

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

**EL MÉTODO DE LAS C MEDIAS DIFUSO MEJORADO CON
APLICACIONES EN LA METEOROLOGÍA**

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DE
INGENIERO MATEMÁTICO**

**CRISTIAN FERNANDO TERÁN LÓPEZ
ahoraeselmomento84@hotmail.com**

**Director: Doctor LUIS HORNA
luis.horna.epn.edu.ec**

Quito, Febrero de 2011.

DECLARACIÓN

Yo CRISTIAN FERNANDO TERÁN LÓPEZ, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

CRISTIAN FERNANDO TERÁN LÓPEZ

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Cristian Fernando Terán López, bajo mi supervisión.

Doctor Luis Horna

DIRECTOR

AGRADECIMIENTOS

A La Escuela Politécnica Nacional por permitirme adentrarme en ese maravilloso mundo de la ciencia y la investigación.

A todos mis profesores de la carrera de Ingeniería Matemática que día a día aportan al desarrollo científico del país.

A mi director de Tesis al Doctor Luis Horna por ser un educador ejemplar que se identifica permanentemente con los estudiantes para satisfacer sus inquietudes.

A todos mis compañeros y amigos que pasamos difíciles momentos para alcanzar nuestras metas.

DEDICATORIA

A mis padres CARLOS y CECILIA por ser mis grandes amigos que siempre me han apoyado a lo largo de mi vida brindándome mucho amor y comprensión.

A mis grandes amores Kathy y Gabrielito por ser mis compañeros que Dios puso en mi camino.

A mis abuelitos, Jaime Enrique, Graciela y Teresa por ser tan buenos y cariñosos.

A toda mi gran familia que siempre ha estado conmigo en las buenas y en las malas.

Cristian

ÍNDICE DE CONTENIDOS

<i>ÍNDICE DE GRÁFICOS</i>	<i>xiii</i>
<i>ÍNDICE DE CUADROS</i>	<i>xv</i>
<i>PRESENTACIÓN</i>	<i>17</i>
<i>INTRODUCCIÓN</i>	<i>18</i>
<i>RESUMEN</i>	<i>21</i>
<i>CAPÍTULO 1: EL PRONÓSTICO DEL TIEMPO</i>	<i>23</i>
1.1 <i>INTRODUCCIÓN</i>	<i>23</i>
1.2 <i>EL COMPORTAMIENTO ATMOSFÉRICO</i>	<i>24</i>
1.2.1 <i>REGIONES DE LA ATMÓSFERA</i>	<i>25</i>
1.2.2 <i>LA METEOROLOGÍA</i>	<i>25</i>
1.3 <i>LAS VARIABLES ATMOSFÉRICAS</i>	<i>26</i>
1.4 <i>OBTENCIÓN DE DATOS METEOROLÓGICOS</i>	<i>28</i>
1.4.1 <i>PROCEDIMINETO PARA LA OBTENCIÓN DE LOS DATOS DE RADIO SONDEOS</i>	<i>30</i>
1.4.2 <i>LIBRO DE CÓDIGOS DE LOS DATOS DE RADIO SONDEOS</i>	<i>30</i>
1.5 <i>APLICACIÓN Y ANÁLISIS ESTADÍSTICO DE VARIABLES ATMOSFÉRICAS</i>	<i>32</i>
1.5.1 <i>JUSTIFICACIÓN</i>	<i>32</i>
1.5.2 <i>MEDIDAS DESCRIPTIVAS DE LOS DATOS METEOROLÓGICOS</i>	<i>32</i>
1.5.3 <i>ANÁLISIS ESTADÍSTICO DE LA PRESIÓN ATMOSFÉRICA (hPa)</i>	<i>33</i>
1.5.4 <i>ANÁLISIS ESTADÍSTICO DE LA TEMPERATURA (°C)</i>	<i>39</i>
1.5.5 <i>ANÁLISIS ESTADÍSTICO DE LA HUMEDAD RELATIVA (%)</i>	<i>43</i>
1.5.6 <i>ANÁLISIS DE LA VELOCIDAD Y DIRECCIÓN DEL VIENTO (M/S)</i>	<i>48</i>
1.6 <i>EL MODELO ATMOSFÉRICO WRF PARA LA VELOCIDAD Y DIRECCIÓN DEL VIENTO</i> ..	<i>51</i>
1.6.1 <i>METODOLOGÍA DE TRABAJO</i>	<i>51</i>
1.6.2 <i>RESULTADOS OBTENIDOS DEL MODELO WRF</i>	<i>52</i>
1.6.3 <i>CONCLUSIONES Y RECOMENDACIONES DE LOS RADIO SONDEOS UTILIZANDO EL MODELO WRF</i>	<i>55</i>
<i>CAPÍTULO 2: CONTRASTES PARAMÉTRICOS DE LAS VARIABLES METEOROLÓGICAS</i>	<i>56</i>
2.1 <i>CONTRASTE PARA UNA MEDIA. INTERVALO DE CONFIANZA</i>	<i>56</i>
2.1.1 <i>PRUEBA t DE UNA MUESTRA</i>	<i>56</i>
2.1.2 <i>ANÁLISIS PARA UNA T DE UNA MUESTRA DE VARIABLES ATMOSFÉRICAS</i>	<i>57</i>
2.1.3 <i>CONCLUSIONES</i>	<i>59</i>

2.2. <i>COMPARACIÓN DE DOS MEDIAS CON MUESTRAS RELACIONADAS</i>	60
2.2.1 <i>COMPARACIÓN DE DOS MEDIAS CON MUESTRAS RELACIONADAS DE VARIABLES ATMOSFÉRICAS</i>	61
2.2.2 <i>CONCLUSIONES</i>	64
2.3 <i>PRUEBA T PARA MUESTRAS INDEPENDIENTES ANOVA DE UN FACTOR DE VARIABLES ATMOSFÉRICAS</i>	65
2.3.1 <i>CONCLUSIONES</i>	67
2.4 <i>TÉCNICAS DE ESTIMACIÓN CURVILÍNEAS DE LAS VARIABLES ATMOSFÉRICAS</i>	68
2.4.1 <i>ANÁLISIS DE TENDENCIA PARA LA VARIABLE PRESIÓN ATMOSFÉRICA</i>	69
2.4.2 <i>ANÁLISIS DE TENDENCIA PARA LA VARIABLE TEMPERATURA</i>	71
2.4.3 <i>ANÁLISIS DE TENDENCIA PARA LA VARIABLE HUMEDAD RELATIVA</i>	75
2.5 <i>MODELOS ARIMA (p,d,q)</i>	76
2.5.1 <i>INTRODUCCIÓN</i>	76
2.5.2 <i>MODELOS ARIMA (p,d,q)</i>	77
2.5.4 <i>APLICACIÓN DEL MODELO ARIMA A LA VARIABLE HUMEDAD RELATIVA</i>	77
2.6 <i>REGRESIÓN LINEAL CON MÚLTIPLES VARIABLES ATMOSFÉRICAS</i>	83
<i>CAPÍTULO 3: MÉTODOS DE REGRESIÓN DIFUSA PARA VARIABLES DEPENDIENTES E INDEPENDIENTES</i>	87
3.1 <i>INTRODUCCIÓN</i>	87
3.2 <i>MODELOS DE REGRESIÓN DIFUSA PARA OBSERVACIONES DEPENDIENTES E INDEPENDIENTES DIFUSAS</i>	88
3.2.1 <i>INTRODUCCIÓN</i>	88
3.2.2 <i>MODELOS DE REGRESIÓN LINEAL DIFUSA</i>	89
3.2.3 <i>EL MODELO CLÁSICO</i>	90
3.2.4 <i>EL MODELO DE DOS PASOS</i>	91
3.3 <i>MODELOS BASADOS EN ÍNDICES DE COMPARACIÓN</i>	92
3.3.1 <i>MODELOS BASADOS EN LOS ÍNDICES DE NECESIDAD</i>	92
3.4 <i>PARÁMETROS Y MEDIDAS DE BONDAD DE AJUSTE DE LA REGRESIÓN DIFUSA</i>	93
3.5 <i>APLICACIÓN DE LOS MODELOS DE REGRESIÓN DIFUSA A LAS VARIABLES METEOROLÓGICAS</i>	95
3.5.1 <i>RESULTADOS OBTENIDOS PARA UNA RADIO SONDA</i>	97
3.5.2 <i>RESULTADOS OBTENIDOS PARA LOS PROMEDIOS DE RADIO SONDAS EN AMBATO</i>	102
<i>CAPÍTULO 4. APLICACIÓN DE TÉCNICAS DIFUSAS A LAS VARIABLES ATMOSFÉRICAS</i>	107
4.1 <i>INTRODUCCIÓN</i>	107

4.2 AGRUPAMIENTO DIFUSO CON FUNCIÓN OBJETIVO.....	108
4.3 TÉCNICA DE AGRUPAMIENTO DIFUSO C-MEDIAS.....	110
4.3.1 ALGORITMO DE LAS C-MEDIAS DIFUSO (FCM)	111
4.4 TÉCNICA DE AGRUPAMIENTO DE LA C- MEDIAS DIFUSO MEJORADO	113
4.4.1. EL MÉTODO DE CONGLOMERADOS SUBTRACTIVO (MSCI)	118
4.4.2 GENERALIZACIÓN DEL MODELO DIFUSO	122
4.4.3 IMPLEMENTACIÓN DEL MÉTODO DE LAS C-MEDIAS DIFUSO MEJORADO A LA METEOROLOGÍA.....	123
4.4.4 CONCLUSIONES.....	129
CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES.....	131
5.1 CONCLUSIONES	131
5.2. RECOMENDACIONES.....	134
BIBLIOGRAFÍA.....	136
ANEXOS A : CONJUNTOS DIFUSOS VERSUS CONJUNTOS REALES.....	137
A.1 DEFINICIÓN DE CONJUNTO DIFUSO Y REALES	137
A.3 PRINCIPIO DE EXTENSION.....	140
A.4 DISTANCIA DIFUSA ENTRE CONJUNTOS DIFUSOS	142
A.6 CONVEXIDAD DE CONJUNTOS DIFUSOS.....	143
A.7 CLASIFICACIÓN DE NUMEROS DIFUSOS.....	144
A.8 OPERACIONES DEFINIDAS ENTRE CONJUNTOS DIFUSOS DEFINIDOS COMO INTERVALOS	146
A.9. OPERACIONES CON NÚMERO DIFUSOS TRIANGULARES	147
A.10 IMPLEMENTACIÓN EN CÓDIGO MATLAB DE LAS REGRESIONES DIFUSAS.....	148
A.11 IMPLEMENTACIÓN EN CÓDIGO MATLAB MÉTODO DE LAS C-MEDIAS DIFUSO MEJORADO	151

ÍNDICE DE GRÁFICOS

Gráfico 1.1 Desarrollo de actividades en el área del PGA	19
Gráfico 1.2 Esquema de interrelación de los procesos investigativos	20
Gráfico. 1.4 Radio sondas VAISALA	29
Gráfica 1.5 Histograma variable presión atmosférica.....	35
Gráfico 1.6 Diagrama de Cajas presión atmosférica	35
Gráfico 1.7 Diagrama de Cajas Temperatura.....	40
Gráfico 1.8 Histograma con curva normal para la variable Humedad Relativa	44
Gráfico 1.9 Diagrama de Caja Humedad Relativa.....	44
Gráfico 1.10 Trayectoria de una radiosonda	49
Gráfico. 1.11 Comportamiento del viento vs. Altura	50
Gráfico 1.12 Perfil vertical de la zona de estudio (Ambato) mostrando los promedios de las velocidades de viento detectada en 15 radio sondeos.	53
Gráfico. 1.13 Salida del modelo WRF Día 1	54
Gráfico. 1.15 Velocidades verticales del viento en la zona de estudio	54
Gráfico 2.1 Diagrama de cajas Presión.....	58
Gráfico 2.2 Histograma de la Temperatura.....	58
Gráfico 2.3 Histograma de la humedad relativa.....	59
Gráfico 2.4 Caja de Diferencias presión.....	62
Gráfico 2.5 Histograma de diferencias temperatura.....	62
Gráfica 2.6 Histograma de diferencias humedad relativa	62
Gráfico 2.8 Comportamiento de la Presión Atmosférica en el tiempo	69
Gráfico 2.9 Comportamiento de la Temperatura en el tiempo	71
Gráfico 2.10 Temperatura vs Altura.....	72
Gráfico 2.11 Temperatura vs Altura.....	73
Gráfico 2.12 Comportamiento de la Presión Atmosférica en el tiempo	75
Gráfico 2.13 Presión Atmosférica vs Altura	75
Gráfico 2.14 Comportamiento de la Humedad Relativa en el tiempo	78
Gráfico 2.15 Metodología Modelos ARIMA.....	79
Gráfico 2.16 ACF y ACF parcial para la variable humedad relativa	80
Gráfico 2.17 ACF y ACF parcial diferenciada para la variable humedad relativa	80

<i>Gráfico 2.18 Humedad Relativa y el Ajuste Obtenido</i>	82
<i>Gráfico 2.19 Análisis en componentes principales</i>	85
<i>Gráfico 3.1 Región de confianza para los lanzamientos de radio sondas</i>	97
<i>Gráfico 3.2 Regresión Difusa Presión Atmosférica (hPa)</i>	98
<i>Gráfico 3.3 Regresión Difusa Temperatura (°C)</i>	100
<i>Gráfico 3.4 Regresión Difusa Humedad Relativa (%)</i>	102
<i>Gráfico 3.5 Regresión Difusa Promedios Presión (hPa)</i>	103
<i>Gráfico 3.6 Regresión Difusa Promedios Presión (hPa)</i>	104
<i>Gráfico 3.7 Regresión Difusa Promedios Humedad Relativa (%)</i>	105
<i>Gráfico 4.1 Métodos de Clasificación</i>	108
<i>Gráfico 4.2. Grupos de división de datos</i>	110
<i>Gráfico 4.3. Datos Antes y Después del Escalamiento</i>	113
<i>Gráfico 4.4. Grados de Membresía de las variables atmosféricas</i>	126
<i>Gráfico 4.5. Regresión múltiple vs Regresión Difusa (C-Medias Mejorado)</i>	128
<i>Gráfico A.1. Lógica clásica y lógica difusa</i>	137
<i>Gráfico.A.2. Funciones de pertenencia de conjuntos difusos en meteorología</i>	139
<i>Gráfico.A.3 Distancia Difusa entre conjuntos difusos</i>	143
<i>Gráfico A.4 Conjunto difuso convexo</i>	144

ÍNDICE DE CUADROS

<i>Cuadro 1.1 Porcentajes de gases en la atmósfera.....</i>	<i>24</i>
<i>Cuadro 1.2 Regiones de la Atmósfera</i>	<i>25</i>
<i>Cuadro 1.3 Datos de radio sondeos proporcionados por el INAMHI</i>	<i>29</i>
<i>Cuadro 1.4 Libro de Códigos de Datos Atmosféricos</i>	<i>32</i>
<i>Cuadro 1.5 Resúmenes de casos de las variables meteorológicas.....</i>	<i>33</i>
<i>Cuadro 1.6 Estadísticas descriptivas para la variable presión atmosférica</i>	<i>34</i>
<i>Cuadro 1.7 Estimadores robustos centrales.....</i>	<i>36</i>
<i>Cuadro 1.8 Análisis de Percentiles Presión Atmosférica.....</i>	<i>37</i>
<i>Cuadro 1.9 Análisis de Percentiles Altura.....</i>	<i>37</i>
<i>Cuadro 1.11 Análisis descriptivo variable Temperatura.....</i>	<i>39</i>
<i>Cuadro 1.12 Temperatura a 20003 metros.....</i>	<i>39</i>
<i>Cuadro 1.15 Estimadores Robustos para la variable temperatura.....</i>	<i>41</i>
<i>Cuadro 1.16 Análisis Descriptivo de la Temperatura</i>	<i>42</i>
<i>Cuadro 1.17 Estadísticas descriptivas para la variable Humedad Relativa</i>	<i>43</i>
<i>Cuadro 1.18 Análisis de Humedad Relativa</i>	<i>46</i>
<i>Cuadro 1.20 Estimadores Robustos para la Humedad Relativa</i>	<i>46</i>
<i>Cuadro 1.19 Estadísticas descriptivas para las radiosondas lanzadas en Ambato</i>	<i>47</i>
<i>Cuadro 1.21 Análisis de tendencia de la velocidad del viento</i>	<i>48</i>
<i>Cuadro 2.1. Datos proporcionados por el INAMHI</i>	<i>56</i>
<i>Cuadro 2.2 Estadísticos para una muestra presión atmosférica</i>	<i>57</i>
<i>Cuadro 2.4 Estadísticos para una muestra (temperatura).....</i>	<i>57</i>
<i>Cuadro 2.6 Estadísticos para una muestra humedad</i>	<i>57</i>
<i>Cuadro 2.3 Prueba para una muestra Presión.....</i>	<i>58</i>
<i>Cuadro 2.5 Prueba para una muestra temperatura.....</i>	<i>58</i>
<i>Cuadro 2.7 Prueba para una muestra humedad relativa.....</i>	<i>58</i>
<i>Cuadro 2.8 Muestra relacionadas presión atmosférica.....</i>	<i>61</i>
<i>Cuadro 2.10 Muestras relacionadas temperatura.....</i>	<i>61</i>
<i>Cuadro 2.12 Muestras relacionadas humedad relativa.....</i>	<i>61</i>

<i>Cuadro 2.11 Resultados del contraste sobre comparación de dos medias con muestras relacionadas.....</i>	<i>63</i>
<i>Cuadro 2.13 Resultados del contraste sobre comparación de dos medias con muestras relacionadas.....</i>	<i>63</i>
<i>Cuadro 2.14 ANOVA de Variables Atmosféricas.....</i>	<i>66</i>
<i>Cuadro 2.15 Prueba de homogeneidad de la varianza.....</i>	<i>66</i>
<i>Cuadro 2.18 Análisis ANOVA y de Coeficientes para la curva logarítmica.....</i>	<i>70</i>
<i>Cuadro 2.19 Análisis ANOVA y de coeficientes para la curva exponencial.....</i>	<i>71</i>
<i>Cuadro 2.20 Resumen del Modelo.....</i>	<i>72</i>
<i>Cuadro 2.21 Análisis ANOVA y de coeficientes para la curva exponencial.....</i>	<i>74</i>
<i>Cuadro 2.22 Análisis ANOVA y de coeficientes para la curva cuadrática.....</i>	<i>74</i>
<i>Cuadro 2.24 Estadístico de Durbin-Watson.....</i>	<i>76</i>
<i>Cuadro 2.25 Resumen de Análisis Modelos ARIMA.....</i>	<i>81</i>
<i>Cuadro 2.26 Salida del Modelo ARIMA(0,0,1).....</i>	<i>82</i>
<i>Cuadro 2.27 Pronóstico obtenido para la Humedad Relativa.....</i>	<i>83</i>
<i>Cuadro 2.28 Variables atmosféricas y sus correlaciones.....</i>	<i>84</i>
<i>Cuadro 2.29 Modelo de Regresión lineal múltiple.....</i>	<i>85</i>
<i>Cuadro 2.30 ANOVA de variables atmosféricas.....</i>	<i>86</i>
<i>Cuadro 2.31 Resumen de modelo con múltiples variables.....</i>	<i>86</i>
<i>Cuadro 4.1 Total de datos para las variables meteorológicas.....</i>	<i>124</i>
<i>Cuadro 2.30 ANOVA obtenido de variables atmosféricas.....</i>	<i>125</i>
<i>Cuadro 4.2 Estimación de los conglomerados para las variables meteorológicas.....</i>	<i>125</i>
<i>Cuadro 4.3. Combinación 1, obtención del modelo difuso.....</i>	<i>127</i>
<i>Cuadro 4.4. Estimaciones de los coeficientes por cada combinación de variables.....</i>	<i>127</i>
<i>Cuadro A.3 Algunos valores del conjunto imagen del conjunto difuso S.....</i>	<i>141</i>

PRESENTACIÓN

Para llegar a los sitios más difíciles y escondidos del Ecuador con una gran variedad de servicios de telecomunicaciones tales como: internet, telefonía, radio, televisión y además, realizar observaciones de monitoreo constante de las actividades sísmicas volcánicas del Ecuador es necesario diseñar y construir una nave espacial que cumpla con estos fines, entre otros más. Sobre todo en nuestro país que su situación geográfica necesita de un monitoreo constante porque estamos en una zona de muchos volcanes y que en cualquier momento podrían ocasionar desastres naturales.

Este proyecto de diseño y construcción de una plataforma de gran altitud (PGA) con fines de investigación se desarrolla en la ciudad de Ambato y resulta muy beneficioso conocer las condiciones meteorológicas que se suscitan en este sitio, he ahí el planteamiento de este tema que durante el desarrollo de esta investigación pretendemos aplicar ciertas técnicas de regresión difusa y en especial el método de las c-medias difuso mejorado a nuestros datos meteorológicos que se enmarca dentro del reconocimiento de patrones y comparar si el ajuste de nuestra curva para futuras predicciones es válida para cualquier lugar sobre todo en nuestro país. Para ello debemos aplicar técnicas estadísticas tradicionales para verificar y comprobar la calidad de los datos así como realizar las diferentes pruebas de hipótesis.

