

**MODELO MATEMÁTICO PARA EVALUAR LA VALORIZACIÓN ENERGÉTICA  
DE BIOMASAS AGRÍCOLAS MEDIANTE DIGESTIÓN ANAEROBIA**

**KAROLTH ROCÍO ESPINOSA JIMÉNEZ  
IVON DANIELA PEÑA ARIAS**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOQUÍMICAS  
ESCUELA DE INGENIERÍA QUÍMICA  
BUCARAMANGA**

**2013**

**MODELO MATEMÁTICO PARA EVALUAR LA VALORIZACIÓN ENERGÉTICA  
DE BIOMASAS AGRÍCOLAS MEDIANTE DIGESTIÓN ANAEROBIA**

**KAROLTH ROCÍO ESPINOSA JIMÉNEZ  
IVON DANIELA PEÑA ARIAS**

**Trabajo de grado para optar al título de:  
Ingeniero Químico**

**Director:  
Humberto Escalante Hernández, PhD**

**Codirectora:  
María Paola Maradei García, PhD**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISCOQUÍMICAS  
ESCUELA DE INGENIERÍA QUÍMICA  
BUCARAMANGA**

**2013**

*Dedico este primer logro en mi vida profesional a Dios, que siempre guía mi camino y llena mi vida de bendiciones, en Él deposito siempre mi confianza porque hace de los sueños una realidad. Gracias al Espíritu Santo que derramó sobre mí fortaleza y discernimiento en los momentos precisos.*

*A mis padres Luis Manuel y Emilse, que son todo en mi vida y el regalo más hermoso que me dio Dios. Gracias por apoyarme de manera incondicional para alcanzar mis metas y enseñarme que el trabajo constante y honesto siempre trae recompensas, a no dejarme derrotar ante los obstáculos y a creer en mi misma.*

*A mi hermanito Juan Manuel que es mi cómplice, él llena mi vida de alegrías y confía en mí más que nadie, siempre agradezco al cielo por su existencia y alegrar cada día de mi vida con sus ocurrencias.*

*A Fabio por su amor incondicional y apoyo en los buenos momentos e incluso los malos haciendo más fácil mi camino, por creer en mí y animarme siempre a seguir. Gracias por llegar a mi vida y quedarte en mi corazón.*

*A mi familia, en especial a mis tías Rosita y Josefina que siempre tienen un consejo y una palabra de aliento que me ayuda a ver la luz en medio de la oscuridad.*

*A mis amigos y todas las personas que han estado presentes en esta etapa de mi vida ( nombrarlos sería una lista interminable), por compartir conmigo momentos inolvidables y permitirme reír y llorar a su lado, especialmente a Mayrita mi confidente y siempre mejor amiga.*

*Karolth*

*A Dios por todas las oportunidades que ha puesto en mi camino, por brindarme la paciencia, la sabiduría y la fortaleza para afrontar cada uno de los retos impuestos por esta carrera y en especial por este trabajo.*

*A mis padres Daniel y Olga por su incondicional apoyo en cada una de las etapas de mi vida. Por su ejemplo, motivación y entrega, por ser los principales responsables de mi personalidad, los principios y los valores que rigen cada paso de mi vida.*

*A mis hermanos David y Adriann, quienes son mis compañeros de vida, mis primeros cómplices, el complemento y la salida de la rutina, por su infinita paciencia y por dar ese toque de picardía a mis días.*

*A Camilo por ser mi gran amigo, mi amor y mi mayor apoyo, por llenar cada momento de alegría, por ayudarme a mantener la confianza y la voluntad en los momentos más difíciles, por recordarme que cada problema tiene una solución y enseñarme a mantener la calma aún en los momentos de mayor desesperación.*

*A mis amigos y familiares, los protagonistas de momentos inolvidables, por brindarme el apoyo, la confianza e innumerables enseñanzas.*

*Daniela*

## **AGRADECIMIENTOS**

Al Doctor Humberto Escalante Hernández y a la Doctora María Paola Maradei García por su orientación, paciencia, disposición y consejos oportunos.

Al Ing. Reynel Rodríguez de la Vega por su dedicación, compromiso, paciencia y constante colaboración en la elaboración de este proyecto.

A la Doctora Liliana del Pilar Castro Molano por aportar sus amplios conocimientos en el proceso de digestión anaerobia y ayudarnos a comprender y fortalecer las base teórica.

Al Doctor Javier Darío Burgos por mostrarnos la luz en los momentos de desesperación, ayudándonos a resolver los inconvenientes estadísticos y matemáticos que se presentaron.

A la Universidad Industrial de Santander y cada uno de los profesores que se encargaron de guiar nuestro aprendizaje de Ingeniería Química, y todas aquellas personas que hicieron parte de nuestra formación profesional.

A nuestros padres, familiares y amigos por su apoyo incondicional y cada uno de sus aportes en el desarrollo de este trabajo.

## TABLA DE CONTENIDO

	pág.
INTRODUCCIÓN .....	18
Digestión anaerobia .....	21
Propiedades de las biomasas como sustrato para DA .....	21
Parámetros de Operación de la DA .....	23
Biogás: Producto de DA.....	25
Herramientas matemáticas para construcción de un modelo de predicción .....	26
Detección de datos anómalos.....	27
Análisis de datos.....	28
Construcción del modelo .....	29
Parámetros estadísticos .....	30
1. DESARROLLO EXPERIMENTAL .....	31
1.1 Primera etapa.....	31
Definición de las variables que influyen en DA. ....	31
Definición de la técnica multivariante a emplear. ....	32
1.2 Segunda etapa: .....	33
Construcción primaria de la base de datos.....	33
Depuración de la base de datos. ....	33
Construcción del modelo. ....	35
Validación del modelo.....	36
1.3 Tercera etapa: .....	36
Evaluación de las principales biomasas agrícolas de Santander, mediante el modelo construido.....	36
2. RESULTADOS Y ANÁLISIS.....	37
2.1 Construcción del modelo, primera aproximación.....	37
2.2 Construcción del modelo de predicción de RCH4 definitivo .....	38
2.3 Validación del modelo .....	39
2.4 Modelo final para la predicción del rendimiento de metano.....	40
2.5 Análisis del comportamiento del modelo .....	41
2.6 Identificación de los residuos agrícolas de Santander idóneos para aplicar digestión anaerobia.....	42

3.	CONCLUSIONES.....	44
4.	RECOMENDACIONES .....	45
	BIBLIOGRAFÍA.....	46
	ANEXOS.....	53

## LISTADO DE TABLAS

<b>Tabla 1.</b> Composición lignocelulosica de algunos residuos agrícolas .....	23
<b>Tabla 2.</b> Ejemplos de biomاسas, con su rendimiento de metano al aplicar DA. ....	25
<b>Tabla 3.</b> Rangos de los parámetros fijos en el proceso de digestión anaerobia...	32
<b>Tabla 4.</b> Variables seleccionadas para la construcción del modelo de predicción de RCH <sub>4</sub> en DA. ....	32
<b>Tabla 5.</b> Parámetros estadísticos de centralización. ....	35
<b>Tabla 6.</b> Primer conjunto de combinaciones de variables. ....	37
<b>Tabla 7.</b> Evaluación de los primeros modelos planteados.....	37
<b>Tabla 8.</b> Categorización de variables en orden ascendente según significancia...	38
<b>Tabla 9.</b> Evaluación de los modelos, según método stepwise (AIC). ....	39
<b>Tabla 10.</b> Constantes para el modelo definitivo.....	41
<b>Tabla 11.</b> Evaluación del rendimiento de metano a partir de la caracterización de algunas biomاسas residuales de Santander. ....	43
<b>Tabla 12.</b> Base de datos para digestión anaerobia. ....	54
<b>Tabla 13.</b> Matriz de coeficiente de Pearson. ....	60
<b>Tabla 14.</b> Prueba del modelo, variando cada descriptor. ....	63

## LISTADO DE FIGURAS

<b>Figura 1.</b> Efecto del tamaño de partícula en el rendimiento de metano a partir de residuos de fibra de fique.....	24
<b>Figura 2.</b> Equivalencias de biogás con otras fuentes de energía.....	26
<b>Figura 3.</b> Herramientas para la construcción de un modelo de predicción.....	27
<b>Figura 4.</b> Etapas para desarrollo del trabajo. ....	31
<b>Figura 5.</b> Diagrama de densidad antes y después de eliminar datos anómalos, respectivamente.....	34
<b>Figura 6.</b> Diagramas de pariedad para los modelos truncados en la variable número 12, 13 y 14 respectivamente mediante el método Stepwise.....	40
<b>Figura 7.</b> Técnicas multivariante a emplear de acuerdo a la dependencia y relación de las variables .....	53
<b>Figura 8.</b> Histograma Celulosa.....	56
<b>Figura 9.</b> Histograma Hemicelulosa. ....	56
<b>Figura 10.</b> Histograma Lignina. ....	56
<b>Figura 11.</b> Histograma C/N. ....	56
<b>Figura 12.</b> Histograma ST. ....	57
<b>Figura 13.</b> Histograma SV.....	57
<b>Figura 14.</b> Histograma RIS.....	57
<b>Figura 15.</b> Histograma RCH4.....	57
<b>Figura 16.</b> Diagrama de dispersión. ....	59
<b>Figura 17.</b> Diagramas de pariedad de los modelos construidos en la primera aproximación.....	61

## LISTADO DE ANEXOS

<b>Anexo A.</b> Método de selección de la técnica multivariante .....	53
<b>Anexo B.</b> Base de datos digestión anaerobia .....	54
<b>Anexo C.</b> Histogramas .....	56
<b>Anexo D.</b> Diagrama de dispersión y matriz de coeficientes de Pearson .....	59
<b>Anexo E.</b> Diagramas de pariedad para los modelos construidos en la primera aproximación.....	61
<b>Anexo F.</b> Variación de datos para prueba.....	63

## LISTADO DE ABREVIATURAS

<b>AIC</b>	Criterio de información Akakike
<b>AM</b>	Análisis multivariante
<b>C/N</b>	Relación Carbono Nitrógeno
<b>Cel</b>	Celulosa
<b>DA</b>	Digestión Anaerobia
<b>DCM</b>	Conjunto de datos construcción del modelo de predicción
<b>DPM</b>	Conjunto de datos prueba del modelo de predicción
<b>Hem</b>	Hemicelulosa
<b>Lig</b>	Lignina
<b>LOF</b>	Local Outlier Factor
<b>RAC</b>	Residuos agrícola de cosecha
<b>RAI</b>	Residuos agrícolas de industria
<b>RCH<sub>4</sub></b>	Rendimiento de metano
<b>RIS</b>	Relación inóculo/sustrato
<b>ST</b>	Sólidos Totales
<b>SV</b>	Sólidos Volátiles

## RESUMEN

**TITULO:** MODELO MATEMÁTICO PARA EVALUAR LA VALORIZACIÓN ENERGÉTICA DE BIOMASAS AGRÍCOLAS MEDIANTE DIGESTIÓN ANAEROBIA \*

**AUTORES:** Espinosa Jiménez, Karolth Rocío  
Peña Arias, Ivon Daniela \*\*

**PALABRAS CLAVES:** Digestión anaerobia, biomasa residual, potencial energético, rendimiento de metano, modelo de predicción.

### DESCRIPCIÓN:

Santander se destaca por la alta producción de cultivos transitorios como papa, yuca, tabaco, frijol, maíz; y cultivos permanentes como caña, piña, naranja, café y cacao. Las actividades de cultivo, cosecha y procesamiento industrial de estos productos, generan grandes cantidades de residuo; los cuales contienen un alto potencial energético, que puede ser aprovechado mediante diferentes tecnologías; una de estas tecnologías es la digestión anaerobia (DA). Dado este potencial, surge la necesidad de tener una herramienta que permita identificar cuáles de las biomasas de Santander son idóneas para su valorización energética mediante DA. El objetivo de este trabajo fue construir esta herramienta, que consiste en un modelo matemático mediante el cual se predice el rendimiento de metano  $RCH_4$  a partir de las propiedades fisicoquímicas de la biomasa a evaluar y la relación inóculo sustrato (RIS). El modelo obtenido tiene un ajuste del 85% y un p-valor de  $1,962e-07$ . Las variables que presentan una mayor influencia en el rendimiento de metano son la relación carbono-nitrógeno (C/N), la lignina, la celulosa y la RIS. Los mejores rendimientos se obtienen a C/N entre 20 y 32, bajos porcentajes de lignina y altos de celulosa. Se realizó la evaluación de algunas biomasas representativas de Santander, obteniendo  $RCH_4$  ente 0,06 y 0,42  $m^3/kg$  SV y una media de 0,19  $m^3/kg$  SV. Las biomasas que presentan mejores rendimientos son el vástago de plátano, la cáscara de papa, el bagazo de fique y el tomate de descarte.

---

\* Trabajo de Grado

\*\* Facultad de Ingenierías Fisicoquímicas. Escuela de Ingeniería Química.

Director: PhD. Humberto Escalante Hernández. Codirector: PhD. María Paola Maradei García.

## ABSTRACT

**TITLE:** MATHEMATICAL MODEL TO EVALUATE THE AGRICULTURAL BIOMASS ENERGY RECOVERY BY ANAEROBIC DIGESTION \*

**AUTHORS:** Espinosa Jiménez, Karolth Rocío  
Peña Arias, Ivon Daniela \*\*

**KEY WORDS:** Anaerobic Digestion, biomass waste, energetic potential, methane yield, prediction model.

### DESCRIPTION:

Santander is known for the high production of seasonal crops such as potatoes, cassava, tobacco, beans, corn, and permanent crops such as sugarcane, pineapple, orange, coffee and cocoa. The activities of growing, harvesting and industrial processing of these products, generate large amounts of waste, which contain a high potential energy, which can be exploited by different technologies, one of these technologies is the anaerobic digestion (AD). Starting from this potential, it is necessary to have a tool to identify which of the biomass of Santander are suitable for energy recovery by DA. The aim of this work was to build this tool, this is a mathematical model by which predicts the performance of methane RCH<sub>4</sub> from physicochemical properties to assess biomass and substrate inoculum ratio (RIS). The resulting model has a setting of 85% and a p- value of 1.962 e- 07. The variables that have a greater influence on the performance of methane are carbon-nitrogen ratio (C/N), lignin, cellulose and RIS. The best yields are obtained at C/N between 20 and 32, low percentages of lignin and high of cellulose. We performed the evaluation of some representative biomass of Santander, getting RCH<sub>4</sub> being 0.06 and 0.42 m<sup>3</sup>/kg VS and an average of 0.19 m<sup>3</sup>/kg VS. The biomasses improved performance are presented stem banana, potato skins, sisal bagasse and discard tomato.

