de banda y produce un conjunto de coeficientes para las zonas analizadas (Páez & Scott, 2004). En las RPG los parámetros son estimados en todas las zonas del área analizada dada una variable dependiente y un conjunto de una o más variables independientes de las cuales deben conocerse sus localizaciones geográficas (Charlton et al., 2009). Esta metodología se utiliza para explorar la heterogeneidad espacial (Brunsdon et al., 1996), la cual es consecuencia de una compleja relación histórico-medioambiental impuesta por la competencia y la heterogeneidad inherente al medio natural (Zhang et al., 2004). Finalmente, a diferencia de los modelos de efectos mixtos, no son necesarias mediciones adicionales para predecir el valor de los coeficientes de zonas no muestreadas. Las RPG extienden las regresiones lineales analizando la estructura espacial y estimando un modelo diferente para cada localización geográfica dentro del área de estudio (Matthews & Yang, 2012). Mediante la interpolación de los coeficientes predichos por las RPG, es posible realizar mapas de coeficientes del área estudiada. Esto es posible ya que la presencia de correlación espacial hace posible la predicción de variables de interés en localizaciones donde no hay disponible mediciones secundarias (Nanos et al., 2004). En estudios previos se observó que los modelos altura-diámetro obtenidos con las RPG y aplicados a funciones altura-diámetro generalizadas proporcionan resultados que son al menos tan buenos como los modelos lineales mixtos (Quirós-Segovia et al., 2016).

2. Objetivos

El objetivo principal de este proyecto fue analizar varias opciones de calibración de las RPG para modelar la relación altura-diámetro en grandes áreas, para optimizar la parametrización del modelo y finalmente poder mejorar los resultados previos.

3. Metodología

Para realizar este estudio se han seleccionadas masas monoespecíficas de *Pinus halepensis* Mill. en la Región de Murcia. El área estudiada se extiende por las regiones montañosas de Sierra Espuña, Cambrón, Burete, Lavia y Quipar. En esta región, el pino carrasco crece en masas con bajas densidades y relativamente bajas alturas dominantes, probablemente reflejando medias o bajas calidades de sitio (Condés & Sterba, 2008). Los datos fueron obtenidos a partir de 230 parcelas con 2582 pies pertenecientes al Tercer Inventario Forestal Nacional. Para realizar el modelo es necesario conocer las coordenadas de los pies de la masa analizada, lo que se calcula mediante coordenadas polares proporcionadas por el Inventario Forestal. Los datos estudiados fueron divididos en dos grupos. El primer grupo consistió en 152 parcelas con 1697 pies que fue utilizada para desarrollar el modelo. El segundo grupo fue utilizado para validar el modelo, y consistió en 78 parcelas con 885 pies. La tabla 1 muestra las variables de masa de ambos grupos.

Tabla 1. Análisis estadístico de los datos.

Grupo 1 - 152 parcelas / 1697 pies							
	d	h	N	G	dg	H _o	D _o
Min	7.5	2.0	14	0.6	10.2	4.0	11.1
Max	55.7	21.5	1351	27.3	38.3	16.3	39.0
Media	21.4	8.6	361	9.1	20.2	8.8	24.3
Desviación Estándar	2.6	8.3	283	5.6	6.3	2.4	5.9
Grupo 2 - 78 parcelas / 885 pies							
	d	h	N	G	dg	H _o	D _o
Min	17.8	2.4	19	1.3	9.2	3.7	12.5
Max	53.1	15.5	1489	32.4	40.4	12.7	41.0
Media	20.0	8.0	420	9.3	18.9	8.3	23.4
Desviación Estándar	7.4	2.1	334	6.2	6.0	1.9	5.4

La ecuación altura-diámetro generalizada que fue seleccionada fue Cañadas I (Cañadas et al., 1999), (Ecuación 1). Esta función fue elegida por su facilidad de aplicación junto con el hecho de que ha dado resultados positivos en su aplicación en masas de pino carrasco (Cabanillas, 2010; Quirós-Segovia et al., 2016).

$$h = 1.3 + (H_0 - 1.3) \cdot \left(\frac{d}{D_0} \right)^{b_0} \quad (1)$$

Donde h es la altura total (m), d es el diámetro normal (cm), H_0 es la altura dominante (m) obtenida según el criterio de Hart (Hart, 1928), D_0 es el diámetro dominante, calculado como la media de los 100 pies más gruesos por hectárea, y b_0 es el parámetro que será estimado.