El despliegue de esta investigación se desarrolla en el aeropuerto de la Fuerza Aérea, con 1800 metros de longitud. Una de las ventajas para desarrollarse en este lugar es que en este aeropuerto hay muy pocas operaciones.

A lo largo de este estudio mostraremos si las técnicas de la estadística y las técnicas difusas son convenientes utilizarlas para predecir variables meteorológicas en una determinada altura y en un determinado lugar.

INTRODUCCIÓN

Generalmente, siempre nos preocupamos por las posibles condiciones climáticas y solemos preguntarnos, cuál será el tiempo?, Y nos referimos sobre cuál será la temperatura, la humedad, velocidad del viento y el valor de otra serie de variables meteorológicas que caracterizan el estado atmosférico. Por ende, deberíamos conocer el sistema que caracteriza el estado atmosférico. Nos preguntamos, Qué es la Atmosfera? Según, Santiago Gaztelumendi en su artículo de La predicción numérica del tiempo define: *“La atmósfera es básicamente un fluido turbulento, por lo que responde a las leyes de la mecánica de fluidos y obviamente no podemos aplicar ecuaciones tan sencillas”*. He ahí, el problema de la predicción numérica de las variables atmosféricas para la meteorología que implica conseguir una predicción del tiempo para el sitio del lanzamiento de la plataforma del Proyecto de gran altitud (PGA). Por lo que, es necesario e indispensable la determinación de las condiciones atmosféricas en el sitio para el lanzamiento y la sustentabilidad de la aeronave.

Para problemas de predicción, tanto la inteligencia artificial como la estadística, tienen un completo cuerpo teórico desarrollado. Esto puede llevar a una incomunicación entre ellas. Las opiniones de cuándo es más conveniente la inteligencia artificial o la estadística para resolver un determinado problema son diversas. Para nuestra investigación utilizaremos la rama de la Estadística.

Uno de los objetivos es realizar el estudio y análisis de las condiciones atmosféricas y la implementación de un modelo de regresión estadístico matemático adecuado a condiciones tropicales ecuatoriales que simulan la atmósfera en la zona de estudio del PGA en la ciudad de Ambato, que se encuentra ubicado en medio de los tres volcanes principales del Ecuador, estos son: el Chimborazo, el Cotopaxi y el Tungurahua. Ambato tiene las siguientes coordenadas: latitud $1^{\circ}13'28''$; longitud $78^{\circ}37'11''$ y 2500 metros sobre el nivel del mar (referencia google earth). Ver Gráfico 1.1 a una resolución horizontal de 4 Km y con un alcance vertical de hasta 25 Km.

La utilización de técnicas difusas son apropiadas para la adopción de decisiones en condiciones de incertidumbre donde los sistemas a modelar, como la atmósfera, son tan complejas e imprecisas. Entonces, obtener un modelo matemático utilizando herramientas y técnicas de regresión difusa de variables meteorológicas en el medio troposférico ecuatorial para la zona de estudio ubicada en el marco del Proyecto Plataforma de Gran Altitud PGA es el objetivo principal de esta investigación.

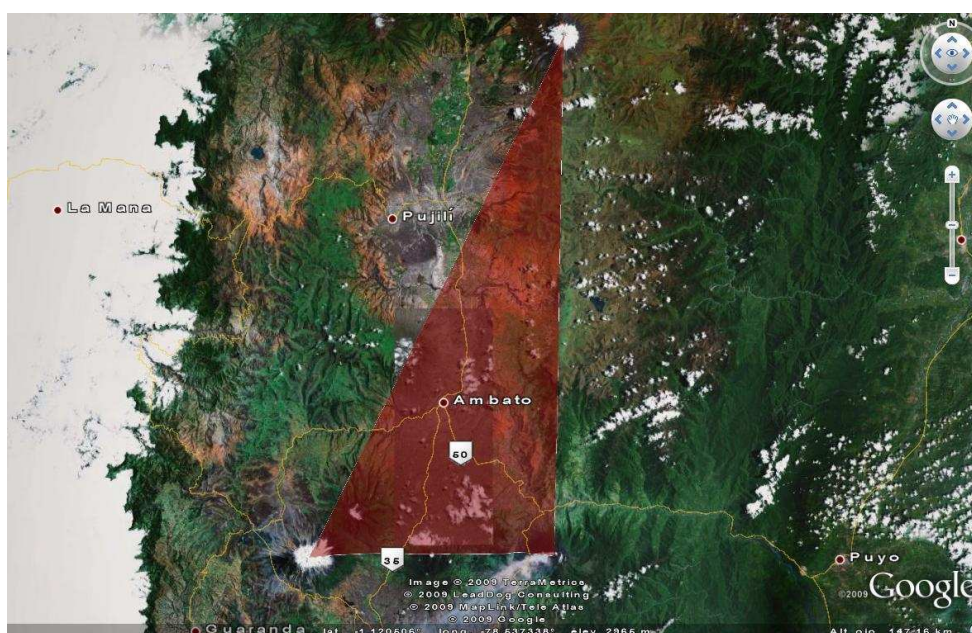


Gráfico 1.1 Desarrollo de actividades en el área del PGA

Hay muchas magnitudes cuantitativas que pueden representarse adecuadamente mediante números difusos: mediciones con márgenes de error; carga eléctrica requerida por una generadora por unidad de tiempo; valor de una acción en un día, una semana o un mes, valor de las monedas frente a otras monedas referenciales, entre otros. Este tipo de datos de mediciones atmosféricas, fundamentalmente cuantificables pero, con una imprecisión e incertidumbre provenientes de diversos componentes del sistema atmosférico justifica la creación de un campo analítico propio para la regresión difusa. Además, la teoría de conjuntos difusos maneja nuevos modelos de medidas, como las medidas de posibilidad y de necesidad, que se traducen en nuevas visiones para afrontar el análisis de regresión.

A continuación presentamos un diagrama (gráfico 1.2) que permite visualizar la interrelación de los procesos a desarrollarse para la obtención del modelo estadístico difuso.

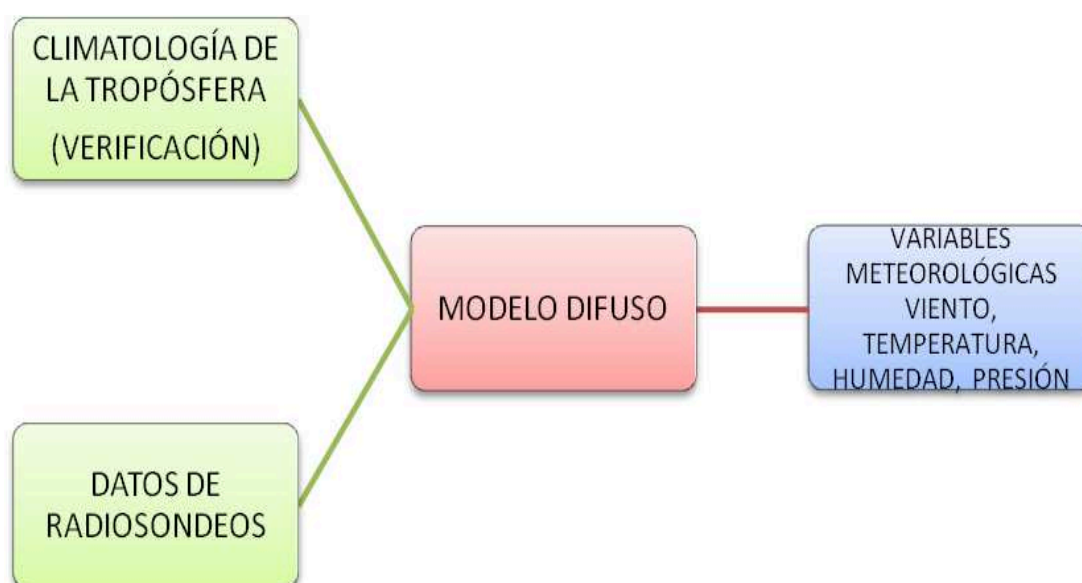


Gráfico 1.2 Esquema de interrelación de los procesos investigativos

Los objetivos que se plantean son:

1. Adecuar técnicas y herramientas difusas que permitan adoptar decisiones en condiciones de incertidumbre de la atmósfera.

2. Verificación de la Climatología generada con los datos de Radio Sondeos obtenidos para este caso, dentro del Proyecto PGA.
3. Generar un modelo regresión lineal múltiple alimentado con los datos de la climatología de la tropósfera y verificado con los datos de los radiosondeos de las variables meteorológicas: dirección y velocidades de viento, temperatura, humedad y presión atmosférica.

RESUMEN

Las técnicas estadísticas y las herramientas difusas, según varias publicaciones, son ampliamente utilizadas en los diferentes campos de la economía, la ingeniería en donde se trabajan con datos cuantitativos, además, estas técnicas y herramientas permiten trabajar en otros campos no cuantitativos como en Ciencias Sociales, Ciencias de la Salud y Biológicas. En esta investigación estamos interesados en explicar el comportamiento cuantitativo de las variables atmosféricas y como una alternativa de solución implementamos ciertas técnicas estadísticas y herramientas difusas para mostrar las condiciones atmosféricas en el área de interés para el PGA. Para ello desarrollamos cinco capítulos que contienen los siguientes temas:

En el capítulo 1, el pronóstico del tiempo damos a conocer sobre las variables de la atmósfera, las formas para obtener los datos de las variables atmosféricas y se realiza un análisis estadístico para visualizar el comportamiento cuantitativo utilizando el análisis de tendencia, la dispersión de los datos, los percentiles y gráficos de diagrama de cajas, muy requerido para el análisis de datos.

En el capítulo 2, el análisis paramétrico de las variables atmosféricas, las pruebas de hipótesis estadísticas para las variables atmosféricas tanto para la media como para la varianza de los datos se analizan en este capítulo.

En el capítulo 3, los métodos de regresión difusa para variables dependientes e independientes se desarrollan para implementar a las variables atmosféricas y mostrar una región de confianza posible para estas variables, verticalmente. Es decir, dependiente de la variable altura.

En el capítulo 4, aplicamos las técnicas difusas a las variables atmosféricas y se utilizan las técnicas de agrupamiento difuso C-Medias y la técnica de agrupamiento difuso mejorado por el método subtractivo para comparar entre estas dos técnicas y la regresión lineal múltiple para obtener el más adecuado.

En el capítulo 5, Conclusiones y Recomendaciones contiene las conclusiones principales que se han extraído durante la implementación y ejecución del proyecto, así como también las investigaciones futuras que se pudieran afrontar tras la realización del mismo.

CAPÍTULO 1: EL PRONÓSTICO DEL TIEMPO

1.1 INTRODUCCIÓN

El objetivo fundamental del proyecto PGA es situar y estabilizar la plataforma de gran altitud a una altura de hasta 20 kilómetros sobre el nivel del suelo en Ambato, por lo que es importante hacer un estudio minucioso del perfil atmosférico en dicha zona es decir, se requiere conocer en forma confiable, el comportamiento de los principales parámetros meteorológicos indispensables para un vuelo de una nave no tripulada y que se quiere que permanezca estacionaria, en el espacio, en un sitio determinado.

Es de suma importancia enfatizar que, debido a que nuestro país se encuentra ubicado en la línea Ecuador, los fenómenos que ocurren son muy especiales, por lo que se amerita un estudio particular del comportamiento y de la estructura de la atmósfera.

Antes de desarrollar la teoría y los modelos estadísticos de los perfiles verticales (con respecto a la altura), es conveniente presentar una idea global de los principales conceptos sobre meteorología, en forma general y tener siempre presente la alta complejidad y comportamiento caótico de la atmósfera en su conjunto.

Debemos saber que los pronósticos meteorológicos han sido objeto de ciertas burlas, en épocas anteriores como también en tiempos modernos. Sin embargo, para predecir el tiempo nuestros antepasados y campesinos se han guiado por cuestiones tradicionales de observación: del cielo, del color de las nubes, de los vuelos de ciertos pájaros. La intensidad de las lluvias, las fases de la luna, la sensación térmica, etc., han transmitido percepciones al ser humano que han sido utilizadas con el fin de fijar, entre otras cosas, las fechas para las siembras. Es decir, no han utilizado instrumentos de precisión para medir los diferentes parámetros meteorológicos, sino tan sólo sensaciones, luego calificadas en rangos de tiempo: excelente, bueno, malo; o, hace mucho frío, frío, calor o mucho

calor, con lo que en forma explícita han utilizado los conceptos elementales de los conjuntos difusos y los números difusos, también conocidos como borrosos o nebulosos (ver anexos).

1.2 EL COMPORTAMIENTO ATMOSFÉRICO

Cuando hablamos de la atmósfera, nos referimos a la capa gaseosa que rodea nuestro planeta compuesta del aire, con partículas sólidas y líquidas en suspensión, o aerosoles y nubes. La composición de la atmósfera y los procesos que en ella se desarrollan tienen gran influencia en la actividad humana y en el comportamiento del medio ambiente en general, afectan los procesos de producción, intercambio y consumo de bienes y servicios, el bienestar y la seguridad de la población, las relaciones sociedad-naturaleza y los procesos en otras esferas del medio natural. De ahí, la necesidad de hacer el seguimiento continuo de la dinámica de la atmósfera, su circulación, las variaciones en su composición y de los fenómenos que en ella ocurren.

Se compone de dos grupos de gases: constantes y variables. Los gases constantes se mantienen en una proporción casi permanente y su participación porcentual en la masa total atmosférica:

GAS	% EN LA ATMÓSFERA
Nitrógeno	78
Oxígeno	20.9
Argón	1.1

Cuadro 1.1 Porcentajes de gases en la atmósfera

Los gases variables son los que cambian en mayor proporción y los más importantes son: el vapor de agua y el dióxido de Carbono (CO₂). Ambos, presentan grandes variaciones estacionales y de largo plazo.

En cuanto a su estructura, el 98% de la masa atmosférica se ubica en una capa delgada cuyo espesor es de unos 30 Km.

1.2.1 REGIONES DE LA ATMÓSFERA

Para efectos prácticos la atmósfera se divide en cuatro regiones según la distribución vertical de la temperatura:

Región	Ubicación	Características
Tropósfera	0 a 16 Km	En esta capa la temperatura disminuye con la altura a una razón de 0,65°/100m. Aquí se originan la mayor parte de los fenómenos determinantes del estado del tiempo. En el Ecuador se extiende hasta unos 17 Km. Se concentra un 80% de toda la masa de la atmósfera
Estratósfera	16 a 50 Km	En la cual la temperatura aumenta hasta alcanzar los 0°C.
Mesósfera	50 a 85 Km	En la cual la temperatura disminuye hasta alcanzar los -95°C
Termósfera	85 a 100Km	Aquí se observa un aumento progresivo de la temperatura.

Cuadro 1.2 Regiones de la Atmósfera¹

El límite superior de la Tropósfera se llama Tropopausa, donde la temperatura deja de disminuir, alcanzando una temperatura próxima a -82 °C. En la Tropósfera (ocurren los fenómenos meteorológicos más significativos) y la Estratósfera, que es donde se centrará nuestro estudio, pues se requiere que la nave sea ubicada a una altura comprendida entre 15 km y 20 km. Además, las radio sondas, que explicaremos más adelante, en promedio, alcanzan esta última altura.

1.2.2 LA METEOROLOGÍA

¹ Eduardo Palomares. Ciencia y Tecnología del Medio Ambiente. Servicio de Publicaciones.

Universidad Politécnica de Valencia. Pág. 63.

La meteorología, como ciencia, es la rama de la física que estudia los fenómenos que se producen en la atmósfera, es decir, entre otros procesos, tiene que ver con el movimiento de la atmósfera, su interacción con los flujos de radiación solar, los procesos termodinámicos que originan la formación de las nubes y la consecuente posterior precipitación, los permanentes intercambios de energía con la superficie terrestre y demás fenómenos eléctricos, reacciones químicas y efectos ópticos que se producen a toda escala, que dependen continuamente del tiempo y del espacio. Claro está, el pronóstico del tiempo ha sido, es y será la actividad más conocida de esta área del conocimiento, tanto como insumo para otras actividades productivas humanas o simplemente para disfrutar de la naturaleza que nos rodea.

Por otro lado, la climatología es una parte de la meteorología que se encarga de estudiar la evolución de las condiciones medias de la atmósfera en una región determinada en intervalos de tiempo relativamente prolongados: diez, treinta y cien años.

Los pronósticos sobre la temperatura en ciudades o regiones se dan en términos de intervalos, es decir, determinando mínimos y máximos o, en otras palabras, rangos de variabilidad, pues precisar un pronóstico es casi imposible, aún en términos probabilísticos.

El constante cambio climático está dejando perplejos a muchos científicos, quienes permanentemente están revisando sus teorías y modelos con el fin de dar una explicación más coherente sobre los fenómenos climáticos que están ocurriendo actualmente sobre nuestro planeta.

1.3 LAS VARIABLES ATMOSFÉRICAS

La complejidad de la modelación de la atmósfera y la obtención de pronósticos confiables en nuestros tiempos ha consolidado y popularizado el excelente “efecto mariposa”, que no es más que dada una pequeña perturbación en la atmósfera, causada por el vuelo de una mariposa, pueda evolucionar y crecer desmedidamente a causa del carácter no lineal y caótico de la atmósfera, y luego

transformarse en una mega perturbación que podría alterar el tiempo de una región muy alejada. Es un formidable símil al “efecto burbuja” en la economía y que hemos sido también testigos en estos últimos años.

Tomando en cuenta, por otro lado, el permanente cambio climático global es imprescindible usar métodos estadísticos y gráficos que permitan describir de una manera más fidedigna los fenómenos atmosféricos y que a su vez concuerde con la percepción humana de éstos. En tal sentido, si se dispusieren de datos adecuados sobre el tiempo y que cubran un número de años razonable, se podría afirmar, con un grado de certeza, si las condiciones son normales o extremas, por lo general, en un período de tiempo de treinta años.

Las propiedades fundamentales de una serie de datos, de los parámetros meteorológicos fundamentales, se expresan en términos de la media, en función del tiempo, y el tamaño de la varianza para determinar la dispersión de los datos. Una tarea fundamental actualmente, consiste en analizar los eventos extremos, pues pueden ser más útiles a la hora de mitigar los efectos catastróficos que causan los fenómenos atmosféricos atípicos, como es el caso del fenómeno de El Niño o La Niña, en nuestro país.

Sin embargo, para los objetivos del proyecto PGA, no se tiene mayor información sobre el perfil atmosférico sobre la zona de Ambato, en función del tiempo, por lo que se recurre a una herramienta que suple, de alguna manera, la cantidad de datos que, además, están sujetos a la incertidumbre e imprecisión como son: el uso de las técnicas estadísticas y las herramientas difusas.

Es importante recalcar que el tiempo y el clima son el resultado de un proceso dinámico y caótico sumamente complejo que requiere mediciones permanentes multidimensionales, lo que su pronóstico ha sido un desafío constante en el medio aficionado y científico. Dentro del contexto del pronóstico del tiempo a escala local, los meteorólogos se han apoyado en datos homogéneos con el fin de obtener más o menos buenos pronósticos para horizontes de tiempo cortos, quedando reservados los modelos dinámicos para los fenómenos del tiempo a mayores horizontes y dominios (regiones) más grandes. Para la simulación de estos modelos es fundamental la utilización de computadoras rápidas y de gran

memoria, pues la actualización de nuevas condiciones iniciales y de frontera más la presencia de grandes bases de datos haría imposible las respectivas corridas secuenciales, en el proceso de análisis y refinamiento del modelo.

Sin embargo, ha habido muy poco interés en utilizar herramientas de la estimación usando los conjuntos difusos y las técnicas estadísticas, donde puedan ser utilizados los nuevos conceptos de números difusos y las regresiones difusas.

1.4 OBTENCIÓN DE DATOS METEOROLÓGICOS

Con el fin de modelar el perfil atmosférico vertical de humedad, temperatura, presión, velocidad y dirección del viento en la zona de interés, se ha planificado el lanzamiento de radio sondas (gráfico 1.4), que son globos llenos de helio y que ascienden a una velocidad constante. Estos globos van equipados con sensores acoplados a un sistema de radio transmisión que envía en forma directa los diferentes datos a una computadora ubicada en tierra. La posición del globo y por ende su velocidad, se calculan mediante el sistema GPS. Los datos son recopilados en el sistema ASCII y luego codificados en las unidades físicas correspondientes. Esta tarea ha estado a cargo del INAMHI, institución estatal ecuatoriana especializada en el ámbito meteorológico y climatológico. Los sitios de lanzamiento son las provincias de Pichincha, Cotopaxi, Tungurahua y Bolívar. Sin embargo los datos que serán utilizados corresponden a la ciudad de Ambato. El cuadro 1.3 muestra los sitios de lanzamiento y las fechas. Cabe indicar que estos globos han alcanzado una altura cercana a los 30 km sobre el nivel del mar.