---

\* Thesis

\*\* Physicochemical Engineering Department. Chemical Engineering. Director: PhD. Humberto Escalante Hernández. Codirector: PhD. María Paola Maradei García.

## INTRODUCCIÓN

Santander es el cuarto departamento con mayor producto interno bruto (PIB) agropecuario en Colombia, alrededor del 9% del suelo es utilizado para la agricultura. Santander se destaca por la producción y variedad de cultivos transitorios como papa, yuca, tabaco, frijol, maíz; y cultivos permanentes como caña, piña, naranja, café y cacao. Según el ranking de producción a nivel nacional para el año 2011, Santander se ubica como el principal productor de tabaco y frijol, el segundo de piña y cacao, y el tercero en papa y caña (Cámara de comercio de Bucaramanga, 2012).

De los procesos de cultivo, cosecha y tratamiento industrial, se generan residuos, agrícolas de cosecha (RAC) y agrícolas industriales (RAI). Gran parte de estos residuos contienen un potencial energético aprovechable. Por ejemplo la fibra de palma de aceite (RAI) que posee 7772,9 MJ/t, las hojas y el cogollo de caña de azúcar (RAC) con 15947,69 MJ/t, el vástago de plátano (RAC) con 2818,81 MJ/t de, el bagazo de caña de azúcar (RAI) con 20393,60 MJ/t, los tallos de café (RAC) con 40921,59 MJ/t, la pulpa de café (RAI) con 7647,87 MJ/t, la tuza de maíz (RAC) con 2809,27 MJ/t y la cascarilla de arroz (RAI) con 2896,68 MJ/t, (Escalante *et al.*, 2010).

El aprovechamiento energético de las biomásas requiere la selección de una tecnología apropiada. Tradicionalmente, se utilizan procesos termoquímicos o bioquímicos. Los procesos termoquímicos implican reacciones químicas irreversibles y temperaturas elevadas. Estos procesos se han aplicado en la transformación de biomásas como raquis de palma de aceite mediante pirólisis (Tarazona, 2013), maíz a partir de combustión (Guerrero & Gamboa, 2009) y cáscara de café por medio de gasificación (Šulc *et al.*, 2011). Por otro lado los procesos bioquímicos requieren la presencia de microorganismos para aprovechar las biomásas, cuyas propiedades fisicoquímicas influyen en el rendimiento del

proceso. Como ejemplos de la aplicación de estos procesos se encuentran: la transformación de bagazo de caña de azúcar mediante fermentación (Almezquita, 2007), cascarilla de arroz a partir de compostaje (Hosseini & Aziz, 2005) y bagazo de fique por medio de digestión anaerobia (Castro, 2012).

La digestión anaerobia (DA) consiste en degradar la materia orgánica, debido a la acción microbiana, para obtener un biogás (compuesto principalmente por metano). Al cuantificar el rendimiento de metano es posible determinar la calidad del biogás. Para un adecuado aprovechamiento de la biomasa por DA, son necesarios bajos niveles de lignina y elevados de celulosa (Appels *et al.*, 2011) (Triolo *et al.*, 2011) y relaciones C/N entre 25 – 32 (Haladová *et al.*, 2011). Es importante un alto contenido de sólidos volátiles (SV), como el rastrojo de maíz que contienen 92,6 % p/p de SV y 97 % p/p de ST (Liew *et al.*, 2011), los tallos de girasol con 88 % p/p y 94 % p/p de SV y ST respectivamente (Monlau *et al.*, 2012); sin embargo si el contenido de celulosa es bajo y el de lignina alto, se obtendrá un bajo rendimiento del proceso.

Los parámetros de operación más relevantes de la DA son la temperatura, el pH, el tamaño de partícula, el tipo de inóculo y la relación inóculo sustrato (RIS). La temperatura se puede establecer dentro de los rangos mesofílicos o termofílicos, siendo los mesofílicos los más utilizados. El pH debe mantenerse dentro 6,8 y 7,2 para garantizar el crecimiento microbiano (Cioabla *et al.*, 2012). El diámetro de partícula ideal está alrededor de los 2 mm (Mshandette *et al.*, 2006). Por otro lado también es importante seleccionar el inóculo más apropiado (lodos PTAR, estiércol de cerdo o ganado, gallinaza, etc...) y una adecuada relación inóculo/sustrato (RIS), la cual debe seleccionarse cuidadosamente para cada biomasa con un estudio previo que permita optimizar el proceso; por ejemplo el estudio realizado para la pulpa de yuca (Panichnumsin *et al.*, 2010), donde se prueban diferentes RIS siendo la mejor de  $1,0 \%SV_{inóculo}/\%SV_{sustrato}$ .

El análisis multivariante (AM) es una metodología útil para evaluar la valorización energética de residuos agrícolas por medio de DA. El AM es un método estadístico que analiza simultáneamente medidas múltiples de cualquier estudio que tenga más de dos variables. Permite resolver problemas de toma de decisión abordándolos de forma sistemática y ordenada. El AM permite que todos los factores que afectan al problema participen en la decisión y suministren una amplia información, facilitando la búsqueda de la solución.

Un modelo matemático que relacione las variables de operación y de respuesta, es una herramienta muy útil en el momento de la toma de decisión. El modelo puede construirse mediante diferentes técnicas multivariantes que dependen de la relación de dependencia y de la linealidad o no del modelo, que se desea construir (Hair *et al*, 1999).

El objetivo de este trabajo fue identificar cuáles de las biomásas procedentes de 16 cultivos de Santander, son idóneas para su valorización energética mediante DA. Para esto es necesario diseñar una herramienta que permita determinar para cada biomasa el rendimiento del proceso en función de la productividad de biogás. Esta herramienta consiste en un modelo matemático que correlacione las propiedades fisicoquímicas de la biomasa y las condiciones de operación con el rendimiento de metano.

Un modelo es la representación matemática de un sistema, por lo tanto es necesaria la elaboración de una base de datos que suministre información sobre el comportamiento de las variables que afectan el sistema. La base de datos permite aplicar un método de regresión que pueda correlacionar las propiedades fisicoquímicas de la biomasa y las condiciones de operación, con el rendimiento de metano. La aplicación de este modelo a las biomásas caracterizadas, permitirá identificar cuáles de estas son aptas para aplicar DA.

## ***Digestión anaerobia***

La digestión anaerobia (DA) se ha utilizado para el tratamiento de residuos orgánicos industriales y desechos agrícolas (*Namsree et al.*, 2012), algunos ejemplos de aplicación de DA en residuos agrícolas con su respectivo rendimiento de metano son: pulpa de yuca con 0,29 m<sup>3</sup>/kg SV (*Panichnumsin et al.*, 2010), tomate con 0,32 m<sup>3</sup>/kg SV (*Jagadabhi et al.*, 2011) y maíz con 0,405 m<sup>3</sup>/kg SV (*Triolo et al.*, 2011)

En DA se descompone la materia orgánica, mediante la actuación de un consorcio microbiano en ausencia de oxígeno. La DA, a pequeña escala y a escala industrial, puede llevarse a cabo en países en desarrollo así como en zonas rurales con un suministro de energía prácticamente nulo (*Appels et al.*, 2011). La DA se lleva a cabo en cuatro etapas: hidrólisis, acidogénesis, acetogénesis y metanogénesis (*Martí 2006*). En la hidrólisis los compuestos de mayor peso molecular (polímeros) se convierten en compuestos orgánicos solubles mediante enzimas producidas por los microorganismos hidrolíticos. En la etapa acidogénica se fermentan las moléculas orgánicas solubles denominadas ácidos grasos volátiles (AGV) (algunos productos puede pasar directamente a la etapa 4). En la acetogénesis, los ácidos grasos volátiles y algunos aromáticos se transforman a productos más sencillos: acetato e hidrógeno, a través de las bacterias acetogénicas. Finalmente en la metanogénesis, ocurre la digestión anaerobia, es decir la formación de metano, mediante las bacterias metanogénicas.

### ***Propiedades de las biomasas como sustrato para DA***

Una biomasa apta para ser digestada debe aportar la fuente de carbono y de energía, permitiendo maximizar el uso de nutrientes como nitrógeno y fósforo (*Ward et al.*, 2008). La biomasa seleccionada debe contar con determinadas características, las cuales pueden ser expresadas en función de:

- **% Sólidos Volátiles (SV):** representan la cantidad de materia disponible para la degradación. Un kg de SV produce 0,5 metros cúbicos de metano (Hills & Roberts, 1981), pero sólo aproximadamente la mitad de los SV añadidos a un digestor pueden ser digeridos (Haladová *et al.*, 2011).
  
- **% Sólidos Totales (ST):** Están directamente relacionados con el contenido de humedad de la biomasa. El contenido de ST, permite clasificar la digestión anaerobia en tres grupos: digestión seca (ST < 10%), digestión semiseca (ST 10%-20%) y digestión húmeda (ST > 20%).
  
- **% Lignina:** Polímero ramificado, amorfo y heterogéneo, que forma un escudo alrededor de la celulosa y la hemicelulosa, impidiendo su degradación (Castro, 2012). La lignina es difícilmente biodegradable por los microorganismos anaeróbicos, lo que provoca una digestión baja (Appels *et al.*, 2011). El contenido de lignina es característico de cada biomasa (Tabla 1).
  
- **% Celulosa:** Es un polisacárido insoluble en agua, compuesto por D-Glucosa, unidas por enlaces glucosídicos  $\beta$  1,4 (Castro, 2012). La celulosa es el principal componente de la biomasa residual agrícola, se ha demostrado que puede ser degradada anaeróbicamente hasta aproximadamente un 80%. (Appels *et al.*, 2011).
  
- **% Hemicelulosa:** Carbohidratos que forman una estructura polimérica compleja, de la unión de diferentes azúcares. La hemicelulosa es más fácil de hidrolizar que la celulosa (Castro, 2012). De sustratos con alta cantidad de hemicelulosa y lignina resultan bajos rendimientos de metano, debido a que la digestión de la hemicelulosa se ve afectada por el tipo y la cantidad de lignina (Jagadabhi *et al.*, 2011). La hemicelulosa junto con la lignina actúan como una capa protectora que hace que la celulosa sea resistente a la digestión enzimática (Triolo *et al.*, 2011).

Tabla 1. Composición lignocelulósica de algunos residuos agrícolas

Biomasa residual	% Celulosa	% Hemicelulosa	% Lignina
Mazorca	45	35	15
Paja de arroz	32,1	24	18
Bagazos	33,4	30	18,9
Bagazo de fique	41,81	22,17	15,56
Paja de trigo	30	50	15
Pasto	45	31,4	12

Fuente: (Castro, 2012)

- **Relación Carbono-Nitrógeno (C/N):** Cuando la relación C/N es demasiado grande, no se da una conversión óptima de carbono en CH<sub>4</sub>; este problema se puede superar con una co-digestión de sustratos con una relación C/N estrecha (Amon *et al.*, 2007). Se recomienda una relación C/N alrededor de 20/1 - 32/1. Diferentes proporciones ocasionan una disminución en la actividad metanogénica disminuyendo la formación de biogás (Haladová *et al.*, 2011).

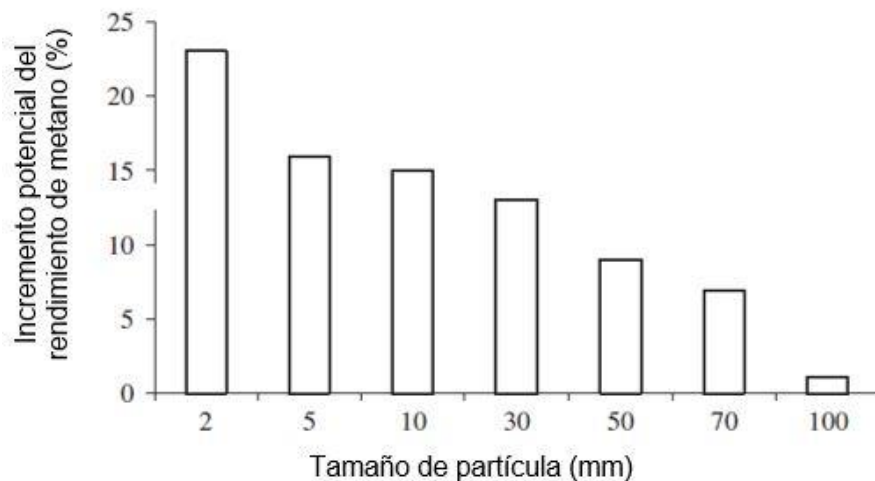
#### Parámetros de Operación de la DA

- **Temperatura:** La DA se puede llevar a cabo a temperaturas psicrófilas, mesófilas o termófilas. La mayoría de estos procesos se realizan en el rango mesófilo, debido a que requieren un menor consumo energético y son más estables (Castro, 2012).
- **pH:** El pH influye directamente en la DA, ya que si es muy bajo, reduce la tasa de crecimiento de los microorganismos metanogénicos. El intervalo de pH más factible para DA oscila entre 6,8 – 7,2 (Ward *et al.*, 2008), sin embargo, es posible obtener buen rendimiento en DA a pH entre 6,5 y 8 (Cioabla *et al.*, 2012).
- **Humedad:** En DA húmeda se necesita poco tiempo para la puesta en marcha del proceso, pero se forma fácilmente escoria flotante. La DA seca necesita más tiempo para alcanzar el más alto contenido de metano, pero el sistema es relativamente estable (Lianhua *et al.*, 2010). La mayoría de procesos

tiene lugar en un entorno acuoso y puede ser procesadas sin ningún tratamiento previo (secado) (Ward *et al.*, 2008).

- **Tamaño de las partículas:** El tamaño de partícula del sustrato influye en la velocidad de DA; a menor tamaño, mayor área de superficie y por ende mayor facilidad para que los microorganismos accedan a la fuente de carbono. Durante la DA de vegetales se mejora la degradación de la fibra y el rendimiento de metano con la disminución del tamaño de partícula (Figura 1) (Mshandete *et al.*, 2006).

Figura 1. Efecto del tamaño de partícula en el rendimiento de metano a partir de residuos de fibra de fique.



Fuente: (Mshandete *et al.*, 2006)

- **Relación inóculo / sustrato (RIS):** El rendimiento del proceso de DA depende de la actividad hidrolítica y metanogénica de los inóculos y del potencial de biometanización del sustrato, por tal motivo es necesario encontrar una RIS apropiada (Castro, 2012). Se han utilizado RIS de 0,2 a 0,5 %SV<sub>inóculo</sub>/ %SV<sub>sustrato</sub> para la digestión de rastrojo de maíz con lodos PTAR, obteniéndose el mejor RCH<sub>4</sub> con una RIS de 0,5 (Liew *et al.*, 2011). Por otro lado se emplearon RIS de 0,25 a 4 %SV<sub>inóculo</sub>/ %SV<sub>sustrato</sub> para pulpa de maíz con estiércol de cerdo, obteniendo los mayores RCH<sub>4</sub> para RIS entre 0,5 y 1 (Panichnumsin *et al.*, 2010).