Los modelos fueron ajustados utilizando diferentes anchos de banda y diferentes números óptimos de pies en el entorno. El ancho de banda óptimo se calcula con el uso de un núcleo fijo, más varios tipos de funciones ponderadas. Un núcleo fijo asume que el ancho de banda en cada centro es una constante a lo largo del área de estudio. El número óptimo de pies en el entorno fue calculado

basado en la optimización del número de pies, este núcleo se adapta a la densidad de los datos analizados. Tanto en el caso del ancho de banda como en el número óptimo de pies, la optimización depende del tipo de función ponderada que se utilice. Estas funciones ponderadas son gaussiana, bisquare y boxcar. Donde una función *gaussiana* es una función de distancias continua entre dos observaciones; *bisquare* es una función discontinua, que da valores nulos a observaciones mayores que una distancia especificada *b*; y *boxcar* es una función simple discontinua que excluye observaciones que son más lejanas que una distancia *b* (Gollini et al., 2015).

$$\begin{array}{ll} \text{Gaussiana:} & w_{ik} = e^{-\left(d_{ik}/b\right)^2} \\ \text{Bisquare:} & w_{ij} = \left[1 - \left(d_{ij}/b\right)^2\right]^2 \quad \text{Donde } d_{ij} \leq b \\ & w_{ij} = 0 \quad \text{Donde } d_{ij} > b \\ \text{Boxcar:} & w_{ij} = 1 \quad \text{Donde } d_{ij} \leq b \\ & w_{ij} = 0 \quad \text{Donde } d_{ij} > b \end{array}$$

Donde w_{ik} es el k -ésimo elemento de la diagonal de la matriz geográfica ponderada, y d_{ik} es la distancia entre las observaciones i y k , y b es el ancho de banda para el núcleo fijo o el número de pies del entorno para el núcleo adaptativo. La librería *GWModel* en lenguaje R (Binbin et al., 2015; R Development Core Team, 2014) ha sido utilizada para ajustar los modelos. Un nivel de $P=0.05$ fue utilizado como test de significancia de las variables de los modelos. La autocorrelación espacial de los residuos fueron analizados utilizando el coeficiente de Moran I (Moran, 1950), con la hipótesis nula de la existencia de autocorrelación espacial cero en la variable ($\alpha=0.05$). La bondad de ajuste fue utilizada para comprobar la mejora de RPG sobre un lineal (Fotheringham et al., 2002).

Las RPG han sido adaptadas para beneficiarse de un gran rango de matrices de distancia, donde se ha demostrado que una matriz de distancia correctamente seleccionada puede mejorar el modelo (Binbin et al., 2014). La distancia puede ser calculada de varias maneras, donde la distancia euclidiana (DE) es la más utilizada (Binbin et al., 2015). Sin embargo, hay otra importante familia de distancias que pueden ser utilizadas variando un parámetro p y un ángulo de rotación θ , estas son las distancias de *Minkowski* (DM). Para calcular la DM óptima se ha de seguir una serie de procedimientos descritos detalladamente por Binbin et al., 2015.