	HORA	LUGAR	DATOS OBTENIDOS	ALTURA MINIMA	ALTURA MÁXIMA
18 de noviembre de 2008.	09H00	Latacunga	4868 datos	2813 metros	24065 metros
29 de mayo de 2008	09H46	Latacunga	2752 datos	2746 metros	21404 metros
14 de noviembre de 2008	09H15	Latacunga	3462 datos	2813 metros	28345 metros
27 de mayo de 2009	08H28	Ambato	3619 datos	2789 metros	27184 metros
27 de mayo de 2009 2	13H28	Ambato	5404 datos	2789 metros	29038 metros

21 de noviembre de 2008	14H51	Ambato	4398 datos	2530 metros	27136 metros
14 de abril del 2009	17H00	Guadalupe	4079 datos	2789 metros	25475 metros
14 de abril del 2009	10H40	Guadalupe	3644 datos	2789 metros	27037 metros
25 de noviembre de 2008	09H09	Ambato	2226 datos	2530 metros	19242 metros
12 de mayo de 2009	06H40	Salinas	5897 datos	4629 metros	23268 metros
13 de mayo de 2009	06H22	Salinas	8297 datos	4629 metros	28528 metros
13 de mayo de 2009	06H00	Quito	8290 datos	2813 metros	28345 metros
10 de abril del 2009	11H38	Quito	412 datos	3058 metros	9547 metros
10 de abril del 2009	13H26	Quito	3568 datos	3058 metros	24777 metros
23 de abril del 2009	09H05	Quito	8432 datos	4629 metros	28352 metros
23 de abril del 2009	16H26	Quito	8293 datos	4629 metros	28508 metros
23 de mayo del 2009.	15H04	Quito	4325 datos	2789 metros	24850 metros
17 de junio del 2009	08H41	Quito	3259 datos	2789 metros	28047 metros
17 de junio del 2009	16H56	Quito	2654 datos	2789 metros	26689 metros
07 de julio del 2009	06H02	Quito	4639 datos	2789 metros	27142 metros
08 de julio del 2009	07H01	Quito	4183 datos	2789 metros	27122 metros
01 de agosto del 2009	10H41	Quito	3012 datos	2789 metros	26670 metros

Cuadro 1.3 Datos de radio sondeos proporcionados por el INAMHI

Las sondas también han proporcionados datos de variables adicionales pero que para nuestros propósitos no son necesarios. Sin embargo, esta gran variedad de datos si serán de mucha utilidad cuando se utilicen los modelos dinámicos de la atmósfera. En el gráfico 1.4 se observa una radio sonda utilizada.



Gráfico. 1.4 Radio sondas VAISALA

1.4.1 PROCEDIMIENTO PARA LA OBTENCIÓN DE LOS DATOS DE RADIO SONDEOS

M.1 Procesamiento de la información proveniente de los radio sondeos del sistema RS Vaisala en una plataforma común Grads, que es un sistema de despliegue gráfico meteorológico en los diferentes tiempos y sitios en los que se han realizado los lanzamientos correspondientes.

M.2 Verificación de los datos obtenidos con radio sondeos comparándolos con la climatología de la tropósfera a nivel mensual, obtenida de las corridas del modelo global WRF (Weather Research and Forecasting Model).

M.3 Determinación de los sesgos y niveles de significancia para determinar los límites de seguridad para cada uno de los parámetros meteorológicos buscados.

1.4.2 LIBRO DE CÓDIGOS DE LOS DATOS DE RADIO SONDEOS

Los cuadros siguientes (cuadro 1.4) muestran los datos meteorológicos obtenidos por las radio sondas lanzadas en Ambato, y como estas se ubican para el procesamiento de los datos. Adicionalmente, proveen de algunas estadísticas descriptivas. Por ejemplo, el libro de códigos para la variable Humedad muestra la ubicación que ocupa la variable en la columna de la base de datos, valores de los percentiles 25, 50 y 75, la media, número de observaciones, entre otros.

TIEMPO_S		Valor
Atributos estándar	Ubicación	1
	Etiqueta	TIEMPO (seg)
	Tipo	Numérico
	Formato	F11
	Medida	Escala
N	Válidos	15647
	Perdidos	0
Tendencia central y dispersión	Media	4255,27
	Desviación típica	2749,608
	Percentil 25	1954,00
	Percentil 50	3910,00
	Percentil 75	6338,00

RADIOSONDA

		Valor	Recuento	Porcentaje
Atributos estándar	Ubicación	3		
	Etiqueta	RADIO SONDAS LANZADAS EN AMBATO		
N	Tipo	N Numérico		
	Formato	F11		
	Medida	Escala		
	Válidos	15647		
Tendencia central y dispersión	Perdidos	0		
	Media	2,33		
	Desviación típica	,985		
	Percentil 25	2,00		
	Percentil 50	2,00		
Valores etiquetados	Percentil 75	3,00		
	1	27 DE MAYO 2009	3619	23,1%
	2	27 DE MAYO 2009 (2)	5404	34,5%
	3	21 DE NOVIEMBRE 2008	4398	28,1%
	4	25 DE NOVIEMBRE 2008	2226	14,2%

ALTURA_M

		Valor
Atributos estándar	Ubicación	2
	Etiqueta	ALTURA (m)
N	Tipo	N Numérico
	Formato	F11
	Medida	Escala
	Válidos	15647
Tendencia central y dispersión	Perdidos	0
	Media	14500,20
	Desviación típica	7225,507
	Percentil 25	8279,00
	Percentil 50	14040,00
	Percentil 75	20203,00

PRESIÓN

		Valor
Atributos estándar	Ubicación	4
	Etiqueta	PRESIÓN (hPa)
N	Tipo	N Numérico
	Formato	F22.2
	Medida	Nominal

TEMPERAT_A

		Valor
Atributos estándar	Ubicación	7
	Etiqueta	TEMPERATU RA (°C)
N	Tipo	N Numérico
	Formato	F11.2
	Medida	Escala
	Válidos	15647
Tendencia central y dispersión	Perdidos	0
	Media	-42,9271
	Desviación típica	28,78807
	Percentil 25	-66,2500
	Percentil 50	-52,2500
	Percentil 75	-17,3500

TEMPERAT		
		Valor
Atributos estándar	Ubicación	5
	Etiqueta	TEMPERATURA (°K)
	Tipo	Numérico
	Formato	F22.2
	Medida	Nominal

HUMEDAD		
		Valor
Atributos estándar	Ubicación	6
	Etiqueta	HUMEDAD RELATIVA (%)
	Tipo	Numérico
	Formato	F11.2
	Medida	Escala
N	Válidos	15647
	Perdidos	0
Tendencia central y dispersión	Media	26,4951
	Desviación típica	31,40961
	Percentil 25	2,0000
	Percentil 50	13,0000
	Percentil 75	46,0000

Cuadro 1.4 Libro de Códigos de Datos Atmosféricos

1.5. APLICACIÓN Y ANÁLISIS ESTADÍSTICO DE VARIABLES ATMOSFÉRICAS

1.5.1 JUSTIFICACIÓN

En cuanto al lugar de lanzamiento de las naves, que estarán a cargo de la FAE en la ciudad de Ambato, es conveniente recalcar que, los modelos verticales de humedad, temperatura, viento y presión se ensayarán en esta posición; Ambato tiene las siguientes coordenadas: latitud 1°13'28"; longitud 78°37'11" y 2500 metros sobre el nivel del mar (referencia google earth). Esta es la razón por la cual no se apreciarán en los datos la zona de inversión térmica, ubicada bajo los 1000 metros sobre el nivel del mar.

1.5.2. MEDIDAS DESCRIPTIVAS DE LOS DATOS METEOROLÓGICOS

Los principales parámetros que van a ser analizados en este estudio son los que se muestran en el cuadro 1.5; los valores percentiles, la tendencia central y la dispersión. Se analizarán también, la asimetría y curtosis y se calculan las

medidas de forma: coeficientes de asimetría y curtosis. Con el fin de interpretar los datos en los diferentes niveles de la altura.

Resúmenes de casos

RADIO SONDAS LANZADAS EN AMBATO		ALTURA (m)	PRESIÓN (hPa)	TEMPERA TURA (°K)	TEMPERA TURA (°C)	HUMEDAD RELATIVA (%)
27 DE MAYO 2009	N	3619	3619	3619	3619	3619
	Media	15420,85	201,9408	224,9571	-48,1929	31,9229
	Mediana	15463,00	126,6000	218,7000	-54,4500	16,0000
	Mínimo	2789	18,40	191,60	-81,55	1,00
	Máximo	27184	753,00	285,90	12,75	100,00
27 DE MAYO 2009 (2)	N	5404	5404	5404	5404	5404
	Media	15859,37	209,0244	228,3713	-44,7787	28,7006
	Mediana	15493,50	126,0500	222,2000	-50,9500	14,0000
	Mínimo	2789	13,70	189,80	-83,35	1,00
	Máximo	29038	750,00	290,40	17,25	100,00
21 DE NOVIEMBRE 2008	N	4398	4398	4398	4398	4398
	Media	13918,34	241,1870	231,2128	-41,9372	18,6166
	Mediana	13209,00	178,2000	219,0500	-54,1000	6,0000
	Mínimo	2530	17,90	187,80	-85,35	1,00
	Máximo	27126	750,60	291,30	18,15	93,00
25 DE NOVIEMBRE 2008	N	2226	2226	2226	2226	2226
	Media	10853,42	313,4105	241,3233	-31,8267	27,8823
	Mediana	10582,50	273,1000	244,1000	-29,0500	13,0000
	Mínimo	2530	67,20	188,20	-84,95	3,00
	Máximo	19242	751,00	297,60	24,45	95,00
Total	N	15647	15647	15647	15647	15647
	Media	14500,20	231,2765	230,2229	-42,9271	26,4951
	Mediana	14040,00	160,0000	220,9000	-52,2500	13,0000
	Mínimo	2530	13,70	187,80	-85,35	1,00
	Máximo	29038	753,00	297,60	24,45	100,00

Cuadro 1.5 Resúmenes de casos de las variables meteorológicas

Por ejemplo, en el cuadro 1.5 se puede visualizar para la radio sonda lanzada el 27 de mayo de 2009, se obtuvieron 3619 datos para las tres variables atmosféricas y se observa que el análisis estadístico para la temperatura en °C se obtiene: la media es -48,19°C, la mediana es -54,45°C, la temperatura mínima es -81,55°C y la temperatura máxima es 12,75°C y así, puede analizarse para los diferentes días y diferentes variables.

A continuación realizamos el análisis estadístico para cada variable atmosférica:

1.5.3 ANÁLISIS ESTADÍSTICO DE LA PRESIÓN ATMOSFÉRICA (hPa)

La presión atmosférica, en un cierto punto, es la fuerza que la columna atmosférica ejerce sobre ese lugar por unidad de área, debido a la influencia de la fuerza de gravedad terrestre. La unidad que se utilizará es el hectopascal: hPa.

En el Cuadro 1.7 se observa que el valor mínimo menor es 13,70 hPa que corresponde al de la radio sonda del 27 de mayo de 2009 y el valor mínimo mayor es 67,20 hPa, que corresponde al de la radio sonda del 25 de noviembre de 2009. Este último valor no refleja un comportamiento inestable para la variable atmosférica ya que los valores mínimos de las otras radio sondas fluctúan entre 13 y 15 hPa, sino implica que este valor apenas es tomado cuando la radio sonda alcanza los 19242 metros y luego desciende y no toma más mediciones.

Para el valor máximo de la presión se observa que alrededor de los 2700 metros la presión se comporta similarmente para cualquier lanzamiento de radio sondas dando un valor en promedio de 752 hPa.

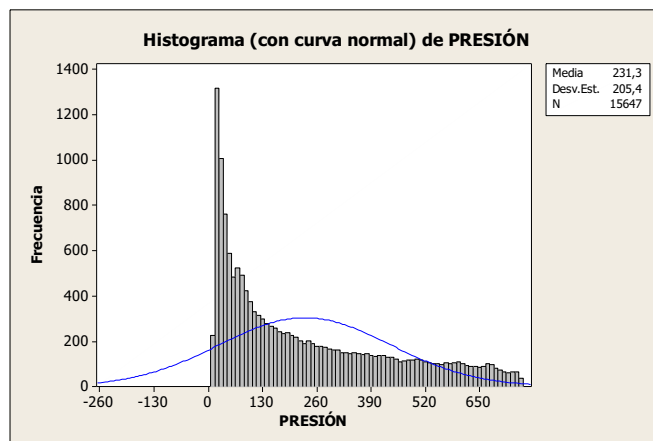
Descriptivos			Estadístico	Error típ.
PRESIÓN	Media		231,28	1,642
	Intervalo de confianza para la media al 95%	Límite inferior	228,06	
		Límite superior	234,49	
	Media recortada al 5%		217,42	
	Mediana		160,00	
	Varianza		42179,020	
	Desv. típ.		205,375	
	Mínimo		14	
	Máximo		753	
	Rango		739	
	Amplitud intercuartil		317	
	Asimetría		,851	,020
	Curtosis		-,461	,039

Cuadro 1.6 Estadísticas descriptivas para la variable presión atmosférica

La amplitud intercuartil en promedio es 316,7; es decir, se encuentran el 50% de los datos, 25% de este valor hacia abajo y 25% de este valor hacia arriba.

La asimetría es 0,851 y el error de la asimetría es 0,02. La distribución de los datos es sesgada (simetría mayor que cero). De la condición anterior concluimos que no existe una asimetría pronunciada. En el gráfico 1.5 se visualiza que no existe una asimetría pronunciada pero, como la asimetría es mayor que 0, los datos se encuentran agrupados a la izquierda

El valor de la curtosis es -0,461 que indica si los datos están concentrados al centro o dispersos, como el valor es negativo la forma es platicúrtica es decir, los datos se encuentran dispersos de su valor promedio.



Gráfica 1.5 Histograma variable presión atmosférica

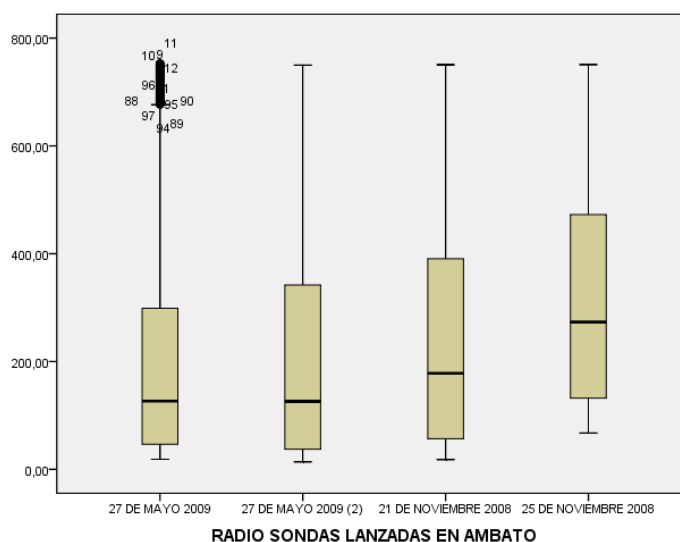


Gráfico 1.6 Diagrama de Cajas presión atmosférica

El gráfico 1.6 muestra un análisis comparativo mediante un diagrama de cajas de los diferentes radiosondeos para la variable presión atmosférica en el cuál concluimos lo siguiente:

1. El 25% de los datos más pequeños obtenidos para los tres lanzamientos se comportan similarmente a diferentes puntos de altura, en el caso de la radiosonda lanzada el 25 de noviembre de 2008 se observa que recién el 25% de los datos se alcanza más arriba del 25% de los datos anteriores debido a que la radio sonda no alcanza hasta una altura similar a la altura de los anteriores lanzamientos que es donde la presión atmosférica desciende.

2. Los valores que se encuentran fuera del máximo de la amplitud intercuartil en el primer lanzamiento son los valores atípicos que, según el Instituto Nacional de Meteorología e Hidrología son valores que se presentan raramente en el Ecuador y que se producen por los factores del clima como son los movimientos de la tierra, la latitud, la disposición del relieve, la altitud, las corrientes marinas, la distancia de Tierras y mares.
3. En general, podemos visualizar por el diagrama de cajas que, el comportamiento de la presión atmosférica en la ciudad de Ambato se comporta similarmente en todos los lanzamientos excepto en el primer lanzamiento que, por información del INAMHI, estos valores no son frecuentes y por ello serán excluidos posteriormente para obtener un buen ajuste de las estimaciones curvilíneas para estas variables (Capítulo 2).

El cuadro 1.7 muestra algunos estimadores robustos centrales para los diferentes radio sondeos que indican trabajar con la Media recortada al 5% de los datos (es decir, quitar el 2,5% de los datos inferiores y el 2,5% de los datos superiores debido a que, estos estimadores tienden a acercarse a la media recortada y no a la media muestral; esto realizamos para dar un mejor pronóstico de cómo se comportaría la presión. La media recortada es 217,42 hPa, ver cuadro 1.6 en el promedio de altura de 2226 metros a 29038 metros, es decir, alrededor de 15632 metros de altura.

Estimadores-M

RADIO SONDAS LANZADAS EN AMBATO	Estimador-M de Huber ^a	Biponderado de Tukey ^b	Estimador-M de Hampel ^c	Onda de Andrews ^d
PRESIÓN (hPa) 27 DE MAYO 2009	146,2268	121,6884	150,6059	121,0474
27 DE MAYO 2009 (2)	149,9079	123,0681	157,8773	122,3004
21 DE NOVIEMBRE 2008	200,3902	194,2876	214,3852	194,0907
25 DE NOVIEMBRE 2008	288,5689	293,1147	301,0616	293,2325

a. La constante de ponderación es 1,339.

b. La constante de ponderación es 4,685.

c. Las constantes de ponderación son 1,700, 3,400 y 8,500.

d. La constante de ponderación es $1,340 \cdot \pi$.

Cuadro 1.7 Estimadores robustos centrales

El cuadro 1.8 muestra el análisis de los percentiles para cada uno de los radio sondeos y el promedio de los percentiles para la variable presión atmosférica. Los

percentiles de interés para el estudio atmosférico del PGA son 50 y 75 a una altura promedio de 14040 metros a 20203 metros (cuadro 1.9).

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Promedio ponderado	PRESIÓN (hPa)	19,6000	25,1000	55,1000	160,0000	371,8000	568,7200	648,7600
Bisagras de Tukey	PRESIÓN (hPa)			55,1500	160,0000	371,7500		

Percentiles

			Percentiles						
			5	10	25	50	75	90	95
Promedio ponderado (definición 1)	PRESIÓN (hPa)	RADIO SONDAS LANZADAS EN AMBATO 27 DE MAYO 2009	21,7000	26,2000	46,5000	126,6000	298,8000	519,9000	618,4000
		27 DE MAYO 2009 (2)	15,3000	18,4000	37,1000	126,0500	341,9250	556,6500	639,0250
		21 DE NOVIEMBRE 2008	22,7000	28,5000	56,6000	178,2000	390,6750	575,7400	657,3150
		25 DE NOVIEMBRE 2008	75,9350	86,2700	131,8500	273,1000	472,6000	620,3500	673,0250
Bisagras de Tukey	PRESIÓN (hPa)	27 DE MAYO 2009			46,5000	126,6000	298,6500		
		27 DE MAYO 2009 (2)			37,1000	126,0500	341,8500		
		21 DE NOVIEMBRE 2008			56,6000	178,2000	390,6000		
		25 DE NOVIEMBRE 2008			131,9000	273,1000	472,5000		

Cuadro 1.8 Análisis de Percentiles Presión Atmosférica

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Promedio ponderado (definición 1)	ALTURA (m)	3879,80	4969,60	8279,00	14040,00	20203,00	25085,60	26674,80
Bisagras de Tukey	ALTURA (m)			8279,50	14040,00	20202,50		

Cuadro 1.9 Análisis de Percentiles Altura

Descriptivos

RADIO SONDAS			Estadístico	Error típ.	
PRESIÓN (hPa)	27 DE MAYO 2009	Media	201,9408	3,20183	
		Intervalo de confianza para la media al 95%	Límite inferior	195,6632	
			Límite superior	208,2184	
		Media recortada al 5%	185,2981		
		Mediana	126,6000		
		Varianza	37101,034		
		Desv. típ.	192,61629		
		Mínimo	18,40		
		Máximo	753,00		
		Rango	734,60		
		Amplitud intercuartil	252,30		
		Asimetría	1,134	,041	
		Curtosis	,231	,081	
		27 DE MAYO 2009 (2)	27 DE MAYO 2009 (2)	Media	209,0244
Intervalo de confianza para la media al 95%	Límite inferior			203,5512	
	Límite superior			214,4976	
Media recortada al 5%	193,1934				
Mediana	126,0500				
Varianza	42121,309				
Desv. típ.	205,23477				
Mínimo	13,70				
Máximo	750,00				
Rango	736,30				
Amplitud intercuartil	304,83				
Asimetría	,979			,033	
Curtosis	-,239			,067	
21 DE NOVIEMBRE 2008	21 DE NOVIEMBRE 2008			Media	241,1870
		Intervalo de confianza para la media al 95%	Límite inferior	235,0502	
			Límite superior	247,3238	
		Media recortada al 5%	228,0187		
		Mediana	178,2000		
		Varianza	43092,726		
		Desv. típ.	207,58787		
		Mínimo	17,90		
		Máximo	750,60		
		Rango	732,70		
		Amplitud intercuartil	334,08		
		Asimetría	,768	,037	
		Curtosis	-,606	,074	
		25 DE NOVIEMBRE 2008	25 DE NOVIEMBRE 2008	Media	313,4105
Intervalo de confianza para la media al 95%	Límite inferior			305,1726	
	Límite superior			321,6485	
Media recortada al 5%	304,8460				
Mediana	273,1000				
Varianza	39281,997				
Desv. típ.	198,19686				
Mínimo	67,20				
Máximo	751,00				
Rango	683,80				
Amplitud intercuartil	340,75				
Asimetría	,510			,052	
Curtosis	-,992			,104	

Cuadro 1.10 Estadísticas descriptivas para las radiosondas lanzadas en Ambato

1.5.4 ANÁLISIS ESTADÍSTICO DE LA TEMPERATURA (°C)

La temperatura es la medida del movimiento molecular o el grado de calor de una sustancia. Las unidades usadas son los grados Kelvin: °K, o grados Celsius: °C.