### Biogás: Producto de DA

Como productos del proceso de digestión anaerobia, se encuentran el efluente depurado y el biogás; siendo este último el que se puede aprovechar energéticamente. El biogás a partir de residuos agrícolas está compuesto por pequeñas proporciones de nitrógeno (0-1%), hidrógeno (0-2%), oxígeno (0-1%) monóxido de carbono (0-1%), sulfuro de hidrógeno (100 – 700 ppm) y amoníaco (trazas); además de dióxido de carbono (30-50%) (Montes, 2008) y mayoritariamente por metano (50 – 70%) (IDAE, 2007).

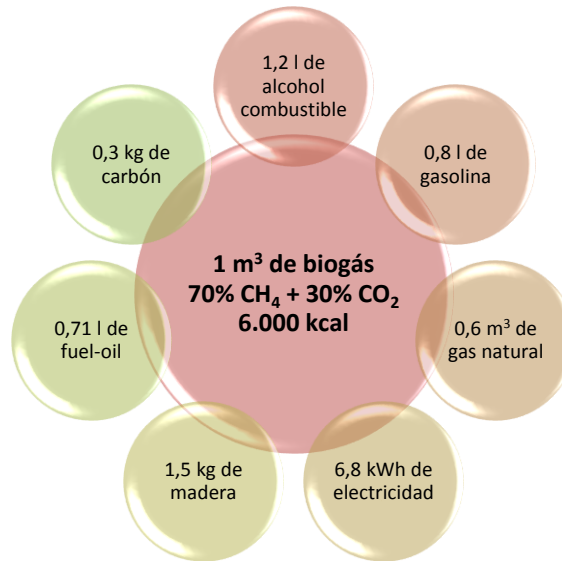
La calidad de biogás se mide mediante el RCH<sub>4</sub>, dado que el metano es el componente mayoritario. Para algunas biomásas de frutas y vegetales, se obtienen RCH<sub>4</sub> de 0,19 hasta 0,5 m<sup>3</sup>/kg SV (Tabla 2). El biogás tiene un poder calorífico de 6,4 kWh/Nm<sup>3</sup>, por lo cual puede competir con otros combustibles como una fuente de energía alternativa. (Figura 2).

*Tabla 2. Ejemplos de biomásas, con su rendimiento de metano al aplicar DA.*

<b>Biomasa Alimentada</b>	<b>Rendimiento de metano m<sup>3</sup>/kg SV</b>
Cáscara de banano	0,277
Mango	0,373
Limón prensado	0,473
Tomate podrido	0,298
Cáscara exterior de Cebolla	0,4
Hojas de Coliflor	0,19
Vástago de Coliflor	0,331
Cáscara de Papa	0,267
Hojas de Zanahoria	0,241
Hojas de Remolacha	0,231

Fuente: (Ward *et al.*, 2008)

Figura 2. Equivalencias de biogás con otras fuentes de energía.



Fuente: (IDAE, 2007)

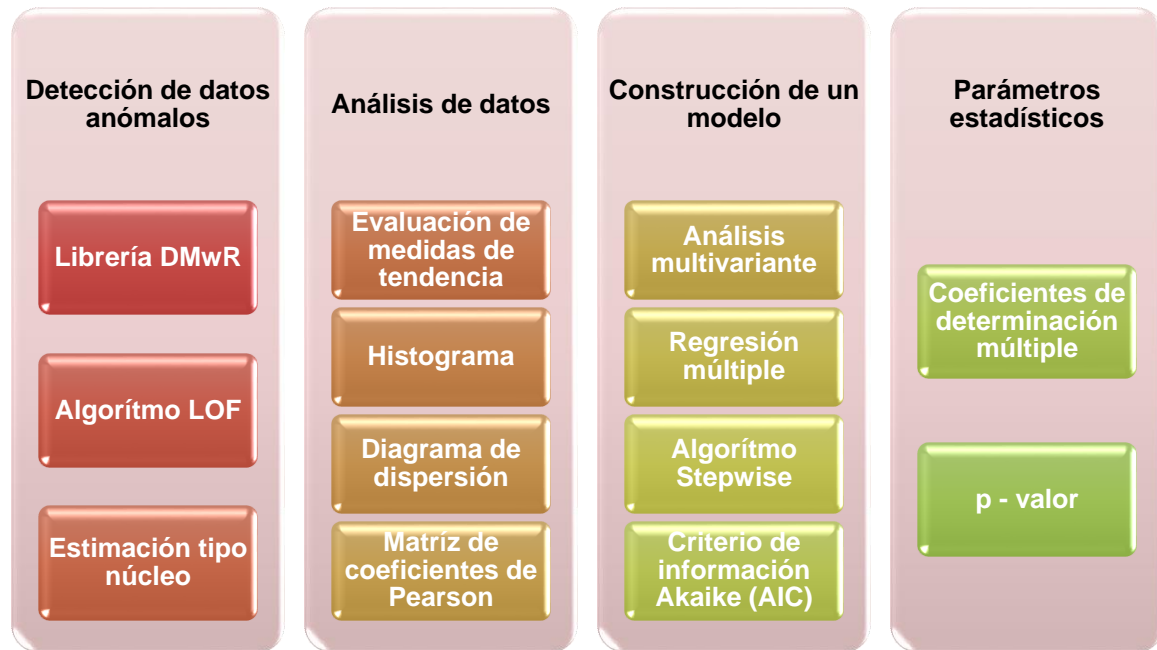
### ***Herramientas matemáticas para construcción de un modelo de predicción***

Las herramientas que pueden ser usadas para la construcción de un modelo de predicción se registran en la Figura 3. Estas herramientas hacen parte del software estadístico R.

**R** versión 2.15.2; es un software libre de GNU dedicado al cálculo estadístico y la generación de gráficas. R contiene funciones agrupadas en paquetes o librerías que son elaborados por diferentes programadores. Dos de dichas librerías son: DMwR (para el análisis de datos anómalos) y Hmisc (para la matriz de correlación).

**RStudio** versión 0.97.318, es una interfaz gráfica de desarrollo integrado (IDE) de código abierto de R. RStudio se puede ejecutar sobre distintas plataformas (Windows, Mac, o Linux), incluso desde la web usando RStudio Server.

Figura 3. Herramientas para la construcción de un modelo de predicción



Fuente: [Autor]

### Detección de datos anómalos

Se entiende por datos anómalos a las observaciones atípicas presentes en un conjunto de datos, las cuales se alejan de la tendencia central y provocan que el conjunto se aparte de una distribución normal. Los datos anómalos deben identificarse y eliminarse para evitar desviación en los resultados.

- **Librería DMwR (Funciones y datos para la “Exploración de datos con R”):** Contiene una serie de funciones y procedimientos encaminados a la exploración de datos con R (Torgo, 2010). Dentro de este paquete se encuentra la función [lofactor()]; la cual obtiene factores atípicos locales utilizando el algoritmo LOF (Breunig et. al., 2000). Es decir, dado un conjunto de datos se produce un vector de factores de valores atípicos locales para cada caso.

- **Algoritmo LOF:** El algoritmo LOF está basado en la densidad y fue creado específicamente para detectar datos anómalos; este da como resultado un valor (Local Outlier Factor) de un objeto  $p$  que representa el grado en que  $p$  es un anómalo (Kuna *et al.*, 2011).
- **Estimación tipo núcleo:** También conocido como estimación tipo Kernel, es un método de estimación no paramétrica de la densidad; es decir cuando no se quiere asumir una hipótesis sobre la distribución de la muestra (Cañette, 2002). En R se lleva a cabo mediante la función genérica de densidad [*density()*].

### Análisis de datos

- **Evaluación de medidas de tendencia:** Son cantidades numéricas calculadas sobre una muestra que resume el estudio realizado. Estos parámetros pueden ser de centralización, como la media aritmética conocida también como promedio, la moda que es el valor que más se repite en la muestra, la mediana es el valor que se encuentra en la mitad al ordenar los datos de menor a mayor. Por otro lado están los parámetros o medidas de dispersión como la varianza y la desviación estándar; estas presentan la variabilidad de una muestra de datos y que tan alejados se encuentran de la media.
- **Histograma:** Es una representación aproximada de la población, en la cual se puede observar la tendencia de los datos y la frecuencia con la que se presentan cada uno de ellos dentro de un rango determinado.
- **Diagrama de dispersión:** Describe las relaciones entre dos variables, representando cada observación por un punto en un plano; estas relaciones se pueden determinar según el acercamiento o agrupamiento de los puntos representados. La agrupación de puntos alrededor de una recta indicaría una relación lineal entre las variables en estudio. Por el contrario, si los puntos

representados se encuentran totalmente dispersos, entonces no existe relación alguna entre las variables representadas.

- **Matriz de correlación de Pearson:** Compuesta por los coeficientes de correlación de Pearson para todo los posibles pares de columnas de una matriz entre dos variables; estos permiten determinar el grado de relación entre dos variables cuantitativas. En la matriz de correlación de Pearson para coeficientes entre 0,65 y 1 se considera una relación directamente proporcional. Para coeficientes entre -0,65 y -1 la relación sería inversamente proporcional. Esta matriz es de gran utilidad en el momento de suponer relaciones y posibles combinaciones de variables al construir el modelo.

#### Construcción del modelo

- **Análisis multivariante:** Es un método estadístico que analiza simultáneamente medidas múltiples de cualquier estudio que tenga más de dos variables (Hair *et al.*, 1999). Las técnicas multivariantes se clasifican en análisis de dependencia y de interdependencia (anexo A).

- **Regresión Múltiple:** Se utiliza para estudiar la posible relación entre una variable dependiente y un conjunto de variables independientes. Un modelo de regresión múltiple tiene la siguiente forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Donde Y es la variable a predecir y  $\beta$  son los parámetros que se desean estimar.

- **Algoritmo Stepwise:** Consiste en la determinación del grado de significancia de cada una de las variables involucradas en el sistema. Se encarga de ir agregando una a una las variables más significativas hasta que el modelo cumpla con las condiciones requeridas de predicción o margen de error establecido. Este algoritmo usa como validación el método de los mínimos cuadrados.

- **Criterio de información Akaike:** Permite la comparación entre dos o más modelos estadísticos que tienen en común una misma variable dependiente y un mismo conjunto de datos.

La función `step()` de R incluye el algoritmo Stepwise y el criterio de información Akaike, con los cuales se obtiene una lista de las variables involucradas en orden de significancia. Una vez organizadas las variables se procede a armar los modelos agregando una a una hasta encontrar un modelo que cumpla con las condiciones deseadas.

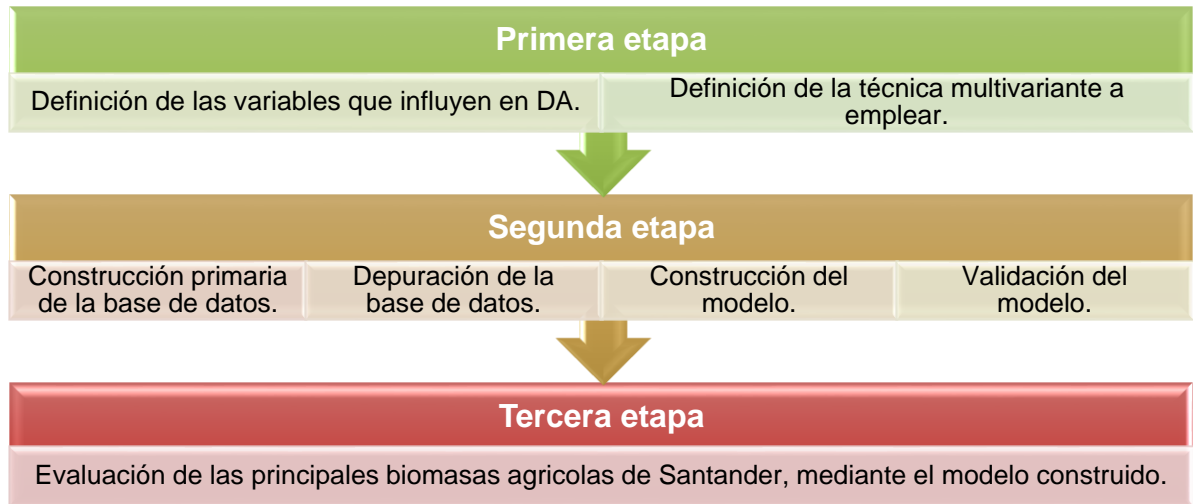
### Parámetros estadísticos

- **Coefficientes de determinación múltiple,  $R^2$ , y  $R^2$  ajustado,** son medidas utilizadas en el análisis de regresión; estas indican el porcentaje de varianza que logran justificar las variables independientes de un modelo. Los valores de  $R^2$  más cercanos a 1 indican mejores ajustes y muestran la relación lineal entre las variables. El valor del  $R^2$  ajustado es ligeramente menor al de  $R^2$  múltiple, porque tienen en cuenta el tamaño de conjunto de datos. Es importante recalcar que estos coeficientes son una medida de ajuste de los datos y no indican por sí mismos la capacidad de predicción de un modelo.
- **p – valor:** Para explicar el p – valor es necesario definir la hipótesis nula, que indica que no existe relación entre las variables. El p – valor es la probabilidad de tener un resultado si la hipótesis nula es verdadera. Si el p – valor es 0, quiere decir que el resultado no apoya la hipótesis nula, es decir, que sí existe relación entre las variables. Por tal motivo, el p – valor debe ser lo más cercano posible a cero, normalmente se permiten valores inferiores a 0,05.

## 1. DESARROLLO EXPERIMENTAL

La metodología empleada para el desarrollo del trabajo, se divide en tres etapas (Figura 4).

Figura 4. Etapas para desarrollo del trabajo.



Fuente: [Autor]

### 1.1 Primera etapa

#### Definición de las variables que influyen en DA.

Se realizó una revisión bibliográfica para definir las variables que influyen en el proceso de DA; estas corresponden a propiedades fisicoquímicas de las biomásas y/o condiciones de operación. La revisión se realizó tomando artículos de los últimos quince años, dónde se reportan ejemplos de aplicación de DA, incluyendo la caracterización de las biomásas y los resultados del rendimiento de metano correspondiente.