Los modelos RPG proporcionan un rango de coeficientes localizados geográficamente a lo largo del área de estudio. Por tanto, a partir de una interpolación de los coeficientes del modelo se puede obtener un mapa de coeficientes. El método elegido fue el de ponderación de distancia inversa (Watson & Philip, 1985). Este método es ampliamente conocido en las ciencias naturales (Chen et al., 2012) y es fácilmente aplicable a las RPG (Tardanico, 2006). Los modelos fueron validados utilizando el mapa de coeficientes conjuntamente con el 'grupo 2' de datos. Con el mapa de coeficientes se obtuvieron las alturas predichas de los pies que conforman el 'grupo 2', mediante la extracción de coeficientes en dichas localizaciones. Posteriormente, la validación fue realizada comparando los valores de altura observadas con los valores de altura predicha mediante el mapa de coeficientes y un análisis estadístico de residuos (Tabla 2).

Tabla 2 – Análisis de residuos. e , valor de los residuos; n , tamaño de muestra; y_i valores observados; \hat{y}_i , valores predichos; \bar{y}_i valores medios.

Error medio	$ME = \frac{\sum(y_i - \hat{y}_i)}{n}$
Error cuadrático medio	$RMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}}$
Error medio absoluto	$MAE = \frac{\sum y_i - \hat{y}_i ^2}{n}$
Coefficiente de determinación	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$

4. Resultados

En la Tabla 3 podemos ver la comparación de los 12 modelos ajustados. Para que un modelo sea considerado válido ha de cumplir los siguientes requerimientos: (1) Tener un índice de Moran I próximo a 0 y con $p > 0.05$, lo que significaría cero autocorrelación espacial en los residuos. (2) Tener un F-test significativo ($p < 0.05$), lo que significa que el modelo RPG mejora una regresión lineal. Según estas imposiciones podemos decir que se observan mejoras cuando se aplican las DM. Siendo los modelos con núcleo fijo y funciones ponderadas gaussiana y bisquare los únicos que cumplen dichos requerimientos. La función ponderada que peores ajustes proporcionó fue boxcar para ambos tipos de núcleos y matrices de distancia.

Tabla 3- Comparación de diferentes calibraciones de RPG. Donde B es el ancho de banda (núcleo fijo) y el número óptimo de pies del entorno (núcleo adaptativo), SRC suma de los residuos cuadrados; R^2 coeficiente de determinación; Moran I es el índice de Moran; AICc criterio de información de Akaike corregido; F-Test, test para comprobar la mejora de RPG sobre regresiones lineales; p , parámetro de Minkowsky y θ es el ángulo de rotación de la matriz de distancias.

		DISTANCIA EUCLIDEANA						DISTANCIA DE MINKOWSKY						
		B	SRC	R^2	Moran I	AICc	F-Test	(p, θ)	B	SRC	R^2	Moran I	AICc	F-test
NÚCLEO FIJO	Boxcar	24779	53.5	0.50	0.08 P=0.00	-1045	< 1.00	(0.25, 0.52)	21358	48.4	0.54	0.05 P=0.00	-1168	<0.01
	Bisquare	2422	45.7	0.57	0.03 P=0.00	-1190	< 0.00	(0.25, 0.52)	13991	44.4	0.58	0.01 P=0.24	-1178	<0.00
	Gauss	775	44.5	0.60	0.02 P=0.05	-1202	< 0.00	(0.25, 0.52)	5593	43.8	0.59	0.01 P=0.24	-1209	<0.00
NÚCLEO ADAPTATIVO	Boxcar	156	51.2	0.52	0.08 P=0.00	-1101	< 1.00	(0.25, 0.52)	76	49.2	0.54	0.05 P=0.00	-1146	<0.01

	Bisquare	65	45.4	0.57	0.03 P=0.01	-1198	< 0.00	(0.25, 1.05)	66	45.1	0.57	0.03 P=0.00	-1216	<0.00
	Gauss	23	47.0	0.56	0.04 P=0.00	-1180	< 0.01	(0.25, 0.52)	23	47.4	0.55	0.04 P=0.00	-1184	<0.01

En la Tabla 4 se puede ver la distribución de coeficientes de los modelos válidos. Ambos ajustes se revelaron prácticamente iguales con valores mínimos (0.12-0.13), medios (0.65-0.66), máximos (1.38-1.39), desviaciones estándar (0.18) y coeficientes de correlación (0.58-0.59) similares. La validación de estos dos modelos fue también similar, mostrando una gran capacidad predictiva, con $R^2=0.71$. El análisis de residuos también mostró alta calidad (ME, -0.1; MAE, 0.8; RMSE, 1.1).