En el cuadro 1.11, se observa que el valor mínimo en promedio es -85,35 °C, esta temperatura se encuentra a la altura de 16954 metros, es decir, que de ahí en adelante la temperatura comienza a ascender y en los 20003 metros la temperatura alcanza entre -67,95°C y -75,55°C con un intervalo de confianza del 95%, ver cuadro 1.12, que es la altura donde se piensa situar al PGA.

Descriptivos			Estadístico	Error típ.
TEMPERAT_A	Media		-42,9271	,23014
	Intervalo de confianza para la media al 95%	Límite inferior	-43,3782	
		Límite superior	-42,4760	
	Media recortada al 5%		-43,8308	
	Mediana		-52,2500	
	Varianza		828,753	
	Desv. típ.		28,78807	
	Mínimo		-85,35	
	Máximo		24,45	
	Rango		109,80	
	Amplitud intercuartil		48,90	
	Asimetría		,521	,020
	Curtosis		-1,006	,039

Cuadro 1.11 Análisis descriptivo variable Temperatura

Resúmenes de casos			
RADIO SONDAS LANZADAS EN AMBATO		ALTURA (m)	TEMPERATURA (°C)
27 DE MAYO 2009	N	1	1
	Media	20003,00	-75,5500
	Mediana	20003,00	-75,5500
	Máximo	20003	-75,55
	Mínimo	20003	-75,55
	Asimetría	.	.
	Curtosis	.	.
27 DE MAYO 2009 (2)	N	1	1
	Media	20003,00	-67,9500
	Mediana	20003,00	-67,9500
	Máximo	20003	-67,95
	Mínimo	20003	-67,95
	Asimetría	.	.
	Curtosis	.	.
21 DE NOVIEMBRE 2008	N	1	1
	Media	20003,00	-68,3500
	Mediana	20003,00	-68,3500
	Máximo	20003	-68,35
	Mínimo	20003	-68,35
	Asimetría	.	.
	Curtosis	.	.
Total	N	3	3
	Media	20003,00	-70,6167
	Mediana	20003,00	-68,3500
	Máximo	20003	-67,95
	Mínimo	20003	-75,55
	Asimetría	.	-1,715
	Curtosis	.	.

Cuadro 1.12 Temperatura a 20003 metros

La amplitud intercuartil en promedio es 48,90. La asimetría es 0,521 y el error de la asimetría es 0,02. Por lo tanto, no se da una asimetría pronunciada. La

distribución es sesgada a la derecha porque el valor de la asimetría es mayor que 0. El valor de la curtosis es -1,006, que indica si los datos están concentrados al centro o dispersos, como el valor es negativo la forma es platicúrtica es decir, los datos se encuentran dispersos del valor promedio.

El cuadro 1.13 se observa los valores de los percentiles, alrededor de los percentiles 75 y 90 en este intervalo, es donde la temperatura comienza a ascender, y es por ellos que se da un valor grande de descenso en estos percentiles por las observaciones tomadas del 75% de los datos al 90% de los datos.

			Percentiles						
			5	10	25	50	75	90	95
Promedio ponderado(definición 1)	TEMPERATURA (°C)	RADIO SONDAS LANZADAS EN AMBATO 27 DE MAYO 2009	-79,0500	-75,8500	-69,7500	-54,4500	-31,1500	-4,6500	1,6500
		27 DE MAYO 2009 (2)	-80,8500	-75,1000	-65,2500	-50,9500	-23,1500	-1,0000	5,8500
		21 DE NOVIEMBRE 2008	-80,8500	-74,8500	-65,5500	-54,1000	-15,5500	2,0500	8,1550
		25 DE NOVIEMBRE 2008	-80,2150	-75,8500	-67,8500	-29,0500	,1750	12,8800	17,0150
Bisagras de Tukey	TEMPERATURA (°C)	RADIO SONDAS LANZADAS EN AMBATO 27 DE MAYO 2009			-69,7500	-54,4500	-31,2000		
		27 DE MAYO 2009 (2)			-65,2500	-50,9500	-23,1500		
		21 DE NOVIEMBRE 2008			-65,5500	-54,1000	-15,5500		
		25 DE NOVIEMBRE 2008			-67,8500	-29,0500	,1500		

Cuadro 1.13 Percentiles Temperatura

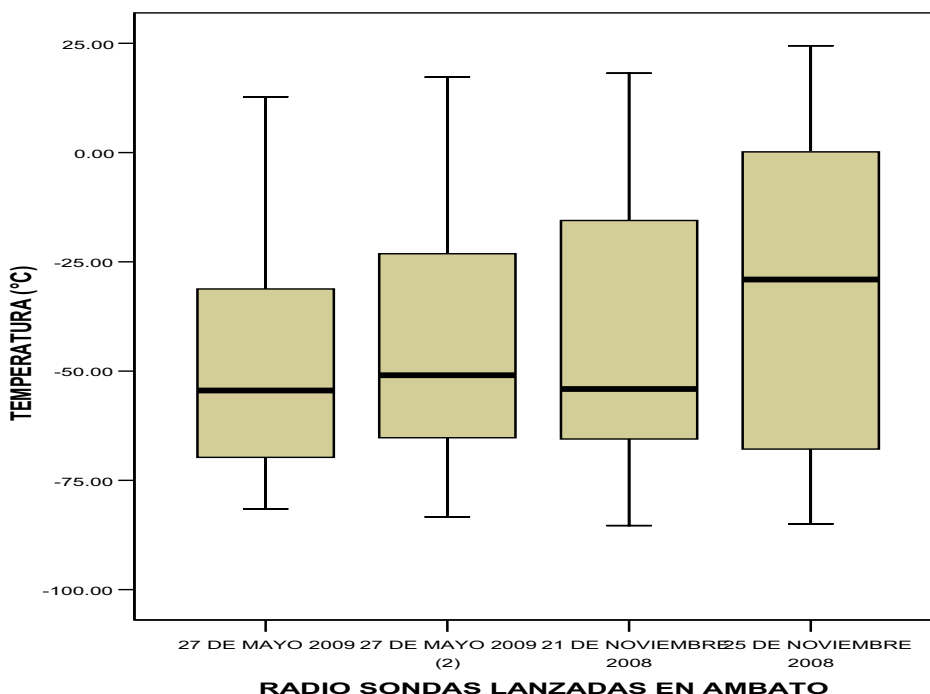


Gráfico 1.7 Diagrama de Cajas Temperatura

En el diagrama de cajas, gráfico 1.7, se observa:

1. Los valores de la temperatura se comportan similarmente en los diferentes días ya que no se visualizan valores atípicos.
2. El valor de la mediana es diferente debido a la cantidad diferente de datos que se obtuvieron por los radio sondeos. No todas las radio sondas alcanza la misma altura.

Estimadores-M

	RADIO SONDAS LANZADAS EN AMBATO	Estimador-M de Huber ^a	Biponderado de Tukey ^b	Estimador-M de Hampel ^c	Onda de Andrews ^d
TEMPERATURA (°C)	27 DE MAYO 2009	-53,6057	-54,6769	-52,1775	-54,6862
	27 DE MAYO 2009 (2)	-49,5098	-50,8345	-48,2266	-50,8307
	21 DE NOVIEMBRE 2008	-46,0587	-46,1103	-44,3803	-46,0746
	25 DE NOVIEMBRE 2008	-31,9387	-31,7384	-31,8267	-31,7365

a. La constante de ponderación es 1,339.

b. La constante de ponderación es 4,685.

c. Las constantes de ponderación son 1,700, 3,400 y 8,500.

d. La constante de ponderación es $1,340 \cdot \pi$.

Cuadro 1.15 Estimadores Robustos para la variable temperatura

El cuadro 1.15 muestra los estimadores robustos para los diferentes días, ellos sugieren trabajar con la media recortada al 5% de los datos, esto implica quitar el 5% de los datos superiores e inferiores para que se puede ajustar de una mejor manera al modelo matemático para la temperatura.

El cuadro 1.16 se muestra en valores numéricos descriptivos obtenidos en los diferentes días por los lanzamientos de radio sondas.

Descriptivos

RADIO SONDAS			Estadístico	Error tip.
TEMPERATURA (°C) 27 DE MAYO 2009	Media		-48,1929	,42830
	Intervalo de confianza para la media al 95%	Límite inferior	-49,0326	
		Límite superior	-47,3531	
	Media recortada al 5%		-49,4425	
	Mediana		-54,4500	
	Varianza		663,870	
	Desv. típ.		25,76567	
	Mínimo		-81,55	
	Máximo		12,75	
	Rango		94,30	
	Amplitud intercuartil		38,60	
	Asimetría		,771	,041
	Curtosis		-,598	,081
	27 DE MAYO 2009 (2)	Media		-44,7787
Intervalo de confianza para la media al 95%		Límite inferior	-45,4923	
		Límite superior	-44,0652	
Media recortada al 5%			-45,7199	
Mediana			-50,9500	
Varianza			715,956	
Desv. típ.			26,75735	
Mínimo			-83,35	
Máximo			17,25	
Rango			100,60	
Amplitud intercuartil			42,10	
Asimetría			,600	,033
Curtosis			-,795	,067
21 DE NOVIEMBRE 2008		Media		-41,9372
	Intervalo de confianza para la media al 95%	Límite inferior	-42,7921	
		Límite superior	-41,0822	
	Media recortada al 5%		-42,5486	
	Mediana		-54,1000	
	Varianza		836,391	
	Desv. típ.		28,92043	
	Mínimo		-85,35	
	Máximo		18,15	
	Rango		103,50	
	Amplitud intercuartil		50,00	
	Asimetría		,424	,037
	Curtosis		-1,153	,074
	25 DE NOVIEMBRE 2008	Media		-31,8267
Intervalo de confianza para la media al 95%		Límite inferior	-33,2534	
		Límite superior	-30,4001	
Media recortada al 5%			-31,8958	
Mediana			-29,0500	
Varianza			1178,158	
Desv. típ.			34,32430	
Mínimo			-84,95	
Máximo			24,45	
Rango			109,40	
Amplitud intercuartil			68,02	
Asimetría			-,016	,052
Curtosis			-1,482	,104

Cuadro 1.16 Análisis Descriptivo de la Temperatura

1.5.5 ANÁLISIS ESTADÍSTICO DE LA HUMEDAD RELATIVA (%)

La humedad relativa se le conoce como el vapor de agua contenido en la atmósfera. Puede hablarse de humedad específica o de humedad relativa, que no es más que la razón entre la presión parcial de vapor de agua y su presión de vapor de saturación. Se la expresa en porcentaje: %.

Procedemos a realizar el análisis de tendencia. El cuadro 1.17 muestra el resumen estadístico para la variable HUMEDAD. La amplitud intercuartil en promedio para todos los lanzamientos de radio sondas es 44; es decir, se encuentran el 50% de los datos, centrales (entre los cuartiles laterales). La asimetría es 1,085 y el error de la asimetría es 0,02. La distribución es sesgada a la derecha porque el valor de la asimetría es mayor que 0. De la condición anterior concluimos que no existe una asimetría pronunciada para esta variable. De igual manera observando el gráfico 1.8 se visualiza que no existe una asimetría pronunciada.

El valor de la curtosis es -0,4 que indica si los datos están concentrados al centro o dispersos, como el valor es negativo la forma es platicúrtica es decir, los datos se encuentran dispersos de su valor promedio.

Descriptivos			Estadístico	Error típ.
HUMEDAD RELATIVA (%)	Media		26,4951	,25110
	Intervalo de confianza para la media al 95%	Límite inferior	26,0029	
		Límite superior	26,9873	
	Media recortada al 5%		24,1662	
	Mediana		13,0000	
	Varianza		986,564	
	Desv. típ.		31,40961	
	Mínimo		1,00	
	Máximo		100,00	
	Rango		99,00	
	Amplitud intercuartil		44,00	
	Asimetría		1,085	,020
	Curtosis		-,400	,039

Cuadro 1.17 Estadísticas descriptivas para la variable Humedad Relativa

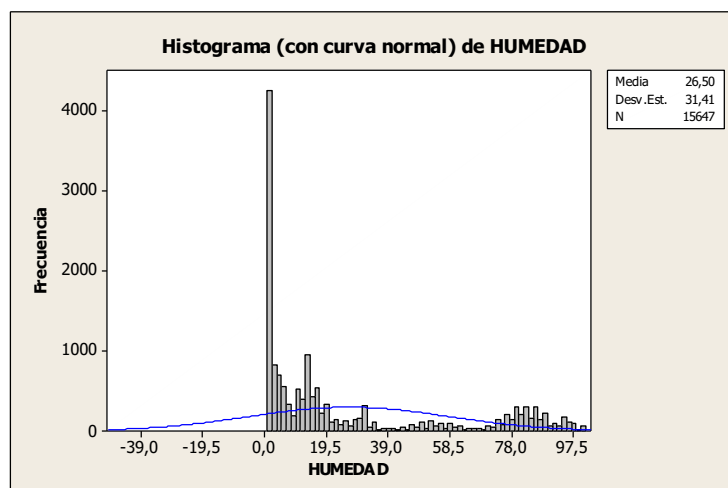


Gráfico 1.8 Histograma con curva normal para la variable Humedad Relativa en (%)

En el análisis descriptivo de las radio sondas de esta variable cuadro 1.17, se observa que el valor mínimo es 1 % y este se encuentra presente en todos los lanzamientos y el mínimo 3% que corresponde a la radio sonda del 25 de noviembre de 2008. Estos valores bajos están dentro de los intervalos del comportamiento de la humedad a una altura de 20000 metros.

Para el valor máximo de la humedad se observa que alrededor de los 2700 m la humedad se comporta similarmente para cualquier lanzamiento de radio sondas, con un valor promedio entre 95 y 100%.

El gráfico 1.9 muestra un análisis comparativo mediante un diagrama de cajas de los diferentes radios sondeos en el cuál se observa lo siguiente:

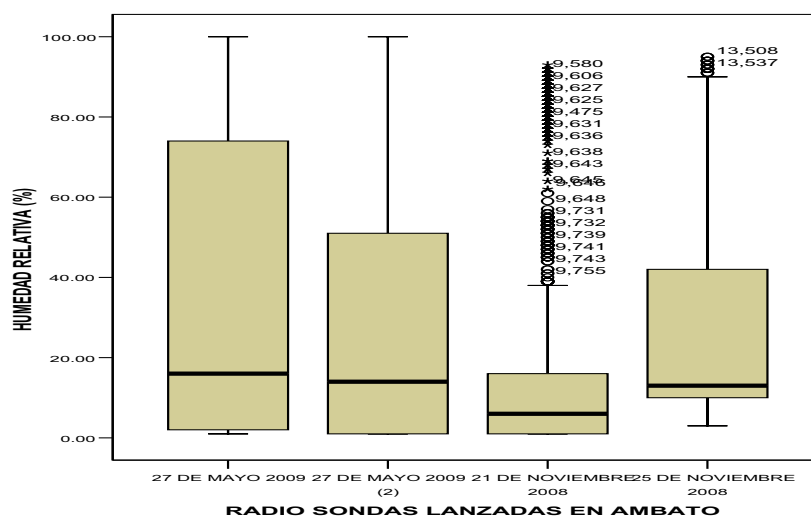


Gráfico 1.9 Diagrama de Caja Humedad Relativa

1. El 25% de los datos mínimos obtenidos para los tres lanzamientos se comportan similarmente a diferentes puntos de altura, en el caso de la radiosonda lanzada el 25 de noviembre de 2008 se observa que recién el 25% de los datos se alcanza más arriba del 25% de los datos anteriores casi al 50% debido a que la radio sonda no alcanza hasta una altura similar a la altura de los anteriores lanzamientos que es donde la humedad relativa desciende y por ende se obtienen más datos bajos.
2. Los valores que se encuentran fuera del máximo de la amplitud intercuartil en el tercer y cuarto lanzamiento son los valores atípicos que, según el Instituto Nacional de Meteorología e Hidrología son valores que se presentan comúnmente en el Ecuador y se encuentran de 18 Km en adelante y que se producen por los factores del clima como son los movimientos de la tierra, la latitud, la disposición del relieve, la altitud, las corrientes marinas, la distancia de tierras y mares.
3. En general, podemos visualizar por el diagrama de cajas que, el comportamiento en promedio de la mediana de la Humedad Relativa en la ciudad de Ambato hasta los 15 Km se comporta similarmente en todos los lanzamientos esto significa que hasta los 18 Km la humedad relativa se comporta similarmente en todos los lanzamientos de radio sondas.

El cuadro 1.18 muestra el análisis de los percentiles para cada uno de los radio sondeos y el promedio de los percentiles para la variable Humedad Relativa. Los percentiles de interés para el estudio atmosférico del PGA son 50 y 75 a una altura promedio de 14040 metros a 20203 metros (cuadro 2.5). El percentil 50 indica que a una altura de 14040 metros la Humedad Relativa máxima que alcanza es 16% y, el percentil 75 indica que a la altura de 20203 metros la Humedad Relativa alcanza hasta un valor de 74%.

Percentiles

RADIO SONDAS LANZADAS EN AMBATO			Percentiles						
			5	10	25	50	75	90	95
Promedio ponderado(definición 1)	HUMEDAD RELATIVA (%)	27 DE MAYO 2009	1,0000	1,0000	2,0000	16,0000	74,0000	90,0000	96,0000
		27 DE MAYO 2009 (2)	1,0000	1,0000	1,0000	14,0000	51,0000	83,0000	89,0000
		21 DE NOVIEMBRE 2008	1,0000	1,0000	1,0000	6,0000	16,0000	82,0000	86,0000
		25 DE NOVIEMBRE 2008	3,0000	3,0000	9,7500	13,0000	42,0000	80,0000	85,0000
Bisagras de Tukey	HUMEDAD RELATIVA (%)	27 DE MAYO 2009			2,0000	16,0000	74,0000		
		27 DE MAYO 2009 (2)			1,0000	14,0000	51,0000		
		21 DE NOVIEMBRE 2008			1,0000	6,0000	16,0000		
		25 DE NOVIEMBRE 2008			10,0000	13,0000	42,0000		

Cuadro 1.18 Análisis de Humedad Relativa

El cuadro 1.19 muestra los valores estadísticos descriptivos que se obtuvieron en los lanzamientos de radio sondeos para la ciudad de Ambato.

El cuadro 1.20 muestra algunos estimadores robustos centrales para los diferentes radio sondeos que indican trabajar con la media recortada al 5% de los datos, es decir, quitar el 2.5% de los datos inferiores y el 2,5% de los datos superiores para esta variable, debido a que estos estimadores tienden a acercarse a la media recortada y no a la media muestral; esto realizamos para dar un mejor pronóstico de cómo se comportaría la humedad relativa. La media recortada es 24.16% cuadro 1.13 en el promedio de altura de 2226 metros a 29038 metros; es decir, alrededor de 15632 metros de altura.

Estimadores-M

RADIO SONDAS LANZADAS EN AMBATO	Estimador-M de Huber ^a	Bponderado de Tukey ^b	Estimador-M de Hampel ^c	Onda de Andrews ^d
HUMEDAD RELATIVA (%) 27 DE MAYO 2009	18,2884	9,9558	18,7022	9,8206
27 DE MAYO 2009 (2)	17,7132	11,8480	17,7511	11,8219
21 DE NOVIEMBRE 2008	7,5633	5,9344	6,3327	5,9346
25 DE NOVIEMBRE 2008	16,3106	11,6490	14,6446	11,6326

a. La constante de ponderación es 1,339.

b. La constante de ponderación es 4,685.

c. Las constantes de ponderación son 1,700, 3,400 y 8,500.

d. La constante de ponderación es 1,340*pi.