De los parámetros de operación para DA, se seleccionaron aquellos que deben ser controlados en un rango específico. La mayoría de procesos de DA reportados

son batch a temperaturas mesofílicas. En algunos artículos, el pH de las biomásas encontradas fue ajustado mediante pretratamientos para llevarlo al rango ideal (6,5 – 8). Los parámetros seleccionados y sus rangos se presentan en la Tabla 3.

*Tabla 3. Rangos de los parámetros fijos en el proceso de digestión anaerobia.*

Variable	Rangos en los que se construyó el modelo
Temperatura	35 °C – 40 °C
pH	4,7 – 7,9
Humedad	0 % – 100 %

Fuente: [Autor]

Las variables seleccionadas como las más influyentes en el proceso de DA, se presentan en la Tabla 4, donde se indica la abreviatura a utilizar de ahora en adelante, las unidades en que se encuentra y el tipo de variable (entrada o salida).

*Tabla 4. Variables seleccionadas para la construcción del modelo de predicción de RCH4 en DA.*

Descripción	Abreviatura	Tipo	Unidades
Celulosa	Cel	Variable de entrada	% p/p
Hemicelulosa	Hem	Variable de entrada	% p/p
Lignina	Lig	Variable de entrada	% p/p
Relación Carbono/Nitrógeno	C/N	Variable de entrada	%ST <sub>C</sub> / %ST <sub>N</sub>
Relación Inóculo/Sustrato	RIS	Variable de entrada	%SV <sub>inóculo</sub> / %SV <sub>sustrato</sub> .
Sólidos Totales de la Biomasa	ST	Variable de entrada	% p/p
Sólidos Volátiles de la Biomasa	SV	Variable de entrada	% p/p
Rendimiento de metano	RCH4	Variable de salida	m <sup>3</sup> /kg SV

Fuente: [Autor]

#### Definición de la técnica multivariante a emplear.

El RCH4 es la única variable de salida del modelo y depende de las demás variables presentadas en la Tabla 4. Dado que todas las variables que influyen en el proceso son cuantitativas y presentan un comportamiento no lineal, se

seleccionó la regresión múltiple como técnica multivariante. El algoritmo para la toma de esta decisión se encuentra en el Anexo A.

## **1.2 Segunda etapa:**

### Construcción primaria de la base de datos.

Mediante la revisión bibliográfica se registraron en una hoja Excel 83 observaciones, tomadas de 15 artículos de los últimos 15 años. En las columnas de la tabla Excel se registraron las variables seleccionadas (Tabla 4) y en las filas las observaciones encontradas para cada biomasa, con su respectiva caracterización fisicoquímica y rendimiento de metano.

Dado que la información de los artículos estaba reportada en diferentes unidades, se hizo la respectiva conversión para utilizar puntos de comparación. Algunos rendimientos de metano se encontraban en unidades de L/kg ST, por tal motivo se unificaron todas a L/kg SV. Para un mejor análisis del rendimiento de metano y comparación con experimentos reportados en la literatura, se transformaron a m<sup>3</sup>/kg SV.

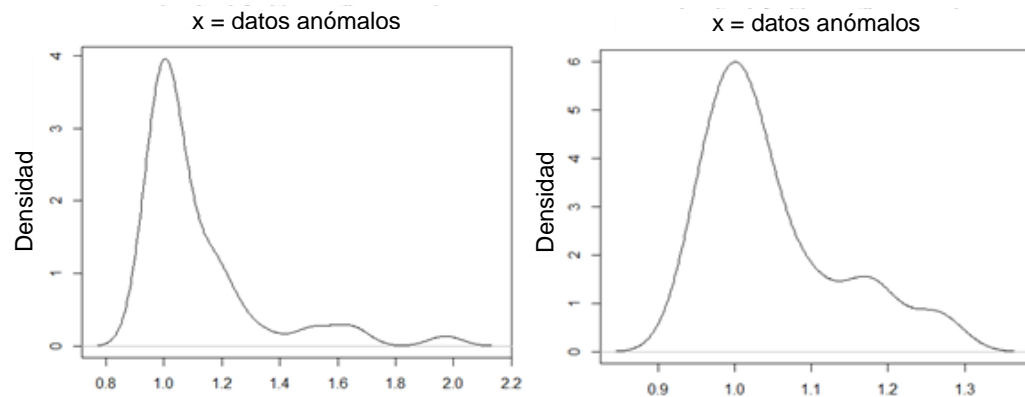
### Depuración de la base de datos.

La base de datos que se construyó inicialmente contó con 83 observaciones y al realizar el análisis de depuración, se redujo a 51 observaciones. Se eliminaron los datos más incompletos y aquellos registros duplicados para una misma biomasa. En el Anexo B se presenta la base de datos final, donde se reporta la fuente de la cual se tomó cada observación y las variables que fueron seleccionadas de otra fuente o imputadas.

Dentro del conjunto de biomazas se consideran atípicas las observaciones que poseen elevados rendimientos de metano ( $> 0,5 \text{ m}^3/\text{kg SV}$ ); para alcanzar estos valores es necesario someter la biomasa a pretratamientos. Así mismo, elevadas o mínimas relaciones C/N se consideran anómalas, puesto que estas suponen exceso y ausencia de material disponible para la degradación respectivamente.

Al aplicar el algoritmo LOF para detección de datos anómalos se descartaron 6 datos, debido a que no reportaban la relación C/N (Anexo B) y requirieron calcularse mediante imputación o lectura en otra fuente. La base de datos definitiva estuvo compuesta por 45 observaciones. En la Figura 5 se observa que la base de datos inicialmente tenía varios datos anómalos y al realizar el filtrado, el conjunto de datos presenta distribución normal.

*Figura 5. Diagrama de densidad antes y después de eliminar datos anómalos, respectivamente.*



Fuente: [Autor]

Al finalizar la depuración de la base de datos se dividieron de manera aleatoria las observaciones en dos conjuntos, uno con el 70% de los datos para la construcción del modelo (DCM) y el otro con el 30% restante para prueba (DPM). El primer conjunto formado por 32 observaciones, y el segundo con 13.

A partir de la base de datos final, se determinaron los parámetros estadísticos de centralización para cada variable que se condensan en la Tabla 5; esto con el fin de estudiar los intervalos en los que se encontraba cada dato y asegurar que la distribución tienda a estar normalizada. Lo anterior se apoyó con la realización del histograma de cada variable que se encuentran en el Anexo C.

*Tabla 5. Parámetros estadísticos de centralización.*

	<b>Cel</b>	<b>Hem</b>	<b>Lig</b>	<b>C.N</b>	<b>SV</b>	<b>ST</b>	<b>RIS</b>	<b>RCH4</b>
<b>Min.</b>	17,21	17,02	2,58	16,12	2,77	6,20	0,30	0,028
<b>Median</b>	47,78	37,26	9,46	38,97	28,40	30,20	0,90	0,283
<b>Mean</b>	43,40	41,64	14,96	41,80	42,77	48,51	0,92	0,250
<b>Max.</b>	60,26	72,77	50,33	76,33	94,60	98,00	2,40	0,439

Fuente: [Autor]

### Construcción del modelo.

El modelo involucra las relaciones entre variables, para lo cual se realizó el diagrama de dispersión de cada variable respecto a las demás y la matriz de coeficiente de Pearson (Anexo D). A partir de los anteriores análisis, fue posible suponer relaciones entre variables, con las cuales se diseñaron diferentes modelos combinando variables acorde a lo indicado por los coeficientes de Pearson:

$$RSTV = \frac{ST}{SV} \quad LH = \frac{Lig}{Hem} \quad LCH = \frac{Cel}{Hem * Lig}$$

Para seleccionar un buen modelo de las combinaciones realizadas fue necesario estudiar cada uno de los valores estadísticos. Asegurar que el R-ajustado debe ser mayor a 0,8, el p-valor inferior a 0,05. Se deben evitar los problemas de sobreajuste, con el fin de que obtener valores de RCH4 coherentes, es decir que el rendimiento de metano de valores positivos y dentro de los rangos en los cuales se construyó el modelo.

Se implementó el algoritmo stepwise para realizar una selección más apropiada de las combinaciones de variables, mediante el criterio de información Akaike (AIC). El modelo se construyó agregando variable a variable de acuerdo al orden indicado por el criterio (AIC) y evaluando los diferentes parámetros estadísticos.

#### Validación del modelo.

En esta etapa se utilizó un conjunto de datos de prueba, para evaluar la precisión del modelo. Mediante la construcción de un diagrama de paridad, se graficaron los RCH4 calculados por el modelo contra los RCH4 de la base de datos, graficando también la recta  $Y=X$  y el margen de error del 20%. El modelo final seleccionado, correspondió al encontrado dentro del margen de error, esto se traduce en un modelo capaz de predecir.

### **1.3 Tercera etapa:**

#### ***Evaluación de las principales biomásas agrícolas de Santander, mediante el modelo construido.***

Se organizaron las propiedades de las 16 biomásas a evaluar en una hoja Excel, de manera que se eliminaron los datos que estaban fuera de los rangos en los cuales fue construido el modelo. Las 16 biomásas se evaluaron mediante el modelo construido, para determinar con cuales se obtendrá un mejor rendimiento de metano al aplicar DA.

## 2. RESULTADOS Y ANÁLISIS

### 2.1 Construcción del modelo, primera aproximación.

En la Tabla 6 se presentan las combinaciones de variables planteadas inicialmente. En la Tabla 7 se presentan el  $R^2$ , el  $R^2$ -ajustado para cada modelo planteado y su respectiva observación. Los modelos 8 y 9 cumplen con los requisitos básicos, R-ajustado mayores a 0,8, variables significativas y sin sobreajuste.

Tabla 6. Primer conjunto de combinaciones de variables.

Modelo No.	Combinaciones planteados
M-1	$RCH4 \propto \text{Cel} * \text{Hem} * \text{Lig} * \text{C.N} * \text{SV} * \text{ST} * \text{RIS}$
M-2	$RCH4 \propto \text{Cel} + \text{Hem} + \text{Lig} + \text{C.N} + \text{SV} + \text{ST} + \text{RIS}$
M-3	$RCH4 \propto \text{Cel} + \text{Hem} + \text{Lig} + \text{C.N} + \text{RSTV} + \text{RIS}$
M-4	$RCH4 \propto \text{LCH} + \text{C.N} + \text{RSTV} + \text{RIS}$
M-5	$RCH4 \propto \text{LH} + \text{Cel} + \text{C.N} + \text{RSTV} + \text{RIS}$
M-6	$RCH4 \propto \text{Cel} * (\text{Hem} + \text{Lig}) + \text{C.N} + \text{RSTV} + \text{RIS}$
M-7	$RCH4 \propto \text{C.N} + \text{RSTV} + \text{RIS} + \text{RSTV} * (\text{Hem} + \text{Lig} + \text{Cel})$
M-8	$RCH4 \propto \text{C.N} * (\text{Cel} + \text{Hem} + \text{Lig}) + \text{RSTV} * (\text{Cel} + \text{Hem} + \text{Lig}) + \text{RIS}$
M-9	$RCH4 \propto \text{C.N} * (\text{Cel} + \text{Hem} + \text{Lig} + \text{RIS}) + \text{RSTV}$

Fuente: [Autor]

Tabla 7. Evaluación de los primeros modelos planteados.

Modelo No.	R-cuadrado	R-ajustado	Observaciones
M-1	1,0000000	----	Sobreajustado
M-2	0,8284555	0,7784217	Variables no significativas
M-3	0,8266989	0,7851066	Variables no significativas
M-4	0,6036652	0,5449490	R-ajustado < 0,8
M-5	0,7891516	0,7486039	R-ajustado < 0,8
M-6	0,8608918	0,8125063	Variables no significativas
M-7	0,8438798	0,7800124	Variables no significativas
M-8	0,9259287	0,8791468	No predice valores lógicos
M-9	0,9332012	0,9013923	No predice valores lógicos

Fuente: [Autor]

Aunque los modelos 8 y 9 cumplen con los parámetros estadísticos exigidos, al realizar el diagrama de pariedad (Anexo E) y evaluar su capacidad de predicción, estos no logran predecir valores lógicos. En algunos casos al aplicar el modelo se obtuvieron rendimientos negativos o exageradamente altos y con muchos puntos fuera del margen de error.

## 2.2 Construcción del modelo de predicción de RCH4 definitivo

Para la construcción del modelo definitivo, fue necesario agregar todas las combinaciones posibles entre las variables (producto de unas con otras y entre sí mismas). Se hicieron 35 combinaciones en total, de las cuales se muestran las 17 combinaciones más significativas, de acuerdo al criterio AIC (Tabla 8).

Tabla 8. Categorización de variables en orden ascendente según significancia.

Nomenclatura	Combinación de variables	AIC
X30	Lig * SV	-277,89
X27	Hem * SV	-277,87
X18	SV * ST	-277,37
X20	ST * RIS	-275,28
X28	Hem * ST	-275,05
X19	SV * RIS	-274,44
X16	Lig ^ 2	-273,74
X2	Hem	-273,72
X15	Hem ^ 2	-273,71
X12	Hem * C/N	-268,91
X4	C/N	-268,90
X13	Lig * C/N	-268,89
X10	Cel * C/N	-268,89
X1	Cel	-267,99
X8	Cel * Hem	-267,99
X9	Cel * Lig	-265,86
X14	Cel ^ 2	-265,73

Fuente: [Autor]

En la Tabla 9, se registra la evaluación de los modelos construidos mediante el algoritmo Stepwise (agregando variable por variable en orden de significancia). De acuerdo a los parámetros estadísticos, los mejores modelos son los que contienen

de 8 a 14 variables, de los cuales aquellos que tienen mejor ajuste dentro del margen de error son los truncados en la variable 12, 13 y 14.