Tabla 4 – Distribución de coeficientes para los modelos óptimos DM. Donde Min es el valor mínimo de los coeficientes; Media es la media de los coeficientes; Max es el valor máximo de los coeficientes; DS es la desviación estándar de los coeficientes; R^2 es el coeficiente de determinación.

Distancias de Minkowsky (DM)		Min	Mean	Max	SD	R^2
Núcleo fijo	Bisquare	0.12	0.65	1.39	0.18	0.58
	Gauss	0.13	0.66	1.38	0.18	0.59

La distribución de coeficientes de ambos modelos es muy parecido, por lo tanto, también lo es su mapa de coeficientes. Un mapa representativo de la regresión DM con núcleo fijo y función ponderada gaussiana se puede observar en la Figura 1. Se observa que el mapa refleja cierta heterogeneidad en sus coeficientes.

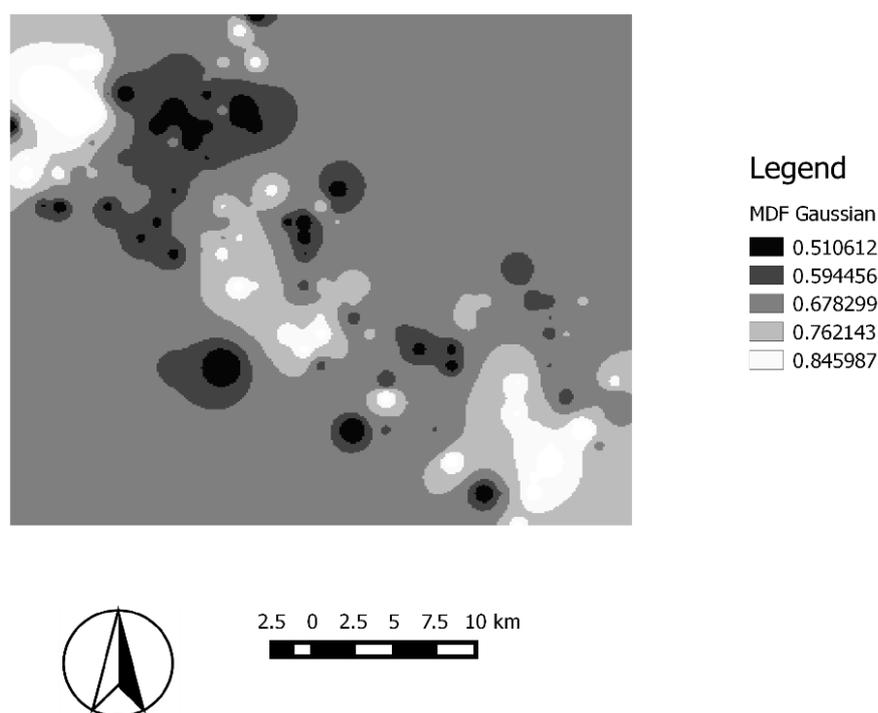


Figura 1. Mapa de coeficientes para el modelo de Distancias de Minkowsky con núcleo fijo y función ponderada gaussiana.