Cuadro 1.20 Estimadores Robustos para la Humedad Relativa

Descriptivos

RADIO SONDAS			Estadístico	Error típ.	
HUMEDAD RELATIVA (%) 27 DE MAYO 2009	Media		31,9229	,58649	
	Intervalo de confianza para la media al 95%	Límite inferior	30,7730		
		Límite superior	33,0728		
	Media recortada al 5%		29,9892		
	Mediana		16,0000		
	Varianza		1244,850		
	Desv. típ.		35,28242		
	Mínimo		1,00		
	Máximo		100,00		
	Rango		99,00		
	Amplitud intercuartil		72,00		
	Asimetría		,793		,041
	Curtosis		-1,071		,081
27 DE MAYO 2009 (2)	Media		28,7006	,42414	
	Intervalo de confianza para la media al 95%	Límite inferior	27,8691		
		Límite superior	29,5321		
	Media recortada al 5%		26,6628		
	Mediana		14,0000		
	Varianza		972,171		
	Desv. típ.		31,17965		
	Mínimo		1,00		
	Máximo		100,00		
	Rango		99,00		
	Amplitud intercuartil		50,00		
	Asimetría		,859		,033
	Curtosis		-,700		,067
21 DE NOVIEMBRE 2008	Media		18,6166	,42092	
	Intervalo de confianza para la media al 95%	Límite inferior	17,7914		
		Límite superior	19,4419		
	Media recortada al 5%		15,7237		
	Mediana		6,0000		
	Varianza		779,197		
	Desv. típ.		27,91410		
	Mínimo		1,00		
	Máximo		93,00		
	Rango		92,00		
	Amplitud intercuartil		15,00		
	Asimetría		1,758		,037
	Curtosis		1,445		,074
25 DE NOVIEMBRE 2008	Media		27,8823	,61000	
	Intervalo de confianza para la media al 95%	Límite inferior	26,6861		
		Límite superior	29,0785		
	Media recortada al 5%		25,9232		
	Mediana		13,0000		
	Varianza		828,300		
	Desv. típ.		28,78020		
	Mínimo		3,00		
	Máximo		95,00		
	Rango		92,00		
	Amplitud intercuartil		32,25		
	Asimetría		1,098		,052
	Curtosis		-,432		,104

Cuadro 1.19 Estadísticas descriptivas para las radiosondas lanzadas en Ambato

1.5.6 ANÁLISIS DE LA VELOCIDAD Y DIRECCIÓN DEL VIENTO (M/S)

El viento es el movimiento (horizontal) del aire que fluye respecto de la superficie de la tierra. Se estudiarán dos de sus aspectos fundamentales, la velocidad y la dirección. La unidad utilizada para la velocidad o intensidad es metro por segundo: m/s, para la velocidad y, para la dirección que, a su vez, utiliza un código numérico que indica el ángulo desde donde sopla el viento con respecto al norte, de conformidad con la convención siguiente: 0° = norte; 90° = este; 180° = sur; 270° oeste.

Para analizar estas variables muy inestables es posible que las técnicas estadísticas y las herramientas difusas no sean el mejor método para predecir el comportamiento de estas variables a futuro. Sin embargo, es necesario conocer el comportamiento de estas variables entre 18 a 21 km de altura, de interés particular del PGA. Por lo que se realiza un análisis estadístico de esta variable en los diferentes niveles de la altura debido a su comportamiento inestable y caótico, con los datos proporcionados por los radio sondeos en la ciudad de Ambato, gráfico 1.11. En promedio, el cuadro 1.21 muestra el análisis de tendencias del comportamiento de la velocidad del viento de 18 a 20 Km.

Descriptivos			Estadístico	Error tip.
VELOCIDAD_Y	Media		3,3919	,11787
	Intervalo de confianza para la media al 95%	Límite inferior	3,1606	
		Límite superior	3,6232	
	Media recortada al 5%		3,6352	
	Mediana		3,9500	
	Varianza		16,590	
	Desv. típ.		4,07305	
	Mínimo		-11,98	
	Máximo		9,98	
	Rango		21,96	
	Amplitud intercuartil		5,63	
	Asimetría		-,903	,071
	Curtosis		,661	,141
	VELOCIDAD_X	Media		7,0263
Intervalo de confianza para la media al 95%		Límite inferior	6,4840	
		Límite superior	7,5685	
Media recortada al 5%			7,1410	
Mediana			4,6700	
Varianza			91,202	
Desv. típ.			9,54999	
Mínimo			-10,80	
Máximo			21,99	
Rango			32,79	
Amplitud intercuartil			17,10	
Asimetría			-,027	,071
Curtosis			-1,313	,141

Cuadro 1.21 Análisis de tendencia de la velocidad del viento

Del cuadro se concluye que a esa altura la velocidad del viento es en la componente y es un mínimo de $-11,98$ m/s y máximo $9,98$ m/s en cambio para la componente x la velocidad mínima es $-10,80$ m/s y máxima $21,9$ m/s.

En el gráfico 1.10 se puede ver lo complicado de la trayectoria y la variabilidad del viento que desplaza la radio sonda en la altura. Se puede dar como primer ajuste un área en la cual según las posibles direcciones y velocidades del viento (para las cuales se puede dar rangos entre los que estas varían) se dé el caso de que la sonda termine dentro de un sitio pronosticado.

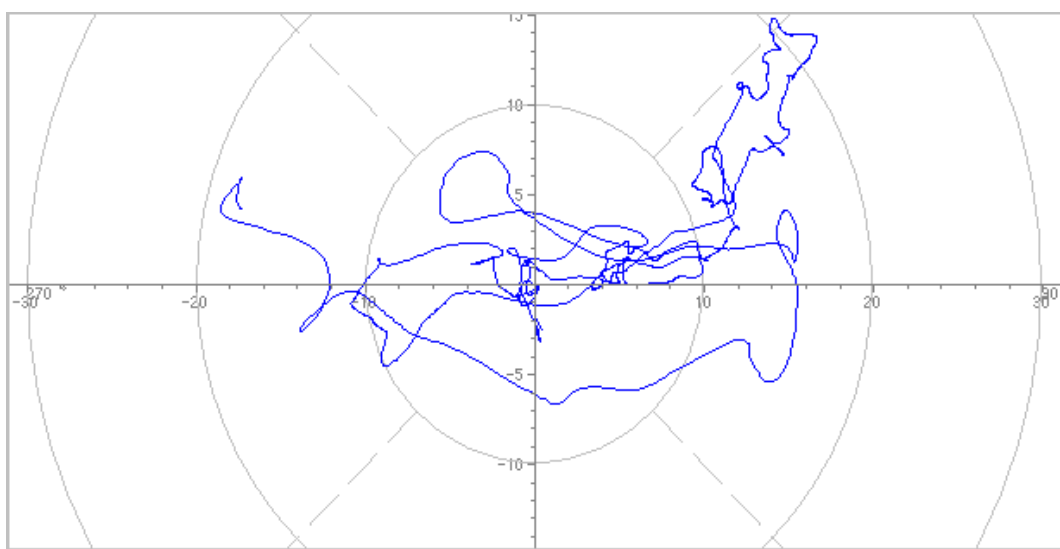


Gráfico 1.10 Trayectoria de una radiosonda

Los datos obtenidos no implican que siempre tomarán ese valor sino que resulta del promedio obtenido en los días medidos; considerar una precisión adecuada para la velocidad y dirección del viento es un proceso complejo que mediante técnicas estadísticas es imposible predecir ya que se deben considerar distintos principios físicos y técnicas numéricas, así como también disponer de una buena caracterización del estado inicial del sistema. Actualmente, se han implementado modelos atmosféricos, herramientas matemáticas construidas en base a la resolución de sistemas de ecuaciones diferenciales parciales de carácter físico y dinámico que gobiernan los movimientos en la atmósfera. Por lo general, estas ecuaciones son no lineales y por tanto, imposibles de resolver exactamente por lo que se recurre a técnicas numéricas que brinda soluciones bastante aproximadas. Además, debido a la complejidad y exceso de eventos a tomar en cuenta como

turbulencias, difusión, radiación, intercambio de calor, tipo de suelo, orografía, fuerza de Coriolis, entre otros, los modelos suelen recurrir a parametrizaciones que simplifiquen los cálculos; inclusive muchos procesos físicos no se pueden modelar explícitamente debido a que todavía no se entienden lo suficientemente bien como para representarlos en formato de ecuación, o bien, porque no se cuenta con los datos apropiados.

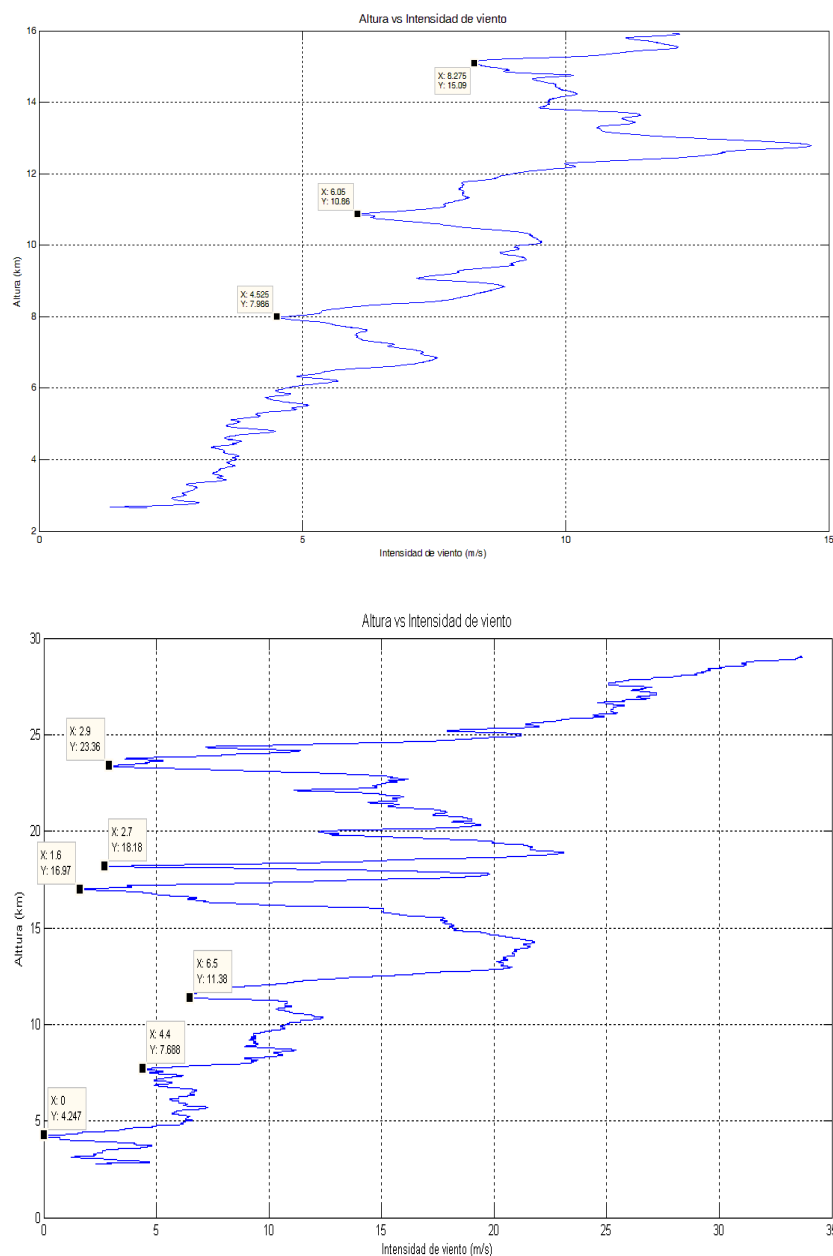


Gráfico. 1.11 Comportamiento del viento vs. Altura

Sin embargo, a pesar de estas herramientas numéricas y el asombroso poder del cálculo de los computadores, la naturaleza se escapa de todas las ecuaciones y métodos numéricos creados.

Las cartas del tiempo son herramientas útiles que permiten: determinar la dirección y velocidad del viento, la ubicación de los centros de alta y baja presión, la posición de los frentes de mal tiempo, variables de mayor importancia para nuestro estudio.

La interpretación de las cartas del tiempo, se basan en la idea de curva de nivel. Las principales curvas de nivel que se utilizan en meteorología son las isotermas, curvas conformadas por puntos que tienen igual temperatura, con el fin de diferenciar claramente las regiones frías de las calientes. Otras curvas de nivel fundamentales son las isóbaras, líneas de igual presión que, a su vez, permiten identificar las zonas de alta presión y las de baja presión y que también proporcionan información respecto a la dirección e intensidad del viento.

1.6 EL MODELO ATMOSFÉRICO WRF PARA LA VELOCIDAD Y DIRECCIÓN DEL VIENTO

Este modelo ha sido desarrollado tanto para servir en pronóstico operacional como para la investigación atmosférica, es el resultado de la colaboración del National Oceanic and Atmospheric Administration NOAA- NCEP y la comunidad de investigación meteorológica de los EE.UU. Es un modelo no hidrostático². Trabaja sobre un sistema operativo basado en Linux. Este modelo se alimenta tanto de datos satelitales como de estaciones meteorológicas o de radio sondeos.

1.6.1 METODOLOGÍA DE TRABAJO

² es decir, que toma en cuenta los movimientos verticales que se dan en la atmósfera y permite una asimilación cuatridimensional de los datos.

El proceso de caracterizar o “reconstruir” la atmósfera, en este caso para la zona del proyecto PGA, es bastante arduo debido al comportamiento inestable de ésta en el Ecuador. Para ello se analizarán los datos históricos troposféricos mes a mes de al menos 5 años, de los diferentes niveles según la atmósfera estándar, con el modelo de mesoescala WRF. La atmósfera en la zona de estudio reconstruida con WRF se validará con los datos obtenidos por las radiosondas realizados. Al momento se han lanzado 22 radiosondas pero, solamente 4 corresponden a la ciudad de Ambato.

1.6.2 RESULTADOS OBTENIDOS DEL MODELO WRF

En el proceso de análisis de los datos de radiosondas, uno de los parámetros más importantes es el comportamiento de la velocidad y dirección del viento en la Tropósfera. Para ello, se ha logrado migrar del sistema DigiCora original de radio sondeos a una plataforma unificada (gráfico. 1.12) con el propósito de visualizar en conjunto los radio sondeos efectuados dentro del marco del proyecto en la zona de estudio.

Analizando este perfil promedio de los radio sondeos se observan en los niveles superiores que los vientos provienen mayormente del oeste, con una velocidad de hasta 15 m/s, mientras que en los niveles netamente inferiores la dirección del viento varía tanto del Este a Oeste con velocidades inferiores menores a 8 m/s.

Una vez corrido el modelo WRF para 2 días diferentes en la zona de estudio observamos que a una altura de 17 Km la dirección del viento proviene del Noreste con una velocidad de 10 m/s en el Día 1 a nivel de 79° de longitud Oeste y centrada a 1.5 ° de latitud Sur. Mientras que, en el Día 2 se observa que esta velocidad del viento proviene del este a la misma altura con velocidades de 10 m/s.

Para poder entender los procesos descendentes y ascendentes a partir de las salidas del modelo WRF que se generan en la atmósfera se determinan las velocidades verticales en la zona de estudio observándose que en la zona 78°, 79° se producen mayores velocidades verticales ascendentes. Gráfico 1.12.

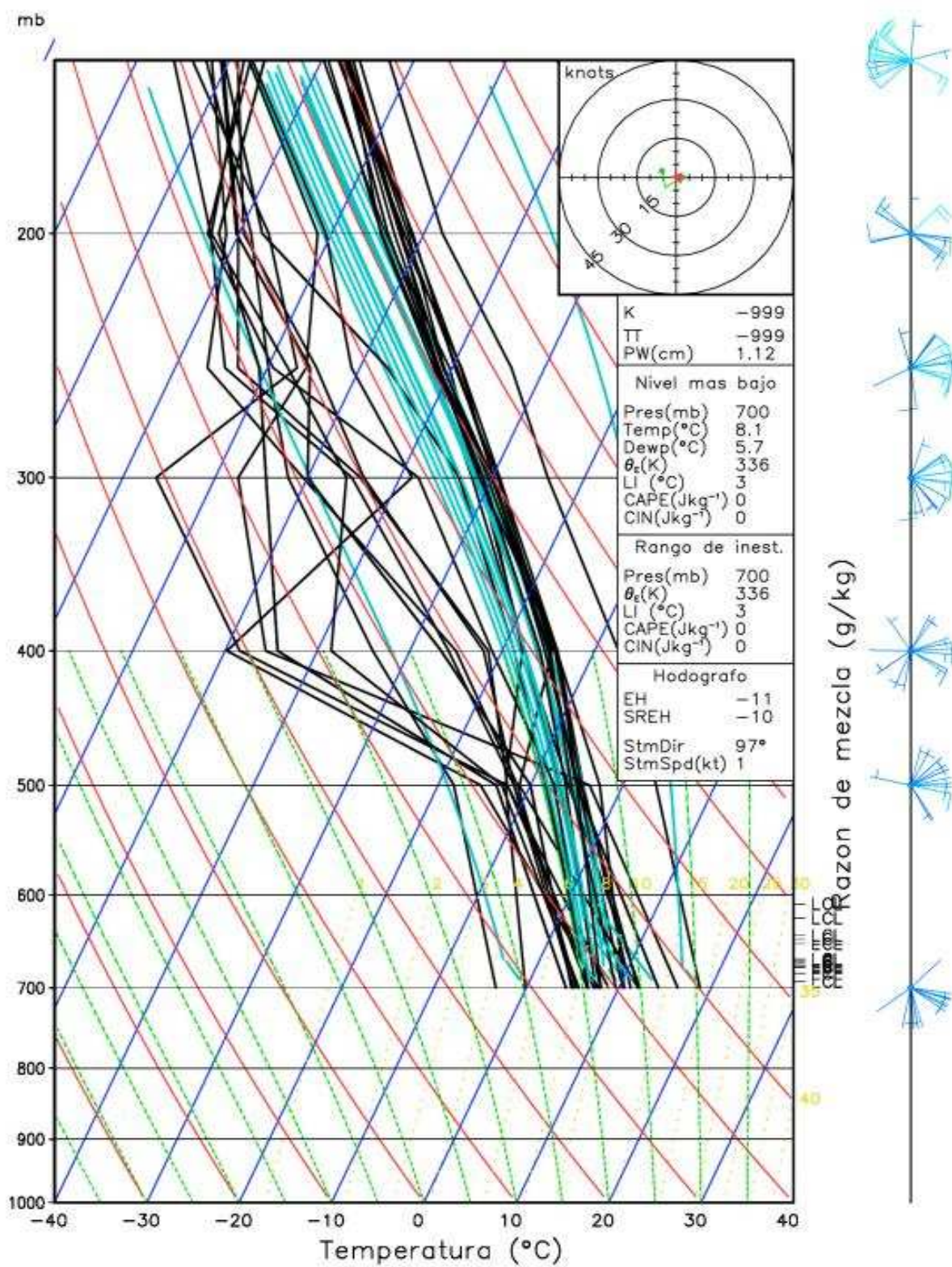


Gráfico 1.12 Perfil vertical de la zona de estudio (Ambato) mostrando los promedios de las velocidades de viento detectada en 15 radio sondeos.