*Tabla 9. Evaluación de los modelos, según método stepwise (AIC).*

No. de Variables	R-cuadrado	R-ajustado	Observaciones
2	0,74845888	0,73111121	No predice valores lógicos
3	0,77023119	0,74561310	No predice valores lógicos
4	0,79257112	0,76184092	No predice valores lógicos
5	0,84759823	0,81829020	Variables poco significativas
6	0,84908693	0,81286780	Variables poco significativas
7	0,85838829	0,81708488	Variables poco significativas
8	0,89888244	0,86371112	No predice valores lógicos
9	0,89967040	0,85862648	No predice valores lógicos
10	0,90240469	0,85593074	No predice valores lógicos
11	0,90266723	0,84913421	No predice valores lógicos
12	0,90953885	0,85240550	Valido
13	0,91359622	0,85119349	Valido
14	0,91709453	0,84881945	Valido
15	0,91714861	0,83947543	Variables poco significativas
16	0,92880963	0,85287323	No predice valores lógicos

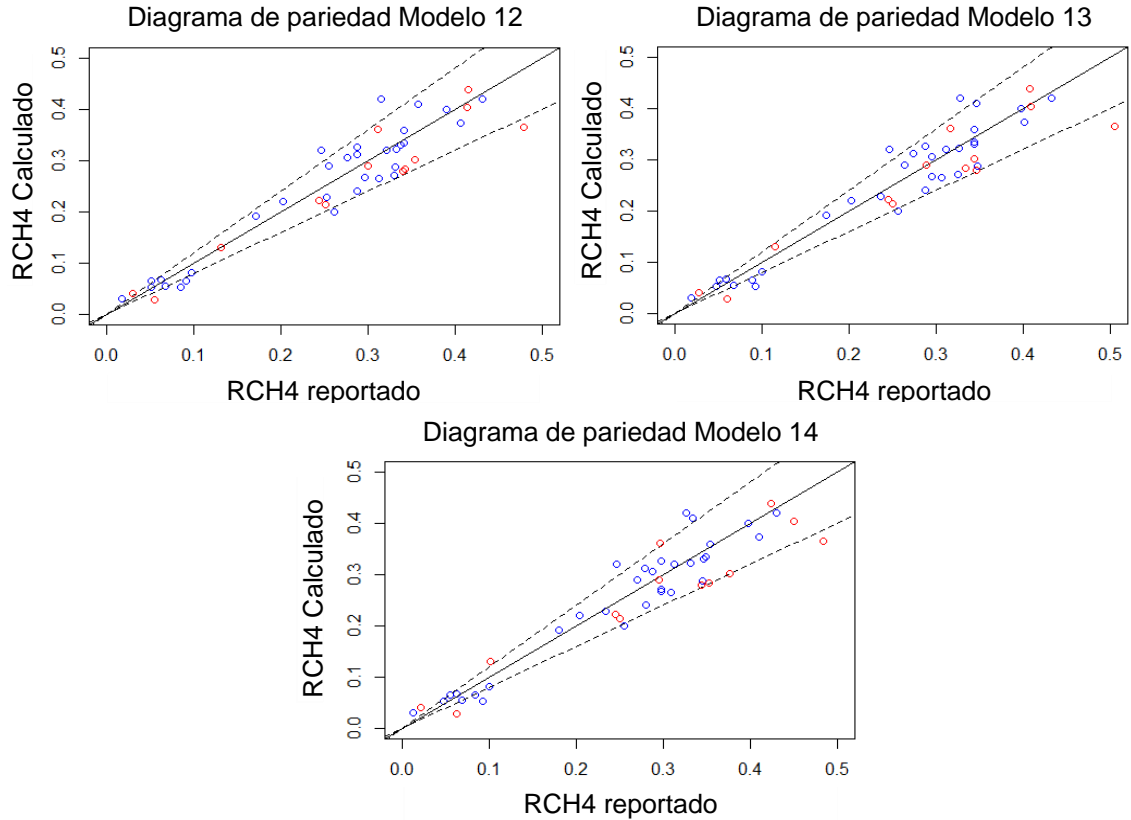
Fuente: [Autor]

### **2.3 Validación del modelo**

Para los 3 mejores modelos se presentan sus respectivos diagramas de pariedad (Figura 6); los puntos azules corresponden al conjunto de datos DCM y los rojos al conjunto DPM.

Finalmente el modelo seleccionado tiene 12 variables, ya que además de ser el de mejor  $R^2$ -ajustado, sus variables son altamente significativas. El modelo de 13 variables aleja algunos puntos del margen de error permitido. El modelo con 14 variables tiene un buen ajuste, sin embargo disminuye la significancia de las variables y el  $R^2$ -ajustado.

Figura 6. Diagramas de pariedad para los modelos truncados en la variable número 12, 13 y 14 respectivamente mediante el método Stepwise.



Fuente: [Autor]

## 2.4 Modelo final para la predicción del rendimiento de metano

A continuación se presenta la ecuación mediante la cual es posible predecir el RCH4 que se obtendrá a partir de una biomasa al aplicar DA batch a temperaturas mesofílicas, donde las constantes se registran en la Tabla 10. Este modelo tiene un  $R^2$  de 0,9095, un  $R^2$  – ajustado de 0,8524 y un p –valor de 1,962e-07.

$$RCH4 = k_0 + Cel * (k_1Cel + k_2Lig + k_3Hem + k_5C/N + k_4) + C/N * (k_7 + k_6Lig + k_8Hem) + Hem * (k_9Hem + k_{10}) + k_{11}Lig^2 + k_{12}SV * RIS$$

Tabla 10. Constantes para el modelo definitivo.

<b>Coficiente</b>	<b>Valor</b>
$k_0$	-7,789496e+02
$k_1$	-8,600956e-01
$k_2$	-7,822496e-01
$k_3$	-9,378314e-01
$k_4$	9,380062e+01
$k_5$	8,260653e-01
$k_6$	8,261205e-01
$k_7$	-8,261179e+01
$k_8$	8,261080e-01
$k_9$	-7,789521e-02
$k_{10}$	1,557993e+01
$k_{11}$	7,790081e-02
$k_{12}$	5,150244e-04

Fuente: [Autor]

## 2.5 Análisis del comportamiento del modelo

Para probar la capacidad de predicción del modelo elegido, se realizó un proceso de verificación en el cual se evalúa el RCH4 respecto a una sola variable, mientras los demás se mantienen fijos (Anexo F). Estas variaciones se hicieron manteniendo cada variable dentro de su valor máximo y mínimo según la Tabla 5.

El RCH4 es muy sensible a la composición lignocelulósica por lo que su análisis debe hacerse teniendo en cuenta el contenido de Lig, Cel y Hem; esto debido a que la suma de estas tres variables es igual a 100%. Una buena cantidad de Cel es favorable siempre y cuando los contenidos de Hem y especialmente de Lig sean bajos. La combinación óptima de estas variables es cuando el porcentaje más alto lo tiene la Cel, el intermedio la Hem y por último el contenido más bajo lo presente la Lig que es quien se encarga de impedir la degradación.

En cuanto a la relación C/N se observa que su aumento es desfavorable para el RCH4, debido a que esto representa un alto contenido de carbono que no se

convierte completamente en metano. Por otro lado es importante tener en cuenta el rango ideal de C/N (20-30/1), debido a que en este aspecto la variable de salida se está comportando inversamente respecto a la relación C/N, por lo que a altos contenidos de nitrógeno, se están obteniendo rendimiento satisfactorios que no son realmente confiables.

Es de esperarse que a mayor cantidad de SV se obtengan mayores RCH<sub>4</sub>. También se observa que el RCH<sub>4</sub> es poco sensible respecto a esta variable debido a la baja significancia que tienen los SV en el modelo según el análisis estadístico. Pese a lo anterior, es importante aclarar que los SV son de gran importancia en el proceso de DA y que dentro de estos se encuentran la Cel, Hem y la Lig; por lo tanto su importancia si se está teniendo en cuenta dentro del modelo final.

## ***2.6 Identificación de los residuos agrícolas de Santander idóneos para aplicar digestión anaerobia***

Los residuos que presentan un mayor RCH<sub>4</sub> son vástago de plátano, cáscara de papa y bagazo de fique con 0,42 m<sup>3</sup>/kg SV y 0,36 m<sup>3</sup>/kg SV y 0,28 m<sup>3</sup>/kg SV respectivamente Tabla 11. Otros productos con RCH<sub>4</sub> superiores a 0,2 m<sup>3</sup>/kg SV son: descarte de tomate, papa y cebolla junca; hojas de cebolla junca, cáscaras de mandarina y naranja, hojas y tallos de papa y mora, poda de naranja y rastrojo de tomate.

Los datos de las biomásas residuales se filtraron, para que estuvieran dentro de los rangos indicados en la Tabla 5. Las biomásas sombreadas en la Tabla 11 son aquellas que aunque no están dentro de los rangos, si están muy cercanos a los valores mínimo y máximo.

Tabla 11. Evaluación del rendimiento de metano a partir de la caracterización de algunas biomasas residuales de Santander.

ID	Cultivo	Residuo	Cel	Hem	Lig	C,N	SV	RIS	RCH4
9	Banano	Rechazo	36,79	22,48	40,73	49,42	73,61	1,50	0,12
10	Café	Pulpa	36,66	30,38	32,96	48,87	85,72	1,50	0,17
12	Café	Tallos	40,48	24,21	35,31	76,09	89,35	1,50	0,07
13	Caña_Azúcar	Cogollo	40,92	24,71	34,37	54,41	81,49	1,50	0,12
19	Palma	Fibra	15,36	35,04	49,60	50,42	85,85	1,50	0,13
20	Palma	Raquis	52,29	18,57	29,14	55,26	82,56	1,50	0,09
23	Plátano	Vástago	37,52	57,46	5,02	23,97	41,07	1,50	0,42
24	Cacao	Hoja	24,85	34,69	40,46	37,63	77,03	1,50	0,17
25	Cacao	Tallos	33,49	40,68	25,82	77,94	86,44	1,50	0,17
27	Cacao	Cacota	34,50	24,31	41,19	78,04	84,43	1,50	0,06
28	Fique	Bagazo	52,43	27,80	19,76	26,95	95,42	1,50	0,28
29	Limón	Poda	44,53	28,67	26,80	50,90	82,72	1,50	0,18
30	Limón	Cáscara	39,21	25,17	35,62	49,72	81,94	1,50	0,13
31	Mandarina	Poda	36,32	33,98	29,70	55,93	84,48	1,50	0,18
32	Mandarina	Cáscara	46,76	26,71	26,53	40,52	81,45	1,50	0,20
33	Naranja	Poda	36,71	32,44	30,85	40,51	83,09	1,50	0,20
34	Naranja	Cáscara_Int	47,71	26,31	25,98	52,67	89,00	1,50	0,17
35	Naranja	Cáscara_Ext	49,00	29,00	22,00	47,23	81,86	1,50	0,21
36	Tomate	Rastrojo	50,86	22,71	26,44	23,42	78,24	1,50	0,22
37	Tomate	Descarte	47,46	27,99	24,55	22,66	92,77	1,50	0,27
38	Papa	Cáscara	47,69	36,48	15,83	29,94	87,33	1,50	0,36
39	Papa	Descarte	43,09	31,72	25,18	36,20	84,36	1,50	0,24
40	Papa	Hojas_Tallos	50,64	25,36	24,00	32,18	82,54	1,50	0,22
43	Yuca	Hoja	28,23	20,15	51,61	14,65	85,40	1,50	0,14
45	Mora	Corona	29,53	29,48	40,99	20,84	86,99	1,50	0,19
46	Mora	Descarte	26,37	28,43	45,20	34,72	79,32	1,50	0,15
47	Mora	Hojas_Tallos	40,64	29,30	30,06	42,02	91,04	1,50	0,20
50	Papaya	Hojas_Tallos	61,96	20,28	17,76	39,46	80,61	1,50	0,19
51	Cebolla_Junca	Hoja	66,23	14,82	18,96	15,66	92,16	1,50	0,23
52	Cebolla_Junca	Descarte	67,08	15,26	17,65	19,04	90,68	1,50	0,23

\*Las variables resaltadas aunque se encuentra fuera del rango, están muy cercanas de los límites.

Fuente: [Autor]

### 3. CONCLUSIONES

Se construyó un modelo matemático para digestión anaerobia batch a temperaturas mesofílicas, con el que se puede predecir del rendimiento de metano a partir de las características fisicoquímicas de las biomásas y la relación inóculo sustrato. Este modelo predice aceptablemente el RCH<sub>4</sub>, a partir de lignina, celulosa, hemicelulosa, C/N, SV y RIS. Este modelo permite tener una aproximación de resultados reales, para tomar la decisión acerca de la viabilidad de aplicar DA a determinadas biomásas.

Las variables más relevantes en la construcción del modelo y que afectan fuertemente el rendimiento de metano son C/N, lignina y celulosa, consiguiendo los mejores rendimientos a porcentajes bajos de lignina y altos de celulosa y rangos de C/N entre 20 y 32.

Los residuos agrícolas de Santander, que presentarían los mejores rendimientos al aplicar digestión anaerobia son vástago de plátano, cáscara de papa y bagazo de fique con 0,42 m<sup>3</sup>/kg SV y 0,36 m<sup>3</sup>/kg SV y 0,28 m<sup>3</sup>/kg SV respectivamente.

#### **4. RECOMENDACIONES**

Para poder aplicar el modelo, la biomasa debe tener características fisicoquímicas dentro de los rangos de la Tabla 5. La temperatura debe mantenerse entre 35 - 40 °C, y el pH entre 6,5 – 8. Sin embargo esto no significa que las biomasas no sean aptas para transformación mediante digestión anaerobia, puesto que si se realiza un pretratamiento con el fin de ajustar las características fisicoquímicas de la biomasa a los rangos óptimos para digestión, el proceso se puede llevar a cabo normalmente.

Otra forma de optimizar el rendimiento de metano es utilizando co-digestión (mezclar la biomasa con otra de diferentes propiedades fisicoquímicas). La construcción del modelo se realizó para experimentos con un solo tipo de biomasa, es posible que al introducir la caracterización de la mezcla se obtenga un rendimiento de metano aproximado.

## BIBLIOGRAFÍA

ABBASSI-GUENDOUZ A, BROCKMANN D, TRABLY E, DUMAS C, DELGENÈS J.P, STEYER J.P, ESCUDIÉ R, *Total solids content drives high solid anaerobic digestion via mass transfer limitation*, Bioresource Technology, Volume 111, May 2012, Pages 55-61, ISSN 0960-8524.

AMEZQUITA N, *Obtención de etanol por la fermentación alcohólica del hidrolizado enzimático del bagazo de caña de azúcar* (Tesis Pregrado) Bucaramanga, Colombia: Universidad Industrial de Santander; 2007.

AMON T, AMON B, KRYVORUCHKO V, ZOLLITSCH W, MAYER K, GRUBER L, *Biogas production from maize and dairy cattle manure—Influence of biomass composition on the methane yield*, Agriculture, Ecosystems & Environment, Volume 118, Issues 1–4, January 2007, Pages 173-182, ISSN 0167-8809.

APPELS L, LAUWERS J, DEGRÈVE J, HELSEN L, LIEVENS B, WILLEMS K, IMPE J.V, DEWIL R, *Anaerobic digestion in global bio-energy production: Potential and research challenges*, Renewable and Sustainable Energy Reviews, Volume 15, Issue 9, December 2011, Pages 4295-4301, ISSN 1364-0321

BREUNIG M, KRIEGEL H, NG R, SANDER J. *LOF: identifying density-based local outliers*. (2000) In ACM Int. Conf. on Management of Data, pages 93-104.