5. Discusión

Los modelos de DE proporcionaron resultados con una calidad relativamente buena excepto en la función boxcar. Esto puede ser debido al gran ancho de banda de dicha función ponderada, lo que produce que la regresión no capte la heterogeneidad de la zona de estudio. Esto equipara esta regresión a una regresión lineal ya que se dan pesos de uno a una gran cantidad de datos (Gollini et al., 2015). Sin embargo, este comportamiento cambia cuando utilizamos los modelos con DM, mejorando, en general, los ajustes. Estos resultados son acordes con estudios previos donde se utilizó las DM para realizar diversos modelos (Binbin et al., 2015). En este estudio se concluía que las DM proporcionan una estimación más exacta de los coeficientes de los modelos, especialmente cuando la distancia óptima es desconocida, por tanto este método proporciona una visión más real de los procesos no estacionarios (Binbin et al., 2015). El mayor problema que se encuentra con esta metodología es su complejidad de aplicación y cálculo.

La autocorrelación espacial fue positiva y próxima a cero en los modelos de DE, sin embargo puede ser rechazado que exista autocorrelación cero en los residuos de acuerdo con el índice Moran I. El cálculo de las matrices de distancia con DM también fue capaz de mejorar la autocorrelación espacial de los residuos. La autocorrelación espacial obtenida coincide con estudios previos donde fueron estudiadas masas forestales agrupadas aleatoriamente y en clústeres (Zhang et al., 2008). En las que se concluía que la presencia de autocorrelación espacial produce estimaciones sesgadas de los parámetros de los modelos y reduce la eficiencia del modelo (Zhang et al., 2008).

Las RPG tienen muchas ventajas sobre otras metodologías tradicionales: (1) es un instrumento muy útil para explorar la heterogeneidad espacial a varias escalas (Jetz et al., 2005), (2) La heterogeneidad espacial se puede mostrar en un mapa de coeficientes, (3) las RPG funcionan con

todo tipo de datos espaciales, lo que es muy útil en casos de inventarios de recursos forestales, (4) se pueden producir mapas que ayudan a interpretar la heterogeneidad espacial (Goodchild & Janelle, 2004) y (5) las RPG no necesitan de mediciones adicionales para realizar predicciones en áreas no muestreadas.

6. Conclusiones

Diferentes calibraciones de RPG fueron realizadas para obtener una parametrización óptima de los modelos altura-diámetro generalizados. Se observó que la calibración básica (DE) proporciona, en general, buenos resultados. Sin embargo, la autocorrelación espacial de los residuos no fue satisfactoria en estos casos. Por tanto, es altamente recomendable re-calcular las matrices de distancia mediante el método de distancias de Minkowsky.

Cuando existe autocorrelación espacial y heterogeneidad en las variables. Las RPG son capaces de ajustar mejor los datos y predice la respuesta con mayor exactitud.

Con las RPG es posible calcular la altura de los árboles en todas las localizaciones dentro del área de estudio sin necesidad de mediciones adicionales ni calibraciones. Además, las RPG pueden ayudarnos a entender procesos espaciales y pueden tener interesantes aplicaciones en áreas como la gestión y producción forestal. Por ejemplo, sería posible desarrollar modelos más precisos de crecimiento y producción, o evaluar el efecto local de la gestión forestal.

Finalmente, este estudio abre futuras posibilidades de investigación de la autocorrelación espacial y posibles factores que expliquen la heterogeneidad espacial en los bosques.

7. Agradecimientos

Este estudio ha sido financiado por la Internal Grant Agency de la Facultad de Forestales y Tecnología de la Madera, Universidad Mendel de Brno. Con el proyecto número LDF_VP_2016012.

8. Bibliografía

BINBIN, L.; CHARLTON, M.; HARRIS, P.; FOTHERINGHAM, S.; 2014. Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data, *International Journal of Geographical Information Science*.

BINBIN, L.; CHARLTON, M.; BRUNSDON, C.; HARRIS, P.; 2015: The Minkowski approach for choosing the distance metric in geographically weighted regression, *International Journal of Geographical Information Science*.

BINBIN, L.; HARRIS, P.; CHARLTON, M.; BRUNSDON, C.; NAKAYA, T.; GOLLINI, I.; 2015. R-CRAN Package "GWModel". Available at: <http://cran.r-project.org/web/packages/GWmodel/GWmodel.pdf> [Accessed 13 November 2015].