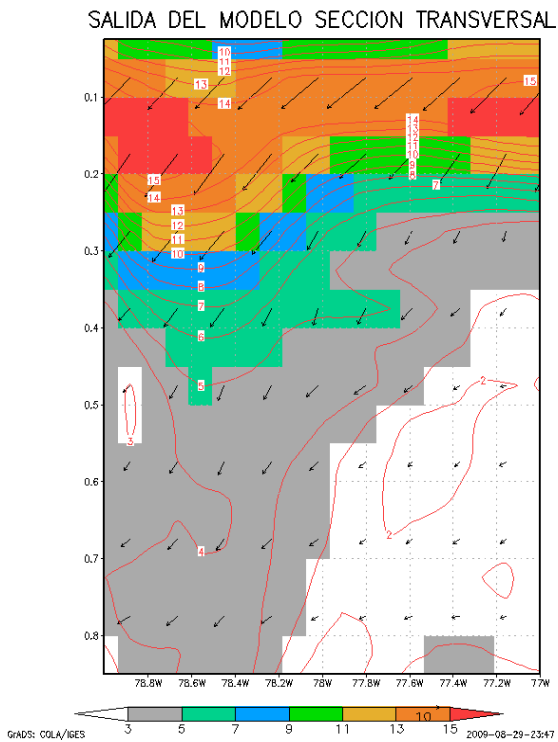


Gráfico. 1.13 Salida del modelo WRF Día 1

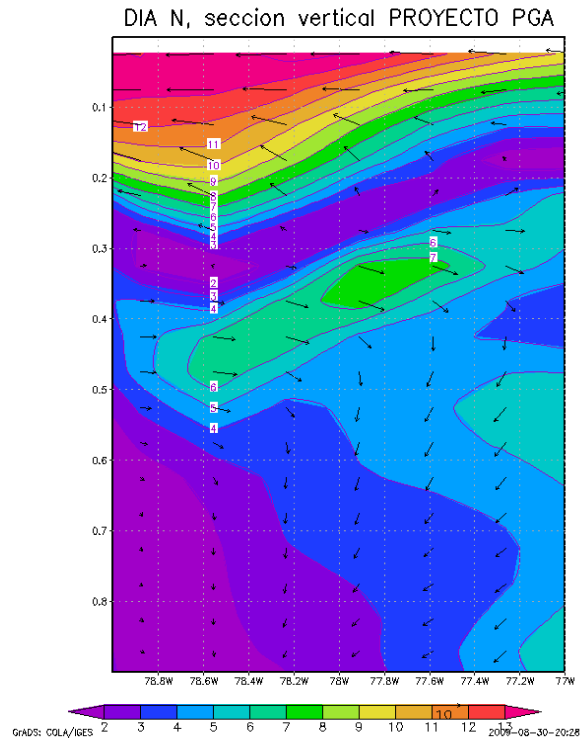


Gráfico. 1.14 Salida del modelo WRF Día 2

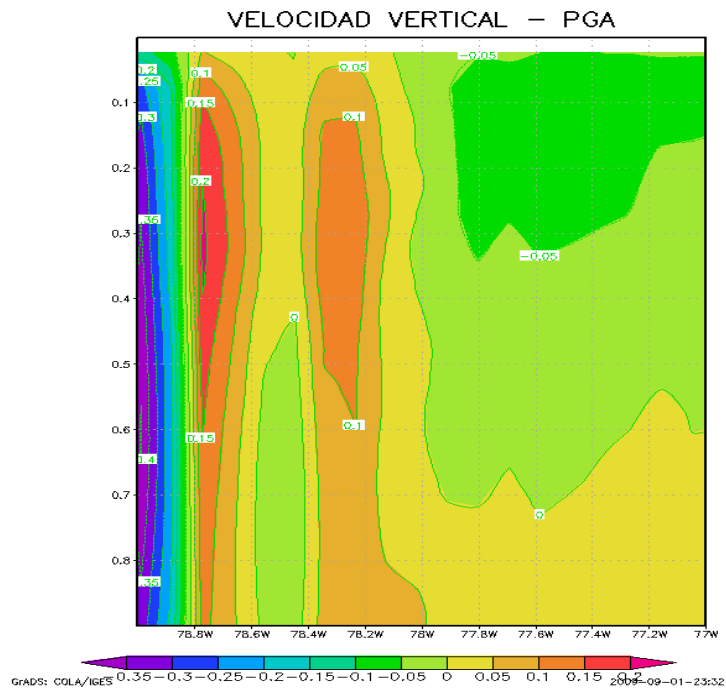


Gráfico. 1.15 Velocidades verticales del viento en la zona de estudio

1.6.3 CONCLUSIONES Y RECOMENDACIONES DE LOS RADIO SONDEOS UTILIZANDO EL MODELO WRF

1. Al analizar en dos días diferentes los datos obtenidos del modelo se observa que a una altura de 17 Km la velocidad del viento es de 15 m/s sesgado hacia las longitudes más occidentales. Se propondría incorporar simulaciones del modelo hasta los 25 Km de altura con el propósito de entender y encontrar láminas con velocidades mínimas del vector viento.
2. Se recomienda que los lanzamientos de las radiosondas deben coincidir con las horas estándar internacionales dadas por la Organización Meteorológica Mundial 00Z y 12Z para efectuar las comparaciones con la salida del modelo.
3. Con estos diagramas se puede analizar la estabilidad de la atmósfera en relación a los vientos cruzados y se observan en los niveles superiores que los vientos provienen mayormente del oeste, con una velocidad de hasta 15 m/s, mientras que en los niveles netamente inferiores la dirección del viento varía tanto del este a oeste con velocidades inferiores menores a 8 m/s
4. Una vez corrido el modelo WRF para dos días diferentes en la zona de estudio observamos que a una altura de 17 Km la dirección del viento proviene del Noreste con una velocidad de 10 m/s en el Día 1 a nivel de 79° de longitud Oeste y centrada a 1.5° de latitud Sur. Mientras, en el día 2 se observa que esta velocidad del viento proviene del Este con velocidades de 10 m/s.
5. Para poder entender los procesos descendentes y ascendentes a partir de las salidas del modelo WRF que se generan en la atmosfera se determinan las velocidades verticales en la zona de estudio observándose que en la zona 78°, 79° se observan mayores velocidades ascendentes.

CAPÍTULO 2: CONTRASTES PARAMÉTRICOS DE LAS VARIABLES METEOROLÓGICAS.

Realizamos los contrastes paramétricos y no paramétricos para comprender el método general del contraste de hipótesis para obtener información a partir de una muestra y tomar decisiones para comprender el significado del p-valor de un contraste en el mismo día y en los diferentes días con los datos de las radiosondas, para averiguar si el comportamiento de las variables atmosféricas son similares y en qué altura son similares, o son completamente diferentes en el tiempo.

2.1 CONTRASTE PARA UNA MEDIA. INTERVALO DE CONFIANZA

2.1.1 PRUEBA t DE UNA MUESTRA

La Prueba t (test de student) de 1 muestra o un intervalo de confianza t para la media permite comparar la media de la muestra con un valor hipotético en relación con la variabilidad en la muestra. Los resultados indican si la diferencia entre los valores es estadísticamente significativa. Por ejemplo, en nuestro caso de estudio el INAMHI en relación a la ubicación geográfica en la que se encuentra nuestro país, considera que las variables meteorológicas se comportan en promedio en la ciudad de Ambato, de acuerdo a los datos del cuadro 2.1 a una altura de 18 Km a 20 Km.

Media					
FUENTE INAMHI	ALTURA (m)	PRESIÓN (hPa)	TEMPERATURA (°K)	TEMPERATURA (°C)	HUMEDAD RELATIVA (%)
Total	18978,44	68,7611	200,0132	-73,1368	10,8756

Cuadro 2.1. Datos proporcionados por el INAMHI

Entonces, realizamos la prueba t de 1 muestra a cada una de las variables atmosféricas para constatar si podemos afirmar estadísticamente que los datos obtenidos por las mediciones de radiosondeos tienden a comportarse similarmente.

2.1.2 ANÁLISIS PARA UNA T DE UNA MUESTRA DE VARIABLES ATMOSFÈRICAS

Para una t de una muestra con dos colas, para las variables atmosféricas se plantea la siguiente hipótesis:

$$H_0: \mu = \mu_0 \text{ versus } H_1: \mu \neq \mu_0$$

Donde μ es la media del valor obtenido de las radiosondas lanzadas en la ciudad de Ambato y μ_0 es la media hipotética dado en la tabla 2.1.

Trabajamos con un intervalo de confianza del 95% para la diferencia entre la media del proceso μ , y el valor de contraste hipotético, μ_0

Para realizar un contraste de hipótesis sobre una media utilizamos el paquete SPSS17.0 y MINITAB que nos permitirá realizar el contraste bilateral de la hipótesis nula.

	N	Media	Desviación típ.	Error típ. de la media
PRESIÓN (hPa)	1194	69,1518	6,95823	,20137

Cuadro 2.2 Estadísticos para una muestra presión atmosférica

	N	Media	Desviación típ.	Error típ. de la media
TEMPERATURA (°C)	1194	-71,9071	2,72096	,07874

Cuadro 2.4 Estadísticos para una muestra (temperatura)

	N	Media	Desviación típ.	Error típ. de la media
HUMEDAD RELATIVA (%)	1194	7,1357	6,27105	,18148

Cuadro 2.6 Estadísticos para una muestra humedad

En el cuadro 2.2 se observa el total de datos que fueron obtenidos de 18 Km a 20 Km (ya que en esta altura es conveniente trabajar para el proyecto). Dado que la media muestral de la variable presión atmosférica es igual a 69,1518 hPa, vamos a constatar la hipótesis nula

$$H_0: \mu = 68,7611 \text{ versus } H_1: \mu_0 \neq 68,7611$$

Estos datos ingresamos en el SPSS 17.0. Los resultados de muestran en el cuadro 2.3.

	Valor de prueba = 68.7611					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
PRESIÓN (hPa)	1,940	1193	,053	,39066	-,0044	,7857

Cuadro 2.3 Prueba para una muestra Presión

	Valor de prueba = -73.1368					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
TEMPERATURA (°C)	15,616	1193	,000	1,22968	1,0752	1,3842

Cuadro 2.5 Prueba para una muestra temperatura

	Valor de prueba = 10.876					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
HUMEDAD RELATIVA (%)	-20,610	1193	,000	-3,74032	-4,0964	-3,3843

Cuadro 2.7 Prueba para una muestra humedad relativa

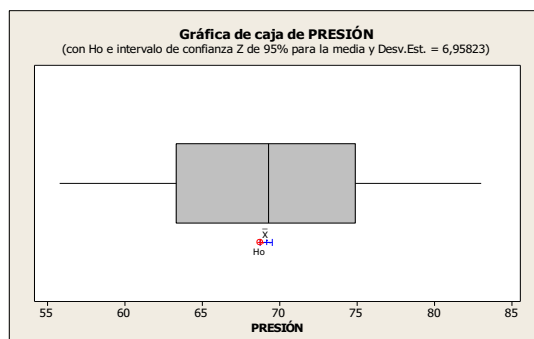


Gráfico 2.1 Diagrama de cajas Presión

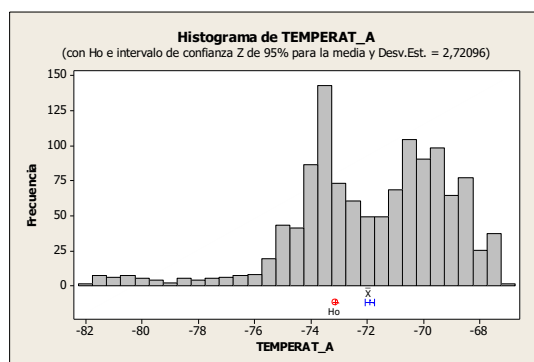


Gráfico 2.2 Histograma de la Temperatura

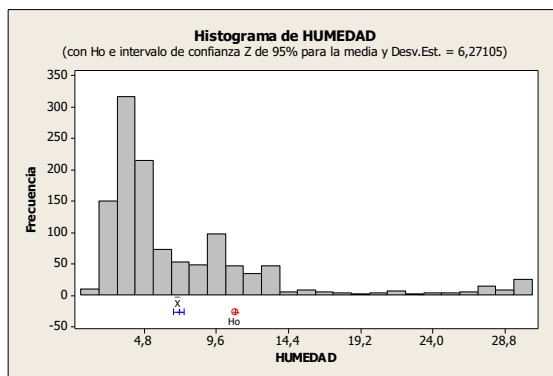


Gráfico 2.3 Histograma de la humedad relativa

El gráfico 2.1 muestra el punto rojo que corresponde a la hipótesis nula que está dentro del intervalo de confianza por lo tanto, se acepta la hipótesis nula.

2.1.3 CONCLUSIONES

1. Para el análisis de la presión atmosférica como el nivel crítico, sig. (*bilateral*), es igual a 0,053 es igual al valor de significación 0,05, entonces, aceptamos la hipótesis nula; es decir, la media de la presión atmosférica de 18 a 20 Km proporcionada por el INAMHI puede considerarse 68,7611 hPa. También obtenemos el resultado del estadístico de contraste:

$$t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} = 1,94 \text{ Los grados de libertad de la t de Student: } gl=1193,$$

La diferencia entre la media muestral y el valor con el que se compara:

$$\bar{x} - \mu_0 = 0.394,$$

Y el intervalo de confianza al 95% para la diferencia entre la media poblacional μ , y la media hipotética, 68,7611:

Intervalo de confianza para la diferencia =(-0,0044 , 0,7857):

De esto último podemos obtener un intervalo de confianza al 95% para la media del los datos de la presión atmosférica de 18 a 20 Km:

$$-0,0044 < \mu - 68,7611 < 0,7857$$

$$68,76 < \mu < 69,55$$

2. Para el análisis de la temperatura como el nivel crítico, sig (bilateral), es igual a 0 menor al nivel de significación $\alpha = 0,05$, entonces, rechazamos la hipótesis nula; es decir, la media de la temperatura de 18 a 20 Km en la ciudad de Ambato no puede considerarse similar a la temperatura $-71,9071^{\circ}\text{C}$.

3. Para el análisis de la humedad relativa como el nivel crítico, sig (bilateral), es igual a 0 menor al nivel de significación $\alpha = 0,05$ entonces, rechazamos la hipótesis nula; es decir, la media de la humedad relativa de 18 a 20 Km en la ciudad de Ambato no puede considerarse similar a la humedad relativa 10,8756 %.

2.2. COMPARACIÓN DE DOS MEDIAS CON MUESTRAS RELACIONADAS.

Se calcula entonces un intervalo de confianza y realizamos una prueba de hipótesis de la diferencia media entre las observaciones relacionadas en la atmósfera de 19 a 20 Km. Una prueba t relacionada hace corresponder respuestas que son dependientes o relacionadas de una manera "en parejas". Esta correspondencia permite explicar la variabilidad entre los pares, y por lo general arroja un término de error más pequeño y, de esta manera, se aumenta la sensibilidad de la prueba de hipótesis o intervalo de confianza.

Como ejemplos típicos de datos pareados figuran las mediciones hechas en gemelos o mediciones del tipo "antes y después". Como en nuestro caso variables observadas "antes y después". Para una prueba t relacionada:

$$H_0: m_d = m_0 \quad \text{versus} \quad H_1: m_d \neq m_0$$

Donde m_d es la media de la población de las diferencias y m_0 es la media hipotética de las diferencias.

Cuando las muestras se extraen de manera independiente de dos poblaciones, se utiliza la prueba t de dos muestras.

2.2.1 COMPARACIÓN DE DOS MEDIAS CON MUESTRAS RELACIONADAS DE VARIABLES ATMOSFÉRICAS.

Se realiza el test de comparación de dos medias con muestras relacionadas esto es, de los datos tomados el 27 de mayo de 2009 a las 08H30 am con los datos tomados el mismo día pero a las 13H30 (ver cuadro 2.1 Datos de Radio sondeos Proporcionados por el INAMHI).

Al realizar la comparación de la Prueba T para muestras relacionadas de estos datos a la altura de 18 a 20 Km de altura tomados en el mismo día se obtiene los siguientes resultados:

Estadísticos de muestras relacionadas

	Media	N	Desviación típ.	Error típ. de la media
Par 1 PRESIÓN_1	62,654930	142	3,1421314	,2636818
PRESIÓN_2	62,225352	142	3,1157217	,2614656

Correlaciones de muestras relacionadas

	N	Correlación	Sig.
Par 1 PRESIÓN_1 y PRESION_2	142	,987	,000

Cuadro 2.8 Muestra relacionadas presión atmosférica

Estadísticos de muestras relacionadas

	Media	N	Desviación típ.	Error típ. de la media
Par 1 TEMPERAT_A	-72,901408	142	1,3385175	,1123259
TEMPERAT_A_2	-72,394366	142	1,2657952	,1062232

Correlaciones de muestras relacionadas

	N	Correlación	Sig.
Par 1 TEMPERAT_A y TEMPERAT_A_2	142	-,228	,006

Cuadro 2.10 Muestras relacionadas temperatura

Estadísticos de muestras relacionadas

	Media	N	Desviación típ.	Error típ. de la media
Par 1 HUMEDAD	3,00	142	,000	,000
HUMEDAD_2	5,01	142	,266	,022

Cuadro 2.12 Muestras relacionadas humedad relativa

En el cuadro 2.8 aparecen los resultados de los estadísticos para cada una de las variables. Las medias muestrales son $\bar{x} = 62,65$ y $\bar{x} = 62,22$; las cuasidesviaciones

típicas muestrales son $S_1 = 3,14$ y $S_2 = 3,12$; los errores típicos de las medias muestrales son $S_1 / \sqrt{n} = 0,26$ y $S_2 / \sqrt{n} = 0,26$, respectivamente.

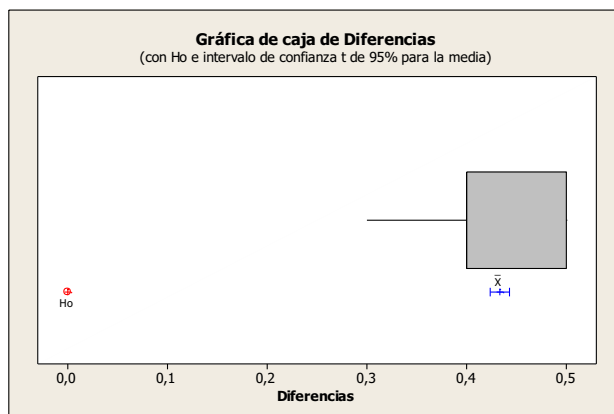


Gráfico 2.4 Caja de Diferencias presión

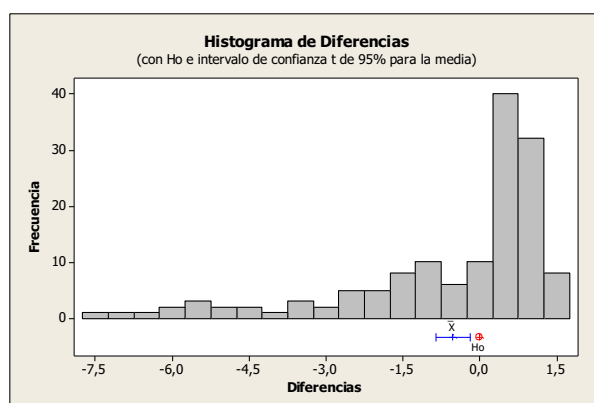


Gráfico 2.5 Histograma de diferencias temperatura



Gráfica 2.6 Histograma de diferencias humedad relativa

Prueba de muestras relacionadas

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación tip.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior				Superior
Par 1	PRESIÓN_1 - PRESIÓN_2	,4295775	,4967681	,0416879	,3471634	,5119915	10,305	141	,000

Cuadro 2.9 Resultados del contraste sobre comparación de dos medias con muestras relacionadas.

Prueba de muestras relacionadas

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación tip.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior				Superior
Par 1	TEMPERAT_A - TEMPERAT_A_2	-,5070423	2,0412292	,1712962	-,8456830	-,1684015	-2,960	141	,004

Cuadro 2.11 Resultados del contraste sobre comparación de dos medias con muestras relacionadas.

Prueba de muestras relacionadas

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación tip.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior				Superior
Par 1	HUMEDAD - HUMEDAD_2	-2,014	,266	,022	-2,058	-1,970	-90,249	141	,000

Cuadro 2.13 Resultados del contraste sobre comparación de dos medias con muestras relacionadas.

El gráfico 2.4 muestra el rechazo de la hipótesis nula, el punto rojo es la media hipotética de la presión atmosférica y este punto se encuentra fuera del intervalo de confianza de la media de la presión atmosférica de 18 a 20 Km.

En la tercera tabla aparecen los siguientes resultados: la media de la variable diferencia $D = X_1 - X_2$ es igual a:

$$\text{Media} = \bar{d} = -0,43,$$

La cuasidesviación típica de la variable diferencia D es igual a:

$$\text{Desviación tip.} = S_d = 0,496,$$

La desviación típica del estadístico \bar{d} es igual a:

$$\text{Error t\acute{p}.} = S_d / \sqrt{n} = 0,48$$

Un intervalo de confianza para la diferencia de las medias poblacionales es:

$$0,35 < \mu_1 - \mu_2 < 0,52$$

El estadístico de contraste para la comparación de las dos medias poblacionales (con muestras relacionadas) es:

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = 10,305,$$

Los grados de libertad son:

$$gl=141$$

y el nivel crítico es:

$$\text{Sig. (bilateral)}=0,263$$

2.2.2 CONCLUSIONES

1. En el cuadro 2.8 aparecen los resultados de la correlación entre dos variables. El valor del coeficiente de correlación lineal de Pearson en la muestra es 0,987. También, aparece en esta tabla el resultado del nivel crítico (Sig.) que es igual a 0, menor que 0.05, por lo que se rechaza la hipótesis nula. Por tanto, se rechaza que las dos variables sean independientes en el proceso de obtención de datos de la presión atmosférica.
2. De los resultados obtenidos en el cuadro 2.11 y realizando un análisis exhaustivo similar al de la presión atmosférica se concluye que se rechaza la hipótesis nula es decir, se rechaza que las dos variables sean independientes en el proceso de obtención de datos de la temperatura.
3. De los resultados obtenidos en el cuadro 2.13 y realizando un análisis exhaustivo similar al de la presión atmosférica se concluye que se rechaza

la hipótesis nula es decir, se rechaza que las dos variables sean independientes en el proceso de obtención de datos de la humedad relativa.

2.3 PRUEBA T PARA MUESTRAS INDEPENDIENTES ANOVA DE UN FACTOR DE VARIABLES ATMOSFÉRICAS.

Un problema frecuente que se presenta es comparar la media de varios procesos o dos tratamientos. Por ejemplo, comparar los datos atmosféricos de un día con los datos obtenidos de otros días para medir si el comportamiento en promedio de estas variables son similares a la altura de 18 a 20 Km en la ciudad de Ambato a las 8H00 am. Para ello se decide hacer un estudio que permita comparar las medias y las varianzas reportadas por los diferentes días.

Para comparar las medias se plantea la hipótesis de igualdad de medias con la alternativa bilateral, puesto que no hay ninguna conjetura del experimentador acerca de qué día se podrá obtener valores altos de estas variables. Luego, el planteamiento es:

$$H_o : \mu_1 = \mu_2 = \dots \mu_k = \mu$$

$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j$$

La cual se desea probar con un nivel de significancia de 5% ($\alpha=0.05$). El estadístico de prueba adecuado para probar la hipótesis de igualdad de media está dado por:

$$t_0 = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

El cuál sigue una distribución t de Student con $n_x + n_y - 2$ grados de libertad, donde S_p^2 es un estimador de la varianza muestral común, suponiendo que dichas varianzas desconocidas sean iguales, y se calcula como

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

Con S_x^2 y S_y^2 las varianzas muestrales de los datos de cada proceso.

Se realiza el cálculo del ANOVA para las variables atmosféricas y se obtienen los resultados del cuadro 2.14.

Utilizamos el método de James-Howell que “Se basa en la distribución del rango estudiantizado y en un estadístico T en el que, tras estimar las varianzas poblacionales suponiendo que son distintas, se corrigen los grados de libertad”.