BROWN D, LI Y, *Solid state anaerobic co-digestion of yard waste and food waste for biogas production*, Bioresource Technology, Volume 127, January 2013, Pages 275-280, ISSN 0960-8524.

CAMARA DE COMERCIO DE BUCARAMANGA, *Boletín de actualidad económica, sector agrícola Santander 2011*, realizado por: Observatorio de competitividad y publicado en septiembre de 2012 - Número 101.

CAÑETTE I. *Estimación de densidades basada en núcleos: algunos elementos básicos*. (2002) Seminario de Reconocimiento de patrones. Seminario de probabilidad y estadística.

CASTRO L. *Proceso integral para la digestión anaerobia del bagazo de fique* (Tesis Doctorado) Bucaramanga, Colombia: Universidad Industrial de Santander; 2012.

CEER (Centro de estudios económicos regionales) – Cartagena. Banco de la República. Documentos de trabajo sobre Economía Regional, “¿Por qué le va bien a la economía de Santander?” por Laura Cepeda Emiliani; 2010

CER (Centro de energías renovables), ministerio de energía, gobierno de Chile, Disponible en: <http://cer.gob.cl/tecnologias/biomasa/bioquimica/digestion-anaerobica/> Acceso el 17 de octubre de 2013.

CIOABLA A.E, IOANA I, DUMITREL G, POPESCU F, *Comparative study on factors affecting anaerobic digestion of agricultural vegetal residues*; 2012.

DINUCCIO E, BALSARI P, GIOELLI F, MENARDO S, *Evaluation of the biogas productivity potential of some Italian agro-industrial biomasses*, Bioresource Technology, Volume 101, Issue 10, May 2010, Pages 3780-3783, ISSN 0960-8524.

ECN Phyllis 2 (2012). *Database for biomass and waste*. Disponible en <https://www.ecn.nl/phyllis2/> Acceso en agosto de 2013.

ESCALANTE H, ORDUZ J, ZAPATA H, CARDONA M.C, DUARTE M, *Atlas del potencial energético de la biomasa residual en Colombia*, Ediciones UIS, 2010.

GUERRERO J, GAMBOA C. *Diseño conceptual de una planta piloto de combustión para el aprovechamiento energético de la biomasa residual del maíz* (Tesis Pregrado) Bucaramanga, Colombia: Universidad Industrial de Santander; 2009.

HAIR J, ANDERSON R, TATHAM R, BLACK W. *Análisis multivariante de datos*. Quinta edición. Prentice Hall Iberia. Madrid, 1999. ISBN: 84-8322-035-0.

HARRELL F.E JR, DUPONT C, AND MANY OTHERS. *Hmisc: Harrell Miscellaneous. R package version 3.12-2*. Disponible en <http://CRAN.R-project.org/package=Hmisc> Acceso el 12 de septiembre de 2013.

HALADOVÁ D., CUNDR O., PECEN J. *Selection of optimal anaerobic digestion technology for family sized farm use – case study of southwest Madagascar*, 2011.

IDAE (Instituto para la Diversificación y Ahorro de la Energía), “*Biomasa: Digestores anaerobios*”. Madrid, Octubre de 2007.

JAGADABHI P, KAPARAJU P, RINTALA J, *Two-stage anaerobic digestion of tomato, cucumber, common reed and grass silage in leach-bed reactors and upflow anaerobic sludge blanket reactors*, *Bioresource Technology*, Volume 102, Issue 7, April 2011, Pages 4726-4733, ISSN 0960-8524

KALIA V.C, SONAKYA V, RAIZADA N, *Anaerobic digestion of banana stem waste*, *Bioresource Technology*, Volume 73, Issue 2, June 2000, Pages 191-193, ISSN 0960-8524.

KUNA H, PAUTSCH G, REY M, CUBA C, RAMBO A, CABALLERO S, STEINHILBER A, GARCÍA-MARTÍNEZ R, VILLATORO F. *Avances en procedimientos de la explotación de información con algoritmos basados en la densidad para la identificación de outliers en bases de datos*. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. 5 y 6 de Mayo de 2011. Universidad Nacional de Rosario. Rosario. Santa Fe. Argentina. Pág. 248-252. ISBN 978-950-673-892-1.

LEHTOMÄKI A, *Biogas production from energy crops and crop residues*. Jyväskylä studies in biological and environmental science; 1456-9701; 163. 2006. ISBN 951-39-2559-5

LEHTOMÄKI A, VIINIKAINEN T.A., RINTALA J.A., *Screening boreal energy crops and crop residues for methane biofuel production*, Biomass and Bioenergy, Volume 32, Issue 6, June 2008, Pages 541-550, ISSN 0961-9534.

LIANHUA L, DONG L, YONGMING S, LONGLONG M, ZHENHONG Y, XIAOYING K, *Effect of temperature and solid concentration on anaerobic digestion of rice straw in South China*, International Journal of Hydrogen Energy, Volume 35, Issue 13, July 2010, Pages 7261-7266, ISSN 0360-3199

LIEW L.N, SHI J, LI Y, *Methane production from solid-state anaerobic digestion of lignocellulosic biomass*, Biomass and Bioenergy, Volume 46, November 2012, Pages 125-132, ISSN 0961-9534.

MARTÍ N. *Phosphorus Precipitation in Anaerobic Digestion Process*. Boca Raton, Florida, USA; 2006. ISBN: 1-58112- 332-9.

MIÑARRO A. *Estimación no paramétrica de la función de densidad*. 1998 Barcelona, España.

MSHANDETE A, BJÖRNSSON L, KIVAIISI A.K, RUBINDAMAYUGI M.S.T., MATTIASSON B, *Effect of particle size on biogas yield from sisal fibre waste*, Renewable Energy, Volume 31, Issue 14, November 2006, Pages 2385-2392, ISSN 0960-1481

MONLAU F, BARAKAT A, STEYER J.P., CARRERE H, *Comparison of seven types of thermo-chemical pretreatments on the structural features and anaerobic digestion of sunflower stalks*, Bioresource Technology, Volume 120, September 2012, Pages 241-247, ISSN 0960-8524.

MONTES M, *Estudio técnico-económico de la digestión anaerobia conjunta de la fracción orgánica de los residuos sólidos urbanos y lodos de depuradora para la obtención de biogás*. (Tesis Doctorado) Departamento de Ingeniería Civil: Ordenación del Territorio, Urbanismo y Medio Ambiente E.T.S. I. de Caminos, Canales y Puertos; 2008.

NAMSREE P, SUVAJITTANONT W, PUTTANLEK C, UTTAPAP D, RUNGSARDTHONG V, *Anaerobic digestion of pineapple pulp and peel in a plug-flow reactor*, Journal of Environmental Management, Volume 110, 15 November 2012, Pages 40-47, ISSN 0301-4797

ORDUZ J, *Modelo matemático para evaluar el potencial energético de la biomasa residual agrícola en Colombia* (Tesis Pregrado) Bucaramanga, Colombia: Universidad Industrial de Santander; 2011

PANICHNUMSIN P, NOPHARATANA A, AHRING B, CHAIPRASERT P, *Production of methane by co-digestion of cassava pulp with various concentrations of pig manure*, Biomass and Bioenergy, Volume 34, Issue 8, August 2010, Pages 1117-1124, ISSN 0961-9534.

PARAWIRA W, MURTO M, ZVAUYA R, MATTIASSON B, *Anaerobic batch digestion of solid potato waste alone and in combination with sugar beet leaves*, Renewable Energy, Volume 29, Issue 11, September 2004, Pages 1811-1823, ISSN 0960-1481.

ŠULC J, ŠTOJDL J, RICHTER M, POPELKA J, SVOBODA K, SMETANA J, VACEK J, SKOBLJA S, BURYAN P, *Biomass waste gasification – Can be the two stage process suitable for tar reduction and power generation?*, Waste Management, Volume 32, Issue 4, April 2012, Pages 692-700, ISSN 0956-053X.

TARAZONA S, *Optimización del proceso de producción de bio oil a partir de raquis de palma de aceite por pirólisis intermedia* (Tesis Pregrado) Bucaramanga, Colombia: Universidad Industrial de Santander; 2013.

TORGO L. *Data Mining using R: learning with case studies*, (2010) CRC Press ISBN: 9781439810187.

TORGO L. *Data Mining with R, learning with case studies Chapman and Hall/CRC*. Disponible en <http://www.dcc.fc.up.pt/~ltorgo/DataMiningWithR> Acceso el 12 de septiembre de 2013.

TRIOLO J.M, SOMMER S.G., MØLLER H.B., WEISBJERG M.R., JIANG X.Y., *A new algorithm to characterize biodegradability of biomass during anaerobic digestion: Influence of lignin concentration on methane production potential*, Bioresource Technology, Volume 102, Issue 20, October 2011, Pages 9395-9402, ISSN 0960-8524

WARD A.J, HOBBS P.J, HOLLIMAN P.J, JONES D.L, *Optimisation of the anaerobic digestion of agricultural resources*, Bioresource Technology, Volume 99, Issue 17, November 2008,Pages 7928-7940, ISSN 0960-8524.

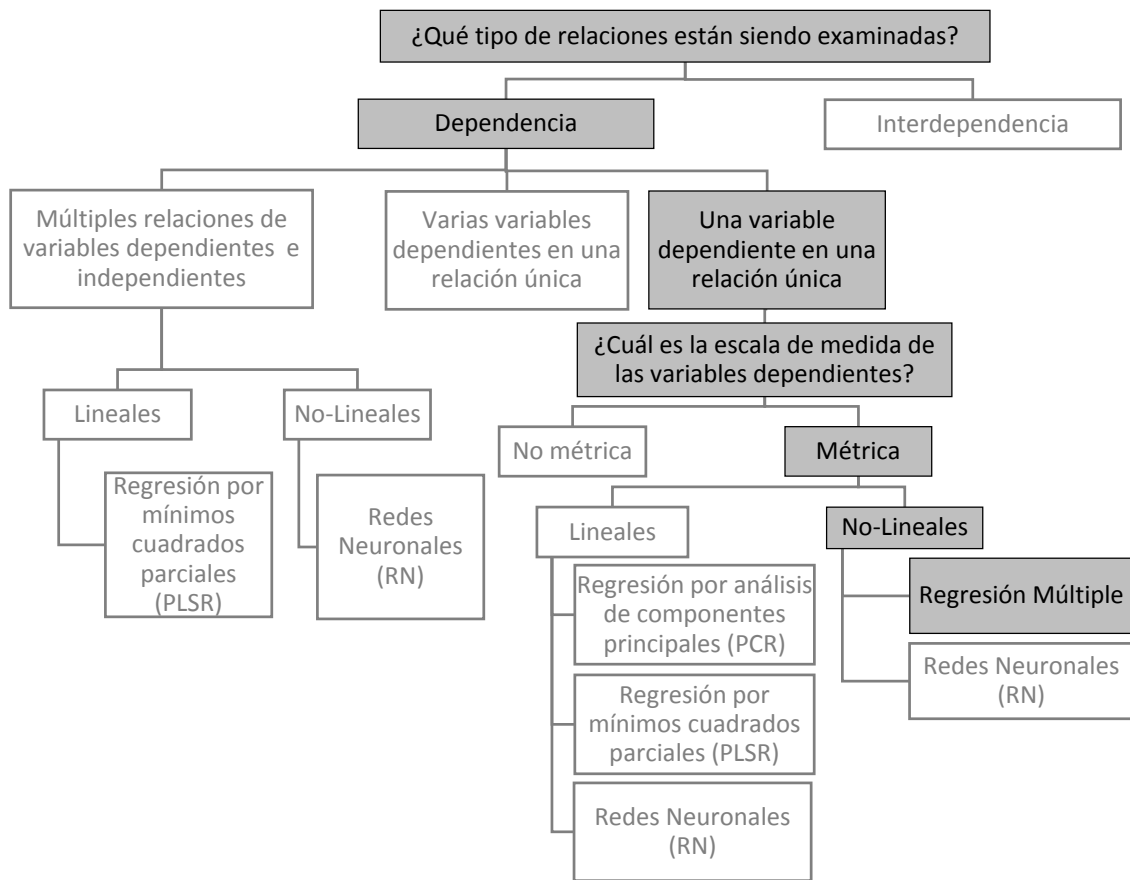
YAO Y, HE M, REN Y, MA L, LUO Y, SHENG H, XIANG Y, ZHANG H, LI Q, AN L, *Anaerobic digestion of poplar processing residues for methane production after alkaline treatment*, Bioresource Technology, Volume 134, April 2013, Pages 347-352, ISSN 0960-8524

## ANEXOS

### Anexo A. Método de selección de la técnica multivariante

La Figura 7 presenta un algoritmo que permite seleccionar la técnica multivariante idónea, de acuerdo al tipo de relación estudiada y la cantidad de variables a utilizar. Las cajas grises, representan la ruta que se seleccionó para el desarrollo de este trabajo.

Figura 7. Técnicas multivariante a emplear de acuerdo a la dependencia y relación de las variables



Fuente: Análisis Multivariante (Hair *et al*, 1999)

## Anexo B. Base de datos digestión anaerobia

Tabla 12. Base de datos para digestión anaerobia.