BRUNSDON, C.; FOTHERINGHAM, S.; CHARLTON, M.; 1996. Geographically Weighted Regression: A method for explore spatial non-stationarity, *Geographical Analysis* 28 (4): 281–298.

CABANILLAS, A.; 2010. Bases para la gestión de masas naturales de (*Pinus halepensis* Mill.) en el Valle del Ebro. Doctoral Thesis. Dept.de Ingeniería Forestal, Univ. Politécnica deMadrid.

CAÑADAS, N.; GARCÍA, C.; MONTERO, G.; 1999. Relación altura-diámetro para *Pinus pinea* L. en el Sistema central. Actas del Congreso de Ordenación y Gestión Sostenible de Montes, Santiago de Compostela I: 139:153.

CHARLTON, M.; FOTHERINGHAM, S.; 2009. Geographically Weighted Regression. White Paper for Science Foundation, Ireland: 1-14. http://www.geos.ed.ac.uk/~gisteac/fspat/gwr/gwr_arcgis/GWR_WhitePaper.pdf [4 July 24 2015].

CHEN, G.; ZHAO, K.; MCDERMID, G.; HAY, G.; 2012. The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data. International Journal of Remote Sensing IX (33): 2909-2924.

CONDÉS, S.; STERBA, H.; 2008. Comparing an individual tree growth model for *Pinus halepensis* Mill. in the Spanish region of Murcia with yield tables gained from the same area. European Journal of Forest Research 127, 253-261.

FOTHERINGHAM, S.; BRUNSDON, C.; CHARLTON, M.; 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, John Wiley & Sons, New York, NY.

GOLLINI, I.; BINBIN, L.; CHARLTON, M.; BRUNSDON, C.; HARRIS, P.; 2015. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. Journal of Statistical Software 63 (17): 1-50.

GOODCHILD, M.F.; JANELLE, D.G.; 2004. Spatially integrated social science. New York, NY: Oxford University Press.

HART, HMF.; 1928. Stamtal en dunning; een orienteerend onderzoek naar de beste plantwijdteen dunning swijze voor den djati. Veenman & Zonen. Wageningen.

JETZ, W.; AHBEK, C.R.; ICHSTEIN J.W.L.; 2005. Local and global approaches to spatial data analysis in ecology. Global Ecol. Biogeogr. 14:97-98.

MATTHEWS, S.; YANG, TC.; 2012. Mapping the results of local statistics: Using geographically weighted regression. Demographic Research 26(6): 151-166.

MORAN, PAP.; 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37 (1): 17-23.

NANOS, N.; CALAMA, R.; MONTERO, G.; GIL, L.; 2004. Geostatistical prediction of height/diameter models. Forest Ecology and Management 195: 221-23524.

PÁEZ, A.; SCOTT, D.M.; 2004. Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal 61:53-67.

QUIRÓS-SEGOVIA, M.; CONDÉS-RUIZ, S.; DRÁPELA, K.; 2016. Comparison of height-diameter models based on geographically weighted regressions and linear mixed modelling applied to large scale forest inventory data. *Forest Systems*, Volume 25, Issue 3.

R DEVELOPMENT CORE TEAM, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org> [15 December 2015].

TARDANICO, R.; 2006. GWR&2D Mapping. Florida International University. <http://www2.fiu.edu/~tardanic/>

WATSON, DF.; PHILIP, GM.; 1985. A Refinement of Inverse Distance Weighted Interpolation. *Geoprocessing* 2:315-327.

ZHANG, L.; BI, H.; CHENG, P.; DAVIS, CJ.; 2004. Modelling spatial variation in tree diameter-height relationships. *Forest Ecology and Management* 189: 317-329.

ZHANG, L.; MA, Z.; GUO, L.; 2008. Spatially assessing model errors of four regression techniques for three types of forest stands. *Oxford Journals* 81(2): 209 -225.