ANOVA						
		Suma de cuadrados	gl	Media cuadrática	F	Sig.
PRESIÓN (hPa)	Inter-grupos	2,124E7	3	7079605,499	173,395	,000
	Intra-grupos	6,387E8	15643	40829,389		
	Total	6,599E8	15646			
TEMPERATURA (°C)	Inter-grupos	397470,088	3	132490,029	164,890	,000
	Intra-grupos	1,257E7	15643	803,503		
	Total	1,297E7	15646			
HUMEDAD RELATIVA (%)	Inter-grupos	410173,503	3	136724,501	142,342	,000
	Intra-grupos	1,503E7	15643	960,532		
	Total	1,544E7	15646			

Cuadro 2.14 ANOVA de Variables Atmosféricas

En el cuadro 2.15 se muestra la prueba de homogeneidad de la varianza con el valor del estadístico para cada variable atmosférica y el grado de significación de la prueba de varianzas iguales al 5% se acepta la hipótesis nula y se concluye que las varianzas de los datos en los diferentes días no son iguales.

Prueba de homogeneidad de varianzas				
	Estadístico de Levene	gl1	gl2	Sig.
PRESIÓN (hPa)	23,933	3	15643	,000
TEMPERATURA (°C)	258,056	3	15643	,000
HUMEDAD RELATIVA (%)	302,422	3	15643	,000

Cuadro 2.15 Prueba de homogeneidad de la varianza

Games-Howell

Variable dependiente	(I) RADIO SONDAS LANZADAS EN AMBATO	(J) RADIO SONDAS LANZADAS EN AMBATO	Diferencia de medias (I-J)	Error típico	Sig.	Intervalo de confianza al 95%	
						Límite inferior	Límite superior
PRESIÓN (hPa)	27 DE MAYO 2009	27 DE MAYO 2009 (2)	-7,08360	4,24808	,341	-17,9993	3,8321
		21 DE NOVIEMBRE 2008	-39,24616*	4,47772	,000	-50,7520	-27,7403
		25 DE NOVIEMBRE 2008	-111,46970*	5,28192	,000	-125,0441	-97,8953
	27 DE MAYO 2009 (2)	27 DE MAYO 2009	7,08360	4,24808	,341	-3,8321	17,9993
		21 DE NOVIEMBRE 2008	-32,16256*	4,19437	,000	-42,9400	-21,3852
		25 DE NOVIEMBRE 2008	-104,38610*	5,04394	,000	-117,3492	-91,4230
	21 DE NOVIEMBRE 2008	27 DE MAYO 2009	39,24616*	4,47772	,000	27,7403	50,7520
		27 DE MAYO 2009 (2)	32,16256*	4,19437	,000	21,3852	42,9400
		25 DE NOVIEMBRE 2008	-72,22354*	5,23881	,000	-85,6871	-58,7600
	25 DE NOVIEMBRE 2008	27 DE MAYO 2009	111,46970*	5,28192	,000	97,8953	125,0441
		27 DE MAYO 2009 (2)	104,38610*	5,04394	,000	91,4230	117,3492
		21 DE NOVIEMBRE 2008	72,22354*	5,23881	,000	58,7600	85,6871
TEMPERATURA (°C)	27 DE MAYO 2009	27 DE MAYO 2009 (2)	-3,41412*	,56207	,000	-4,8584	-1,9698
		21 DE NOVIEMBRE 2008	-6,25570*	,61124	,000	-7,8263	-4,6851
		25 DE NOVIEMBRE 2008	-16,36613*	,84422	,000	-18,5359	-14,1963
	27 DE MAYO 2009 (2)	27 DE MAYO 2009	3,41412*	,56207	,000	1,9698	4,8584
		21 DE NOVIEMBRE 2008	-2,84158*	,56803	,000	-4,3012	-1,3820
		25 DE NOVIEMBRE 2008	-12,95201*	,81348	,000	-15,0429	-10,8611
	21 DE NOVIEMBRE 2008	27 DE MAYO 2009	6,25570*	,61124	,000	4,6851	7,8263
		27 DE MAYO 2009 (2)	2,84158*	,56803	,000	1,3820	4,3012
		25 DE NOVIEMBRE 2008	-10,11042*	,84820	,000	-12,2904	-7,9304
	25 DE NOVIEMBRE 2008	27 DE MAYO 2009	16,36613*	,84422	,000	14,1963	18,5359
		27 DE MAYO 2009 (2)	12,95201*	,81348	,000	10,8611	15,0429
		21 DE NOVIEMBRE 2008	10,11042*	,84820	,000	7,9304	12,2904
HUMEDAD RELATIVA (%)	27 DE MAYO 2009	27 DE MAYO 2009 (2)	3,22231*	,72379	,000	1,3624	5,0822
		21 DE NOVIEMBRE 2008	13,30626*	,72191	,000	11,4512	15,1613
		25 DE NOVIEMBRE 2008	4,04061*	,84621	,000	1,8660	6,2152
	27 DE MAYO 2009 (2)	27 DE MAYO 2009	-3,22231*	,72379	,000	-5,0822	-1,3624
		21 DE NOVIEMBRE 2008	10,08395*	,59755	,000	8,5486	11,6193
		25 DE NOVIEMBRE 2008	,81829	,74297	,689	-1,0911	2,7277
	21 DE NOVIEMBRE 2008	27 DE MAYO 2009	-13,30626*	,72191	,000	-15,1613	-11,4512
		27 DE MAYO 2009 (2)	-10,08395*	,59755	,000	-11,6193	-8,5486
		25 DE NOVIEMBRE 2008	-9,26566*	,74113	,000	-11,1704	-7,3609
	25 DE NOVIEMBRE 2008	27 DE MAYO 2009	-4,04061*	,84621	,000	-6,2152	-1,8660
		27 DE MAYO 2009 (2)	-,81829	,74297	,689	-2,7277	1,0911
		21 DE NOVIEMBRE 2008	9,26566*	,74113	,000	7,3609	11,1704

* La diferencia de medias es significativa al nivel 0.05.

Cuadro 2.16 Comparaciones Múltiples de Datos Atmosféricos

2.3.1 CONCLUSIONES

1. Se observa el valor de significancia sig.=0 (cuadro 2.14) por lo que los datos de las variables atmosféricas presenta alta variabilidad en ellos por lo que no se puede trabajar con varianzas iguales en el proceso de comparar los promedios de las variables en los diferentes días.
2. En nuestro caso, el valor de sig=0 (cuadro 2.15) se rechaza la hipótesis nula y se concluye que no se puede comparar las varianzas de los datos atmosféricos en los diferentes días.

3. Se realizan las comparaciones de los datos de los radio sondeos y aplicando la metodología de Games Howell que permite trabajar, aunque las varianzas no son iguales y se obtiene los resultados del cuadro 2.16 en las que se visualiza que el valor $p=0,341$ mayor que, $\alpha =0,05$ se acepta la hipótesis nula para comparar la medias y se concluye que los datos obtenidos para la variable presión atmosférica en promedio para la ciudad de Ambato a la altura de 18 a 20 Km son iguales. De igual forma, se obtiene en el cuadro 2.15 otro valor p mayor al $\alpha =0,05$ en la que, interpreto de la misma manera pero para la variable humedad relativa.
4. En el análisis ANOVA de un factor fue indispensable ya que nos permitirá obtener un modelo ajustado de predicción para cada variable atmosférica con respecto a la altura considerando los datos más relevantes para obtener el mejor ajuste posible.

2.4 TÉCNICAS DE ESTIMACIÓN CURVILÍNEAS DE LAS VARIABLES ATMOSFÉRICAS

Si queremos implementar un modelo estadístico matemático que simulen las condiciones atmosféricas específicas del lugar destinado para el lanzamiento del PGA se realiza una regresión con términos lineales, polinomiales (de segundo o tercer orden), exponenciales y logarítmicos según el comportamiento de cada una de las variables atmosféricas con un alcance vertical de hasta 20 km de una sola variable predictiva a través de los datos, en la escala real o logarítmica. La regresión polinomial es un método para modelar la curvatura en la relación entre una variable de respuesta (Y) y una variable predictiva (X) mediante la ampliación del modelo de regresión lineal, para que incluya a X^2 y X^3 como predictores.

En el paquete SPSS y MINITAB trabajamos con los métodos de ajuste analítico que es la regresión curvilínea y el objetivo es buscar el mejor ajuste lineal, polinómico, exponencial o logarítmico que explique la evolución de cada una de las variables atmosféricas con respecto a la altura.

2.4.1 ANÁLISIS DE TENDENCIA PARA LA VARIABLE PRESIÓN ATMOSFÉRICA.

El comportamiento de esta variable atmosférica es similar en todos los días, gráfico 2.7 y 2.8 a una altura de 20 Km.

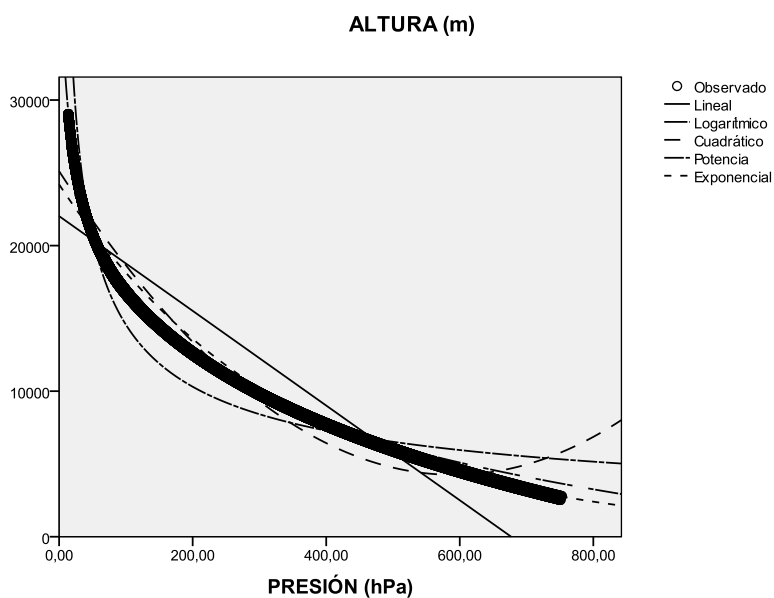


Gráfico 2.7 Presión Atmosférica vs Altura

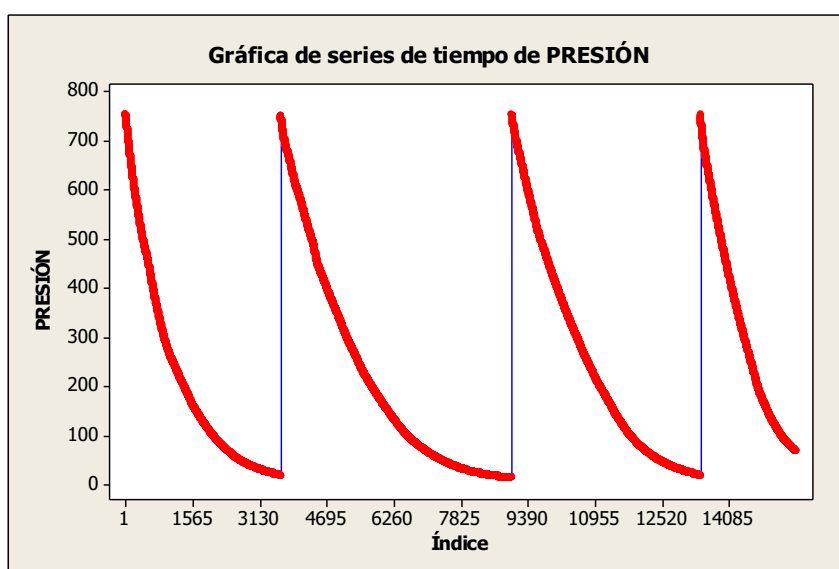


Gráfico 2.8 Comportamiento de la Presión Atmosférica en el tiempo

El gráfico 2.7 muestra el comportamiento de la variable presión atmosférica y los ajuste obtenidos utilizando las funciones lineales, polinomiales y logarítmicas. El

cuadro 2.17 muestra análisis estadístico de cada uno de las estimaciones curvilíneas.

Resumen del modelo y estimaciones de los parámetros

Variable dependiente: ALTURA (m)

Ecuación	Resumen del modelo					Estimaciones de los parámetros		
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2
Lineal	,855	92276,248	1	15645	,000	22024,106	-32,532	
Logarítmica	,997	5342080,783	1	15645	,000	45942,747	-6385,494	
Cuadrático	,954	162403,951	2	15644	,000	25111,497	-70,577	,060
Potencia	,893	130804,269	1	15645	,000	145258,353	-,499	
Exponencial	,984	963979,380	1	15645	,000	24204,354	-,003	

La variable independiente es PRESIÓN (hPa).

Cuadro 2.17 Resumen del modelo y estimación curvilínea de la presión atmosférica

Del cuadro 2.17 se observa que el mejor R^2 está en las estimaciones logarítmica y exponencial 0,997 y 0,984 respectivamente, que indican que el 99% aproximadamente de los datos son explicados por los dos modelos. Sin embargo, uno de los dos modelos debe ser elegido, para ellos realizamos un análisis ANOVA para determinar la curva de mejor ajuste. Este se muestra en el cuadro 2.18 para la curva logarítmica y el cuadro 2.19 para la curva exponencial.

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	8,145E11	1	8,145E11	5342080,783	,000
Residual	2,385E9	15645	152461,275		
Total	8,168E11	15646			

La variable independiente es PRESIÓN (hPa).

Coefficientes

	Coefficients no estandarizados		Coefficients estandarizados	t	Sig.
	B	Error típico			
ln(PRESIÓN (hPa))	-6385,494	2,763	-,999	-2311,294	,000
(Constante)	45942,747	13,957		3291,640	,000

Cuadro 2.18 Análisis ANOVA y de Coeficientes para la curva logarítmica

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	5487,728	1	5487,728	963979,380	,000
Residual	89,064	15645	,006		
Total	5576,792	15646			

La variable independiente es PRESIÓN (hPa).

Coeficientes					
	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Error típico	Beta		
PRESIÓN (hPa)	-,003	,000	-,992	-981,825	,000
(Constante)	24204,354	21,988		1100,806	,000

La variable dependiente es ln(ALTURA (m)).

Cuadro 2.19 Análisis ANOVA y de coeficientes para la curva exponencial

De los cuadros anteriores se observa que el valor de sig = 0 que indican que los parámetros estimados son significativos y para el caso del valor de sig =0 del ANOVA nos indica que la varianza de los diferentes datos son similares. Por lo que, elegimos la regresión curvilínea logarítmica. Entonces, el modelo que explica el 99,7% de los datos de la presión atmosférica con respecto a la altura es:

$$\hat{y} = 45.942,47 - 6.385,494 \ln(\text{presión}(hPa))$$

2.4.2 ANÁLISIS DE TENDENCIA PARA LA VARIABLE TEMPERATURA

El comportamiento de esta variable atmosférica desciende hasta la altura de 17 Km y de ahí en adelante comienza ascender en todos los días gráfico 2.9.

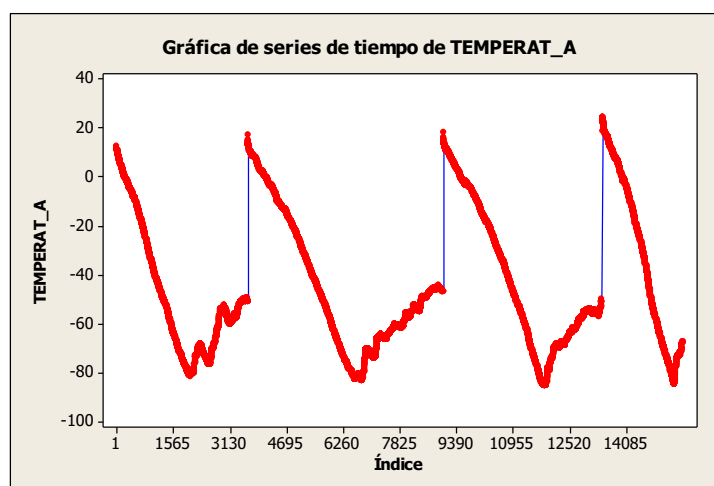


Gráfico 2.9 Comportamiento de la Temperatura en el tiempo

Por lo que, para ajustar de mejor manera los datos, construimos dos modelos para simular el comportamiento de la temperatura, el primer modelo desde el lugar de lanzamiento es decir, desde los 2 Km hasta los 17 Km y el segundo desde los 17 Km hasta los 20 Km.

El gráfico 2.10 corresponde al comportamiento de la temperatura en (°C) desde los 2 Km hasta los 17 Km. La temperatura no es fija como la presión atmosférica sino que presenta algunas variaciones en los diferentes datos obtenidos.

El mejor ajuste para la temperatura de 2 Km a 17 Km es la estimación lineal que puede explicar el 98,5% de los datos. Cuadro 2.20.

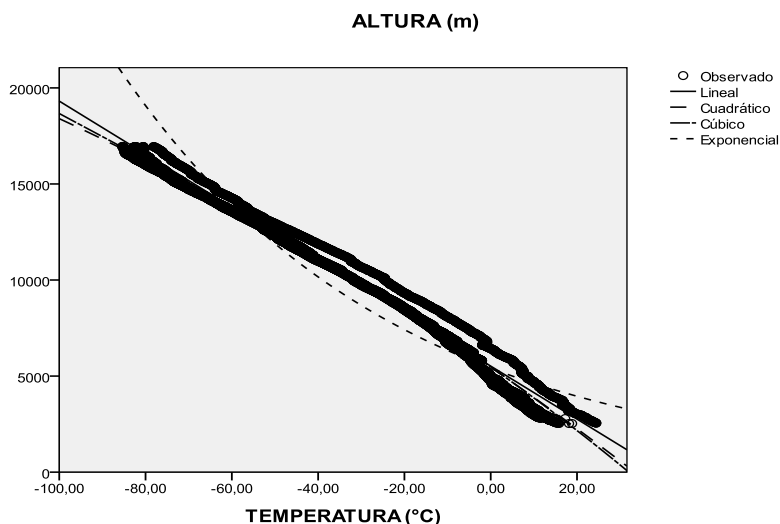


Gráfico 2.10 Temperatura vs Altura

Resumen del modelo

R	R cuadrado	R cuadrado corregida	Error típico de la estimación
,993	,985	,985	497,871

La variable independiente es TEMPERATURA (°C).

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	1,601E11	1	1,601E11	645769,728	,000
Residual	2,413E9	9733	247875,915		
Total	1,625E11	9734			

La variable independiente es TEMPERATURA (°C).

Coefficientes

	Coefficients no estandarizados		Coefficients estandarizados	t	Sig.
	B	Error típico	Beta		
TEMPERATURA (°C)	-137,929	,172	-,993	-803,598	,000
(Constante)	5519,358	7,283		757,832	,000

Cuadro 2.20 Resumen del Modelo

El modelo es:

$$\hat{y} = 5.519.36 - 137.929(\text{temperatura}(^{\circ}\text{C}))$$

Ahora realizamos el análisis desde los 17 Km hasta los 20Km de la temperatura el cual presenta un comportamiento diferente de ascenso en los diferentes días.

Gráfico 2.11.

El mejor ajuste para la temperatura de 17 Km a 20 Km es la estimación exponencial y cuadrática que pueden explicar el 90,6% y 91,5% de los datos. Cuadro 2.22.

Sin embargo, uno de los dos modelos debe ser elegido, para ellos realizamos un análisis ANOVA para determinar la curva de mejor ajuste. Este se muestra en el cuadro 2.20 para la curva exponencial y el cuadro 2.21 para la curva cuadrática.

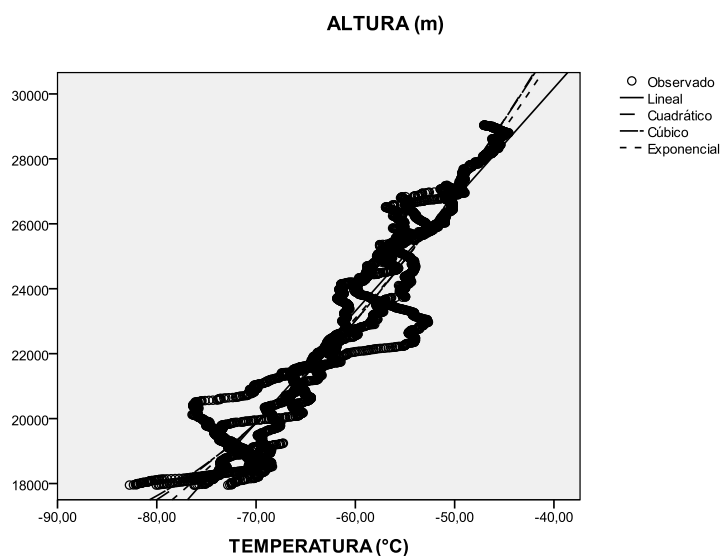


Gráfico 2.11 Temperatura vs Altura

Resumen del modelo y estimaciones de los parámetros

Variable dependiente: ALTURA (m)

Ecuación	Resumen del modelo					Estimaciones de los parámetros			
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2	b3
Lineal	,905	50089,718	1	5237	,000	43930,283	343,678		
Logarítmica ^a	,000	,000		
Cuadrático	,915	28050,917	2	5236	,000	58380,953	824,727	3,922	
Cúbico	,915	28319,072	2	5236	,000	54033,528	596,706	,000	-,022
Exponencial	,906	50687,793	1	5237	,000	56907,423	,015		

La variable independiente es TEMPERATURA (°C).

a. La variable independiente (TEMPERATURA (°C)) contiene valores no positivos. El valor mínimo es -82,75. No es posible calcular el modelo logarítmico ni el de potencia.