ID	SUSTRATO	Cel	Hem	Lig	C,N	SV	ST	RIS	RCH4	REFERENCIA
1	Corn Stover	49,56	28,09	22,35	71,00	92,60	97,00	0,5	0,081	Liew <i>et al.</i> , 2011
2	Corn Stover	49,56	28,09	22,35	71,00	92,60	97,00	0,3	0,052	Liew <i>et al.</i> , 2011
3	Wheat Straw	47,78	26,48	25,74	76,33	92,60	97,80	0,5	0,067	Liew <i>et al.</i> , 2011
4	Wheat Straw	47,78	26,48	25,74	76,33	92,60	97,80	0,3	0,064	Liew <i>et al.</i> , 2012
5	Yard waste	35,06	22,94	42,00	70,00	94,60	98,00	0,5	0,041	Liew <i>et al.</i> , 2012
6	Yard waste	35,06	22,94	42,00	70,00	94,60	98,00	0,3	0,031	Liew <i>et al.</i> , 2012
7	Leaves	26,58	23,09	50,33	44,36	89,20	95,90	0,5	0,055	Liew <i>et al.</i> , 2012
8	Leaves	26,58	23,09	50,33	44,36	89,20	95,90	0,3	0,028	Liew <i>et al.</i> , 2012
9	Sunflower stalks	40,24	24,62	35,15	38,97	88,00	94,00	0,5	0,192	Monlau <i>et al.</i> , 2012
10	Perennial grass	47,79	42,75	9,46	34,72	7,72	28,82	1,0	0,271	Triolo <i>et al.</i> , 2011
11	Perennial grass	46,05	48,43	5,52	36,77	5,20	23,50	1,0	0,411	Triolo <i>et al.</i> , 2011
12	Dried wild grass	55,49	35,29	9,22	52,15	88,02	93,41	1,0	0,306	Triolo <i>et al.</i> , 2011
13	Grass	50,34	45,85	3,81	21,88	3,23	18,74	1,0	0,374	Triolo <i>et al.</i> , 2011
14	Grass	53,56	43,86	2,58	21,88	2,89	18,02	1,0	0,421	Triolo <i>et al.</i> , 2011
15	Grass	49,06	47,15	3,79	21,88	2,77	17,53	1,0	0,439	Triolo <i>et al.</i> , 2011
16	Maize	37,20	59,04	3,76	29,03	14,55	38,75	1,0	0,399	Triolo <i>et al.</i> , 2011
17	Maize	40,85	54,62	4,53	16,12	8,56	29,92	1,0	0,405	Triolo <i>et al.</i> , 2011
18	Maize	36,66	60,36	2,97	39,34	7,45	27,89	1,0	0,361	Triolo <i>et al.</i> , 2011
19	Straw	57,96	34,29	7,75	62,07	90,63	93,69	1,0	0,290	Triolo <i>et al.</i> , 2011
20	Tomate	57,34	36,24	6,42	35,68	7,60	10,00	1,0	0,320	Jagadabhi <i>et al.</i> , 2011
21	Cucumber	60,26	32,45	7,28	50,15	4,50	6,80	1,0	0,240	Jagadabhi <i>et al.</i> , 2011
22	Common reed	47,75	44,94	7,30	55,09	41,00	44,30	1,0	0,220	Jagadabhi <i>et al.</i> , 2011
23	Grass silage	53,69	40,27	6,04	37,42	39,00	41,00	1,0	0,360	Jagadabhi <i>et al.</i> , 2011
24	Maize Tonale	53,31	37,26	9,43	24,20	18,40	19,40	0,8	0,334	Amon <i>et al.</i> , 2006
25	Maize Tonale	39,78	52,85	7,37	39,60	28,30	29,80	0,8	0,283	Amon <i>et al.</i> , 2006
26	Maize Tonale	38,68	52,96	8,36	45,10	41,80	43,10	0,8	0,280	Amon <i>et al.</i> , 2006
27	Maize PR34G13	49,85	37,46	12,68	24,90	17,20	18,00	0,8	0,366	Amon <i>et al.</i> , 2006
28	Maize PR34G13	40,59	50,86	8,55	33,50	26,90	28,20	0,8	0,302	Amon <i>et al.</i> , 2006
29	Maize PR34G13	40,42	50,21	9,37	46,20	41,40	43,00	0,8	0,268	Amon <i>et al.</i> , 2006
30	Maize Tixxus	38,46	54,62	6,92	44,10	28,60	30,20	0,8	0,322	Amon <i>et al.</i> , 2006
31	Maize LZM 600	52,61	36,81	10,58	43,50	17,40	18,10	0,8	0,313	Amon <i>et al.</i> , 2006
32	Maize LZM 600	40,86	50,07	9,06	42,10	27,50	29,00	0,8	0,326	Amon <i>et al.</i> , 2006
33	Maize LZM 600	33,39	59,17	7,44	42,20	46,70	48,00	0,8	0,287	Amon <i>et al.</i> , 2006
34	Potato	17,21	72,77	10,02	35,00	18,05	19,00	2,4	0,200	Parawira <i>et al.</i> , 2004
35	Potato	17,21	72,77	10,02	35,00	18,05	19,00	1,4	0,214	Parawira <i>et al.</i> , 2004

ID	SUSTRATO	Cel	Hem	Lig	C,N	SV	ST	RIS	RCH4	REFERENCIA
36	Potato	17,21	72,77	10,02	35,00	18,05	19,00	0,9	0,320	Parawira <i>et al.</i> , 2004
37	Potato	17,21	72,77	10,02	35,00	18,05	19,00	0,6	0,223	Parawira <i>et al.</i> , 2004
38	Reed canary grass	54,14	36,93	8,93	28,00	28,40	29,50	0,8	0,330	Lehtomäki <i>et al.</i> , 2007
39	Reed canary grass	51,89	35,40	12,71	25,00	38,70	39,90	0,8	0,420	Lehtomäki <i>et al.</i> , 2007
40	Tomato skin and seeds	73,50	18,81	7,70	37,97	31,30	32,00	2,0	0,218	Dinuccio <i>et al.</i> , 2009
41	Barley Straw	54,17	34,72	11,11	70,57	85,34	90,50	2,0	0,229	Dinuccio <i>et al.</i> , 2009
42	Rice Straw	25,02	64,34	10,63	43,69	81,52	88,70	2,0	0,195	Dinuccio <i>et al.</i> , 2009
43	Grape stalks	37,48	25,36	37,16	58,21	28,58	31,10	2,0	0,098	Dinuccio <i>et al.</i> , 2009
44	Maize drying up residues	27,60	67,22	5,19	35,89	79,76	81,80	2,0	0,317	Dinuccio <i>et al.</i> , 2009
45	Grape marcs	25,66	34,77	39,57	28,51	55,69	61,40	2,0	0,116	Dinuccio <i>et al.</i> , 2009
46	Yard waste	42,63	17,02	40,35	55,30	91,70	94,30	1,0	0,052	Brown & Li, 2012
47	Cassava Pulp	51,40	36,14	12,46	23,00	5,20	6,50	1,5	0,266	Panichnumsin <i>et al.</i> , 2010
48	Cassava Pulp	51,40	36,14	12,46	27,00	5,20	6,20	1,0	0,290	Panichnumsin <i>et al.</i> , 2010
49	Cassava Pulp	51,40	36,14	12,46	59,00	5,20	5,70	0,3	0,263	Panichnumsin <i>et al.</i> , 2010
50	Poplar processing residues	49,23	26,42	24,36	25,00	77,30	84,80	2,2	0,130	Yao <i>et al.</i> , 2012
51	Poplar processing residues	49,23	26,42	24,36	25,00	77,30	84,80	1,2	0,065	Yao <i>et al.</i> , 2012

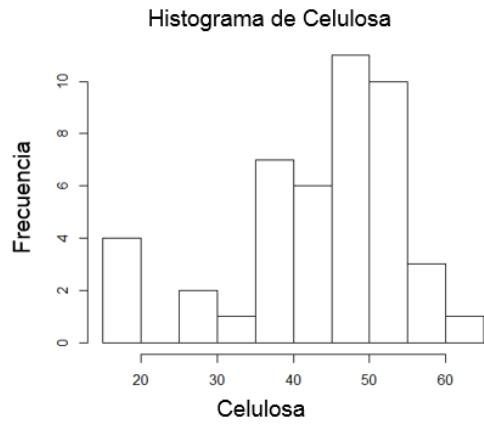
Fuente: [Autor]

\*Los datos en números rojos fueron tomados de (ECN, 2012), de acuerdo a las demás propiedades de las biomásas.

\*Los datos con color de celda azul corresponden a los datos imputados.

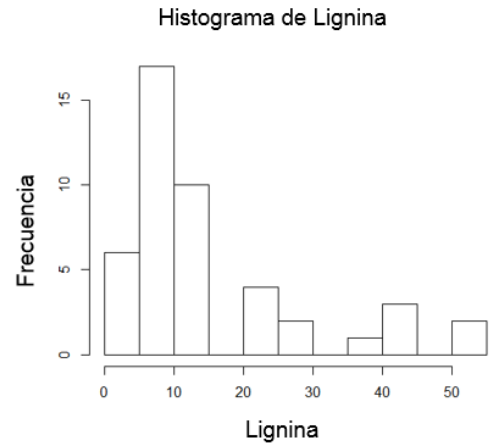
## Anexo C. Histogramas

Figura 8. Histograma Celulosa



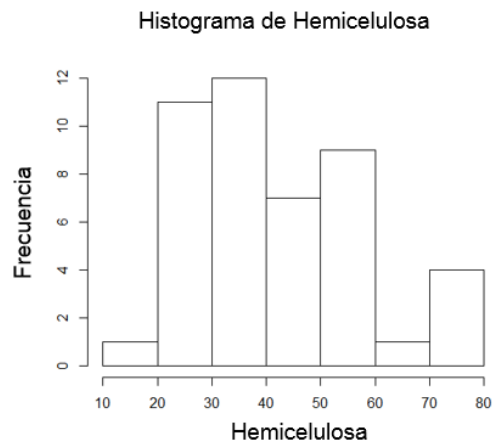
Fuente: [Autor]

Figura 10. Histograma Lignina.



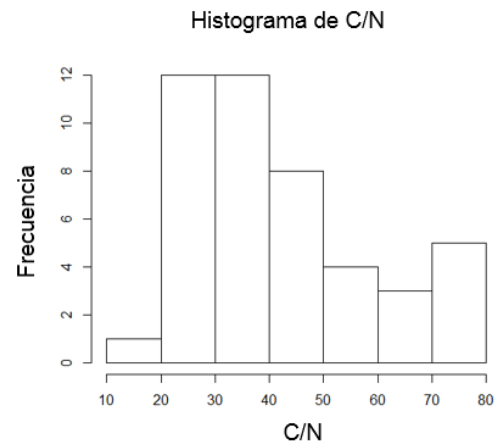
Fuente: [Autor]

Figura 9. Histograma Hemicelulosa.



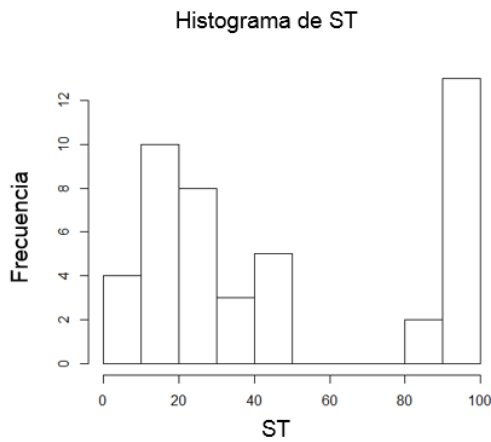
Fuente: [Autor]

Figura 11. Histograma C/N.



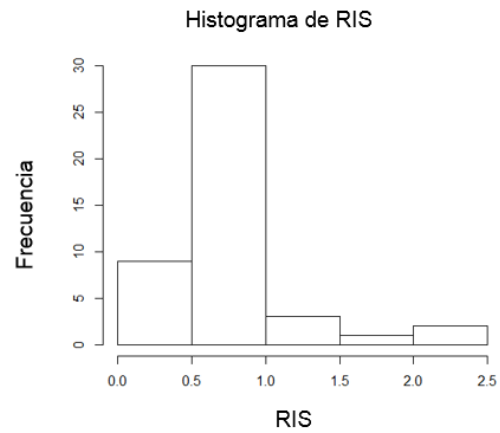
Fuente: [Autor]

Figura 12. Histograma ST.



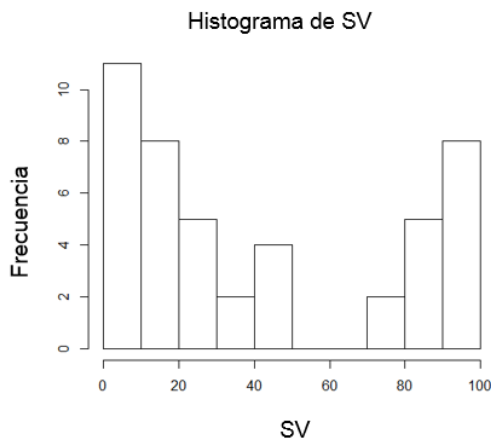
Fuente: [Autor]

Figura 14. Histograma RIS.



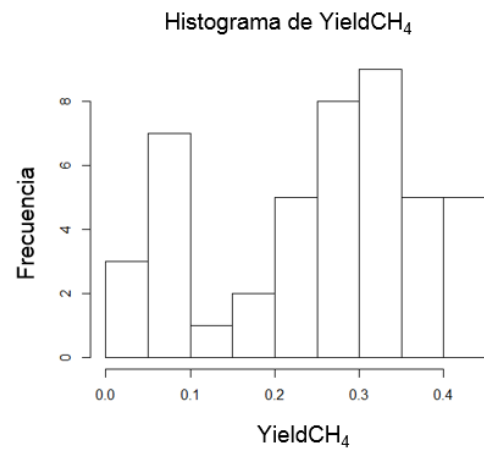
Fuente: [Autor]

Figura 13. Histograma SV.



Fuente: [Autor]

Figura 15. Histograma RCH4



Fuente: [Autor]

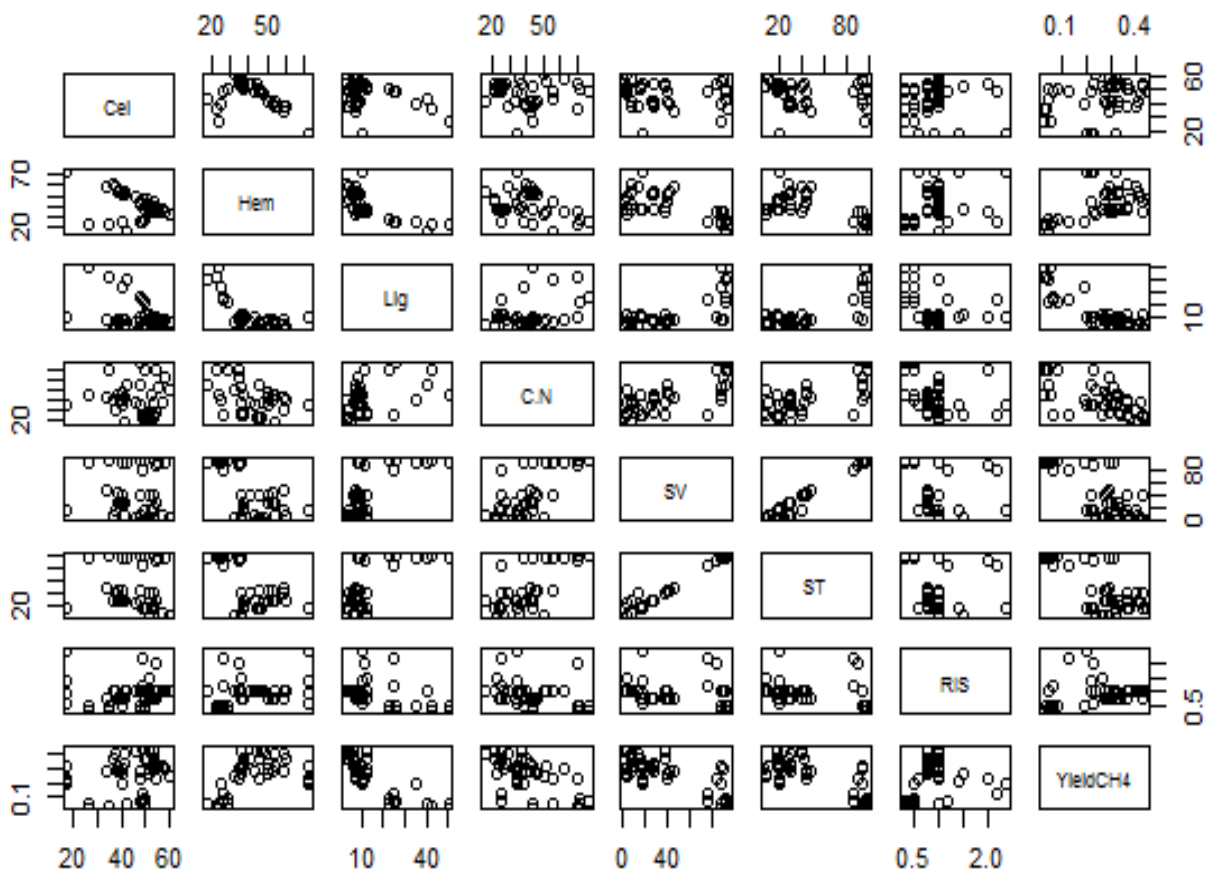
A partir de estos histogramas se observa marcadamente que para ST y SV se tienen un gran conjunto de datos entre 0 y 50 % y otro conjunto pequeño cercano a 100 %; este contenido en los sólidos totales muestra que la mayoría de los experimentos registrados en la base de datos, corresponden a procesos en entorno acuoso y unos pocos de biomazas relativamente secas que en la mayoría de los casos presentan bajos rendimientos de metano y es preferible someterlas a

pretratamientos previos. En cuanto a la relación Carbono/Nitrógeno aproximadamente el 50% de los datos se encuentra alrededor del rango recomendado para el tipo de residuos estudiado y el otro 50% presenta altos contenidos de carbono (lo que sugiere el uso de una co-digestión para ajustar la relación y así mejorar el rendimiento de metano). Por otro lado se observa una gran cantidad de biomásas con alto contenido en celulosa, que al comparar con la base de datos (Anexo B) coincide con la alta densidad de rendimientos por encima de 0,2 m<sup>3</sup>/kg SV presentes en la Figura 15 y los altos contenidos de lignina con los rendimientos deficientes (inferiores a 0,2 m<sup>3</sup>/kg SV). Finalmente el 80% de las biomásas registradas en la base de datos corresponden a procesos con relaciones inóculo/Sustrato entre 0 y 1, siendo la RIS igual a 1 la relación óptima según el análisis bibliográfico.

#### Anexo D. Diagrama de dispersión y matriz de coeficientes de Pearson

A partir del diagrama de dispersión (Figura 16) se evidencia una relación de tipo lineal entre ST y SV, también se intuye relación inversa entre Hem – Lig y RCH4 – Lig las cuales pueden ser lineales, logarítmicas o exponenciales. Esto se evidencia notablemente en la matriz de coeficientes de Pearson (Tabla 13) donde para coeficientes entre 0,65 y 1 se considera una relación directamente proporcional y entre -0,65 y -1 la relación sería inversamente proporcional. Esta matriz es de gran utilidad en el momento de suponer relaciones y posibles combinaciones de variables al construir el modelo.

Figura 16. Diagrama de dispersión.



Fuente: [Autor]

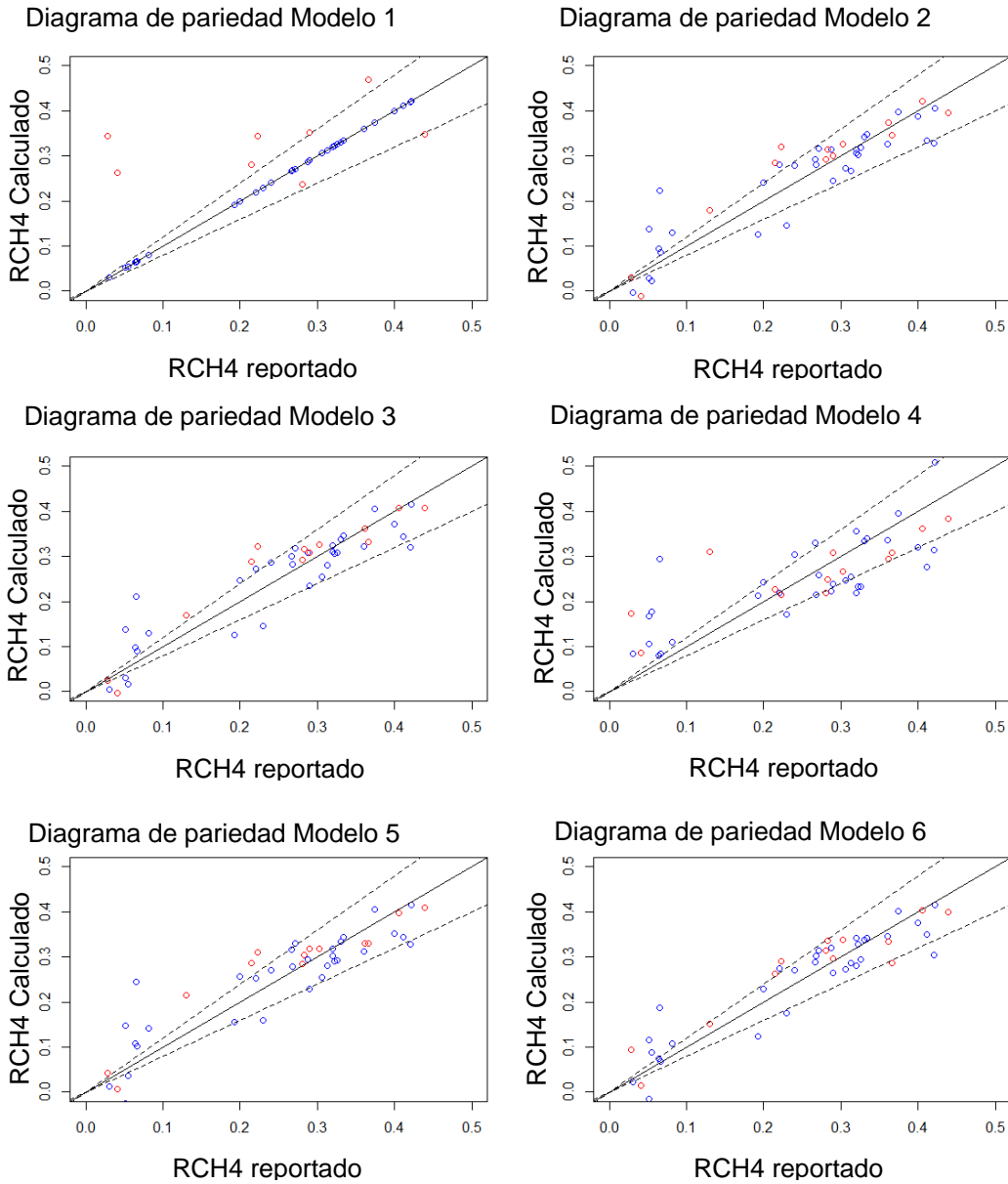
Tabla 13. Matriz de coeficiente de Pearson.

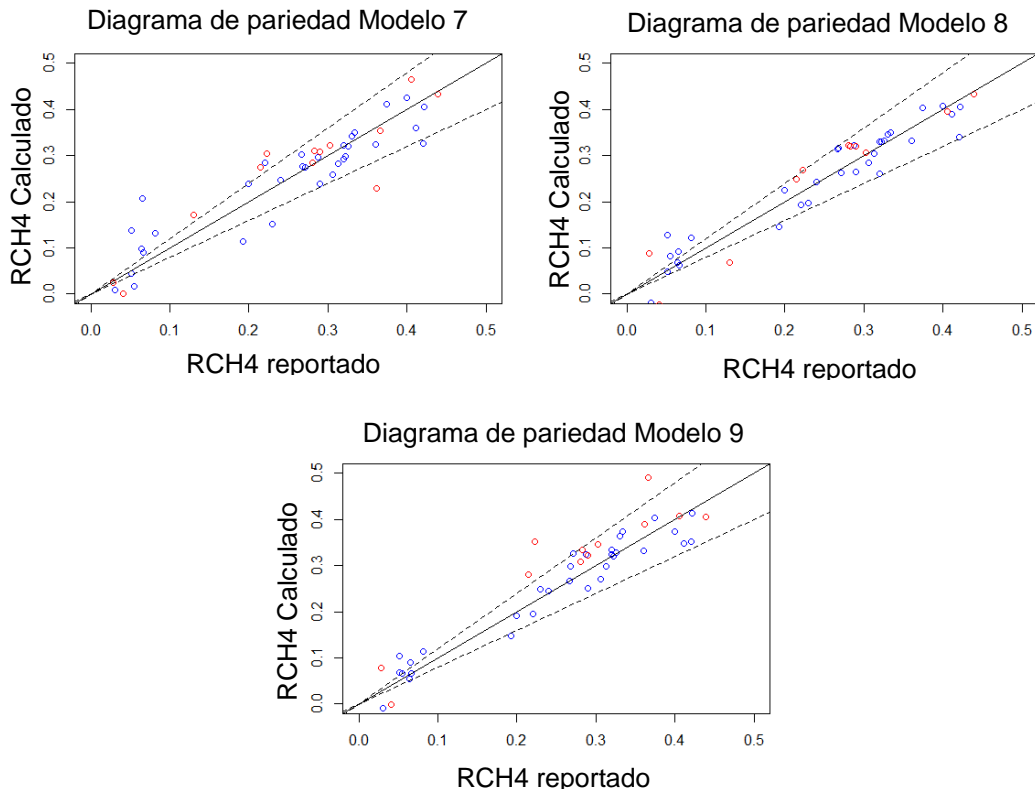
Matriz de coeficientes de Pearson								
	Cel	Hem	Lig	C/N	SV	ST	RIS	RCH4
Cel	1	-0,55	-0,25	0,03	0	0,02	0,02	0,24
Hem	-0,55	1	-0,67	-0,4	-0,66	-0,66	0,3	0,55
Lig	-0,25	-0,67	1	0,44	0,76	0,74	-0,37	-0,84
C/N	0,03	-0,4	0,44	1	0,73	0,7	-0,35	-0,65
SV	0	-0,66	0,76	0,73	1	0,98	-0,28	-0,79
ST	0,02	-0,66	0,74	0,7	0,98	1	-0,28	-0,76
RIS	0,02	0,3	-0,37	-0,35	-0,28	-0,28	1	0,23
RCH4	0,24	0,55	-0,84	-0,65	-0,79	-0,76	0,23	1

Fuente: [Autor]

**Anexo E. Diagramas de pariedad para los modelos construidos en la primera aproximación**

*Figura 17. Diagramas de pariedad de los modelos construidos en la primera aproximación*





\* Los puntos azules corresponden al conjunto de datos DCM y los rojos al conjunto DPM.

Fuente: [Autor]

El modelo 1 se descarta inmediatamente porque en su diagrama de pariedad está sobreajustado, por lo tanto es incapaz de predecir. Los diagramas de los modelos del 2 al 7 presentan demasiados puntos fuera del margen de error; por lo tanto son modelos incapaces de predecir y contienen variables poco significativas. Los diagramas de los modelos 4 y 5 presentan un ajuste deficiente, el cual se evidencia por la alta dispersión de algunos puntos.

Los diagramas de pariedad de los modelos 8 y 9, aparentan un ajuste adecuado, ya que la mayoría de los puntos están dentro del margen de error. Sin embargo, al aplicar cualquiera de estos modelos para predecir el RCH4 de las biomásas de Santander, se obtienen valores ilógicos (RCH4 negativos).

**Anexo F. Variación de datos para prueba**

*Tabla 14. Prueba del modelo, variando cada descriptor.*

Cel	Hem	Lig	C,N	SV	ST	RIS	RCH4
20	30	50	40	45	50	1	0,09
30	30	40	40	45	50	1	0,12
40	30	30	40	45	50	1	0,16
50	30	20	40	45	50	1	0,21
60	30	10	40	45	50	1	0,27
20	60	20	40	45	50	1	0,22
30	50	20	40	45	50	1	0,25
40	40	20	40	45	50	1	0,24
50	30	20	40	45	50	1	0,21
60	20	20	40	45	50	1	0,14
25	65	10	40	45	50	1	0,29
35	55	10	40	45	50	1	0,32
45	45	10	40	45	50	1	0,32
55	35	10	40	45	50	1	0,29
65	25	10	40	45	50	1	0,23
40	55	5	40	45	50	1	0,37
40	45	15	40	45	50	1	0,29
40	35	25	40	45	50	1	0,20
40	25	35	40	45	50	1	0,12
40	20	40	40	45	50	1	0,07
45	40	15	0	45	50	1	0,39
45	40	15	30	45	50	1	0,30
45	40	15	45	45	50	1	0,26
45	40	15	60	45	50	1	0,22
45	40	15	75	45	50	1	0,18
45	40	15	40	5	50	1	0,26
45	40	15	40	25	50	1	0,27
45	40	15	40	45	50	1	0,28
45	40	15	40	65	50	1	0,29
45	40	15	40	85	50	1	0,30
45	40	15	20	5	50	1	0,31
45	40	15	20	25	50	1	0,32
45	40	15	20	45	50	1	0,33
45	40	15	20	65	50	1	0,34
45	40	15	20	85	50	1	0,35
45	40	15	25	5	50	1	0,30
45	40	15	25	25	50	1	0,31
45	40	15	25	45	50	1	0,32
45	40	15	25	65	50	1	0,33
45	40	15	25	85	50	1	0,34

45	40	15	30	5	50	1	0,28
45	40	15	30	25	50	1	0,29
45	40	15	30	45	50	1	0,30
45	40	15	30	65	50	1	0,31
45	40	15	30	85	50	1	0,32
45	40	15	40	45	10	1	0,28
45	40	15	40	45	30	1	0,28
45	40	15	40	45	50	1	0,28
45	40	15	40	45	70	1	0,28
45	40	15	40	45	90	1	0,28
45	40	15	40	45	50	0,25	0,26
45	40	15	40	45	50	0,5	0,26
45	40	15	40	45	50	1	0,28
45	40	15	40	45	50	1,5	0,29
45	40	15	40	45	50	2	0,30

Fuente: [Autor]