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	90,524	1	90,524	50687,793	,000
Residual	9,353	5237	,002		
Total	99,876	5238			

La variable independiente es TEMPERATURA (°C).

Coeficientes

	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Error típico	Beta		
TEMPERATURA (°C)	,015	,000	,952	225,139	,000
(Constante)	56907,423	234,125		243,065	,000

La variable dependiente es ln(ALTURA (m)).

Cuadro 2.21 Análisis ANOVA y de coeficientes para la curva exponencial

ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	4,781E10	2	2,390E10	28050,917	,000
Residual	4,462E9	5236	852166,576		
Total	5,227E10	5238			

La variable independiente es TEMPERATURA (°C).

Coeficientes

	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
	B	Error típico	Beta		
TEMPERATURA (°C)	824,727	20,202	2,283	40,824	,000
TEMPERATURA (°C)** 2	3,922	,164	1,335	23,874	,000
(Constante)	58380,953	611,914		95,407	,000

Cuadro 2.22 Análisis ANOVA y de coeficientes para la curva cuadrática

De los cuadros anteriores se observa que el valor de sig = 0 que indican que los parámetros estimados son significativos y para el caso del valor de sig = 0 del ANOVA nos indica que la varianza de los diferentes datos son similares. Por lo que elegimos la regresión curvilínea cuadrática entonces, el modelo que explica el 91,5% de los datos de la temperatura con respecto a la altura de 17Km a 20 Km es:

$$\hat{y} = 58.380,95 + 824,727(\text{temperatura}(^{\circ}\text{C})) + 3,922(\text{temperatura}(^{\circ}\text{C}))^2$$

2.4.3 ANÁLISIS DE TENDENCIA PARA LA VARIABLE HUMEDAD RELATIVA

El comportamiento de esta variable atmosférica no es similar en todos los días. Gráfico 2.12 y 2.13. Al tratar de buscar una curva que mejor ajuste el comportamiento de esta variable no sería tan apropiado debido a que el comportamiento es caótico y de igual forma el coeficiente de determinación R^2 no explica la mayor cantidad de datos como se observa en el cuadro 2.23.

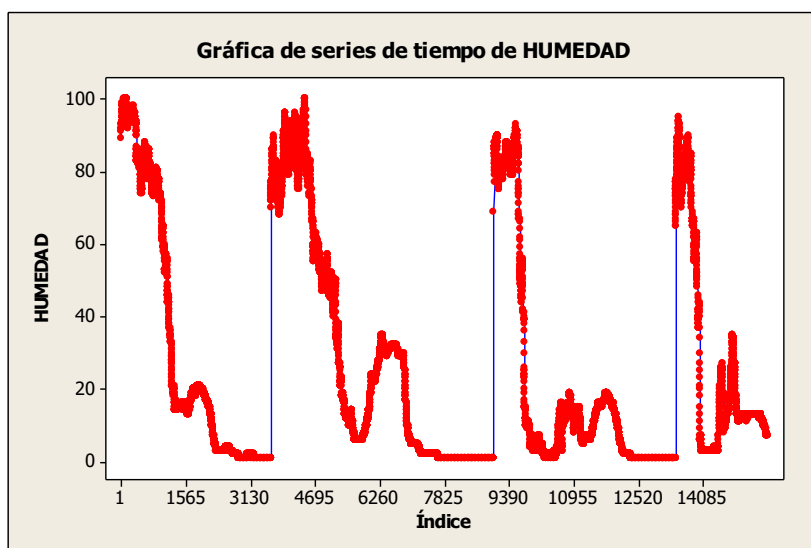


Gráfico 2.12 Comportamiento de la Presión Atmosférica en el tiempo

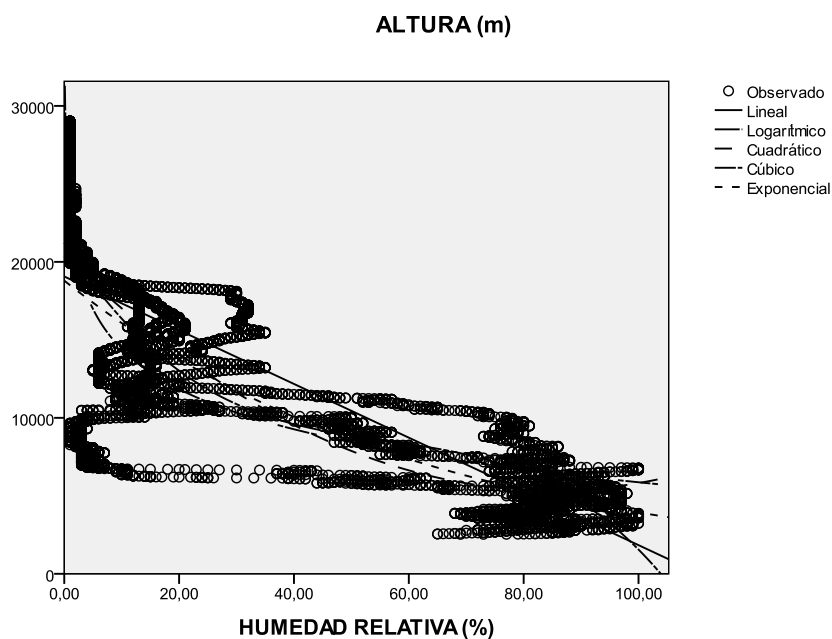


Gráfico 2.13 Presión Atmosférica vs Altura

Resumen del modelo y estimaciones de los parámetros

Variable dependiente: ALTURA (m)

Ecuación	Resumen del modelo					Estimaciones de los parámetros			
	R cuadrado	F	gl1	gl2	Sig.	Constante	b1	b2	b3
Lineal	,561	19969,028	1	15645	,000	19064,138	-172,256		
Logarítmica	,679	33141,069	1	15645	,000	22953,401	-3709,665		
Cuadrático	,595	11485,120	2	15644	,000	20352,807	-353,643	2,083	
Cúbico	,613	8242,053	3	15643	,000	21232,783	-574,510	9,303	-,055
Exponencial	,676	32707,428	1	15645	,000	18799,004	-,016		

La variable independiente es HUMEDAD RELATIVA (%).

Cuadro 2.23 Resumen del modelo regresivo para la Humedad Relativa

Por lo que, podemos aplicar las técnicas autorregresivas ya que, al calcular el estadístico de Durbin Watson³. En el cuadro 2.24 se muestra el valor del estadístico para la serie de datos de la Humedad Relativa y se comprueba que tiene que utilizarse los métodos autorregresivos para buscar un modelo que aproxime el comportamiento de los datos utilizando los modelos ARIMA.

Resumen del modelo^b

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,749 ^a	,561	,561	4789,161	,002

a. Variables predictoras: (Constante), HUMEDAD RELATIVA (%)

b. Variable dependiente: ALTURA (m)

Cuadro 2.24 Estadístico de Durbin-Watson.

2.5 MODELOS ARIMA (p,d,q)

2.5.1 INTRODUCCIÓN

Un modelo ARIMA⁴ que sus siglas en inglés son: *AutoRegresive Integrated Moving Average*, es un modelo estadístico autopredictivo que permite predecir los valores de una variable en función de sus valores pasados sin necesidad de ninguna otra información de variables auxiliares o relacionadas. Las observaciones en un momento dado es modelada en función de los valores anteriores. El modelo ARIMA presenta una ecuación explícita que permite

³ Indica que los datos de una serie no son modelables regresivamente si el valor de este estadístico que se basa en los residuos es menor a 1.5.

⁴ Box y Jenkins son los autores de la modelización ARIMA. Autoregresivo (AR), Integrado (I) y Medias Móviles (MA).

describir un valor como una función lineal de datos anteriores y errores debido al azar. Puede incluir, además, un componente cíclico o estacional. El objetivo consiste en obtener un modelo adecuado, el modelo ARIMA debe contener todos los elementos necesarios, pero los mínimos necesarios para describir el fenómeno en estudio. Según César Pérez en *Econometría de las series temporales* menciona: “Box y Jenkins recomienda como mínimo unas 50 observaciones en la serie temporal. Modelizar una serie temporal consiste en derivar un modelo ARIMA que se ajuste al conjunto de datos dado que pueden utilizarse a las variables atmosféricas, sin embargo, solo utilizamos esta metodología para encontrar un modelo para la variable humedad relativa.

2.5.2 MODELOS ARIMA (p,d,q)

Un modelo $ARIMA(0,d,0)$ es una serie temporal que se convierte en un ruido blanco⁵ después de ser diferenciada d veces. El modelo $ARIMA(0,d,0)$ se expresa mediante: $(1-B)dX_t = a_t$. El modelo general $ARIMA(p,d,q)$ denominado proceso autorregresivo integrado de medias móviles de orden p,d,q , toma la siguiente expresión:

$$(1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p)(1 - B)^d Y_t = (1 - \nu_1 B - \nu_2 B^2 - \dots - \nu_q B^q) a_t$$

Un modelo $ARIMA(p,d,q)$ permite describir una serie de observaciones después de que hayan sido diferenciadas d veces, a fin de extraer las posibles fuentes de no estacionariedad. Si hay alguna componente p,d,q igual a cero, se elimina el término correspondiente de la fórmula general.

En la figura 2.14 que se obtiene de “*Econometría de las Series Temporales*, Cesar Pérez Año, 2006” esquematiza las fases de la metodología de Box Jenkins en modelos ARIMA.

2.5.4 APLICACIÓN DEL MODELO ARIMA A LA VARIABLE HUMEDAD RELATIVA.

⁵ Es un proceso puramente aleatorio.

Aplicamos la metodología de Box_Jenkins para los modelos ARIMA y utilizamos el paquete estadístico STATGRAPHICS para los cálculos que permite trabajar y realizar predicciones con series temporales a partir de la metodología de Box-Jenkins. Ajustaremos a un modelo ARIMA adecuado una serie de 15647 datos correspondiente a la humedad relativa y prediciremos el comportamiento de esta variable a la altura de 18 a 20 Km de altura para próximos días.

Realizamos una presentación gráfica de los datos, gráfico 2.14, para observar la estacionalidad.

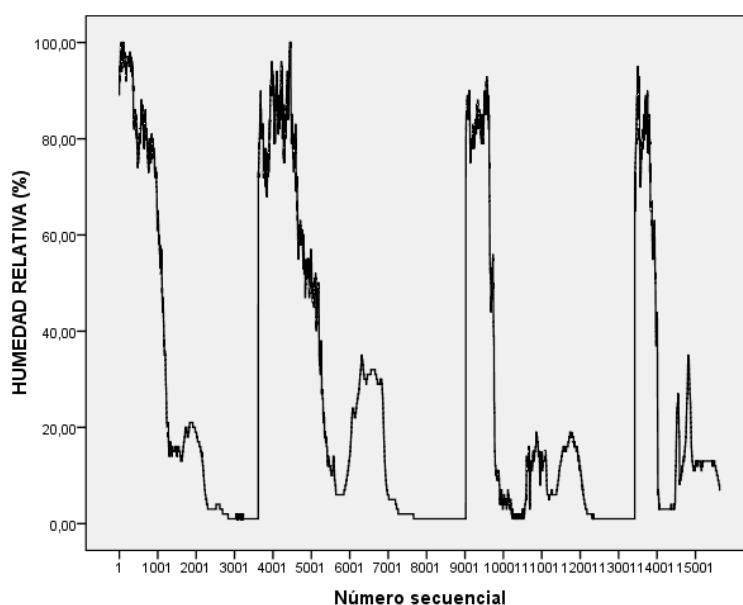


Gráfico 2.14 Comportamiento de la Humedad Relativa en el tiempo

La estacionalidad, así como la estacionariedad también puede detectarse a través de las funciones de autocorrelación y autocorrelación parcial estimadas (FAC y FAC parcial respectivamente) Gráfico 2.16.

Se observa que los coeficientes de la FAC no decaen rápidamente, lo que indica falta de estacionariedad en media. Por lo tanto diferenciamos la serie original y graficamos sus funciones de autocorrelación y autocorrelación parcial estimadas. Gráfico 2.17.

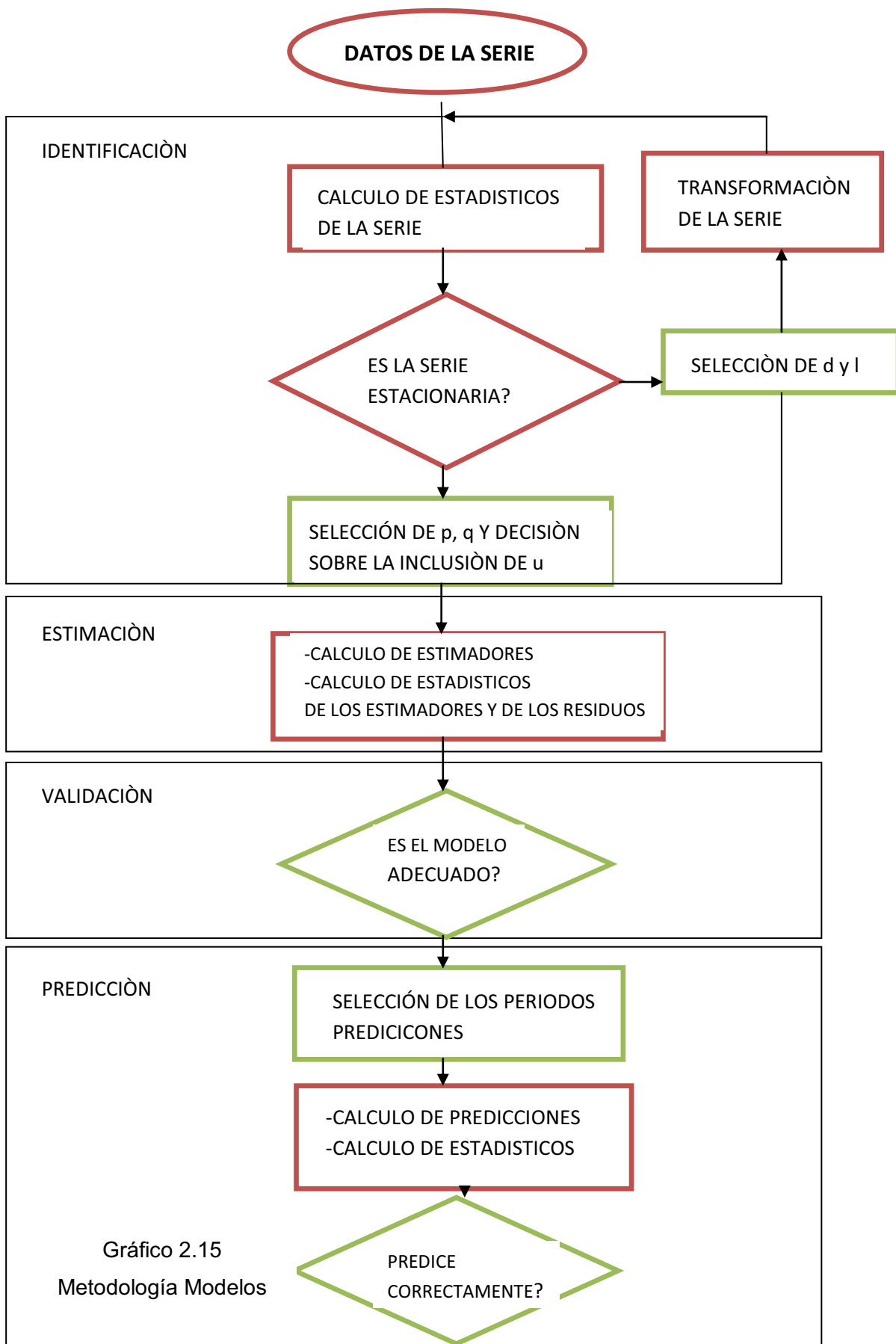


Gráfico 2.15
Metodología Modelos

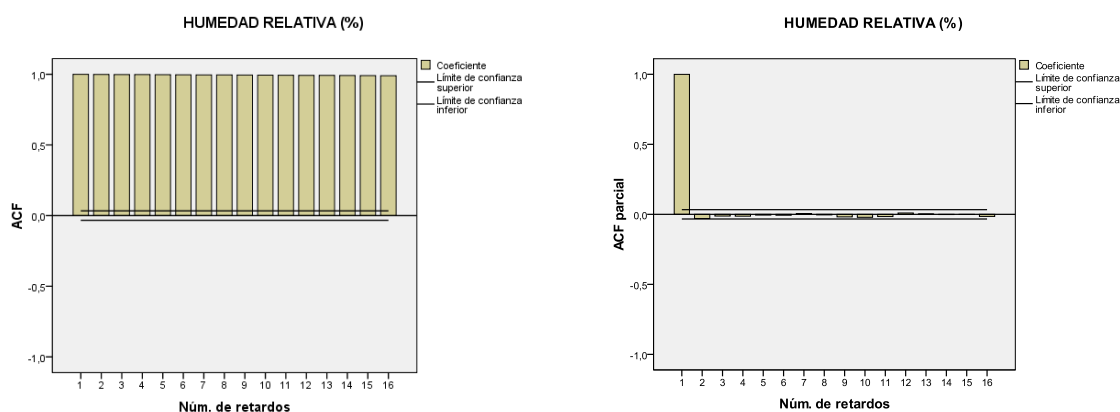


Gráfico 2.16 ACF y ACF parcial para la variable humedad relativa

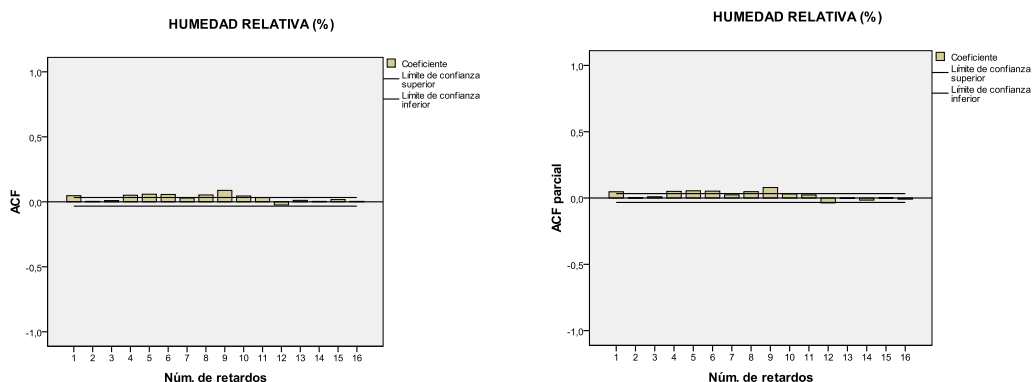


Gráfico 2.17 ACF y ACF parcial diferenciada para la variable humedad relativa

Ahora los retardos significativos de la FAC decaen tan rápidamente, luego, ya no existen problemas de estacionariedad en la serie diferenciada, es decir, la serie diferenciada es $I(0)$ y la serie original es $I(1)$. En cuanto a la identificación de la parte de la media móvil de la serie vemos que 5 retardos se salen de la franja horizontal y que el decrecimiento de los retardos de la FAC parcial no es muy rápido. Luego la parte de media móvil se modelizará como un proceso $MA(0)$ hasta $MA(5)$. Para identificar la parte autorregresiva vemos que, existen 2 cambios significativos de las barras del gráfico de la FAC parcial por lo que, se modelizará como un proceso $AR(1)$ o $AR(2)$. Realizamos los análisis de los modelos para tomar el mejor de ellos, cuadro 2.25.

```

Resumen de Previsión
-----
Diferencia No estacional de orden: 1

Modelo de pronóstico seleccionado: AR
Número de predicciones generadas: 12
Número de periodos retenidos para val

Estadístico      Periodo           Periodo
                  Estimación       Validación
-----
RMSE              1,06757
MAE               0,179507
MAPE              0,996291
ME               -0,00000540778
MPE              -0,0352397

```

```

Resumen de Previsión
-----
Diferencia No estacional de orden: 1

Modelo de pronóstico seleccionado: ARIMA
Número de predicciones generadas: 12
Número de periodos retenidos para valida

Estadístico      Periodo           Periodo
                  Estimación       Validación
-----
RMSE              1,06765
MAE               0,179586
MAPE              0,993394
ME               -0,00000458381
MPE              -0,0385954

```

Cuadro 2.25 Resumen de Análisis Modelos ARIMA

Además, considerando las dos funciones de autocorrelación en conjunto, vemos que sus retardos no se anulan muy bruscamente. Estamos entonces ante una estructura ARMA(0,1) para la serie diferenciada. Se concluye entonces al buscar el mejor modelo que la serie se ajusta a un modelo ARIMA(0,1,1) ya que el MAPE del cuadro 2.25 es el menor de ambos, que significa el porcentaje de error absoluto medio. El cuadro 2.26 se muestra el resumen del mejor modelo obtenido.

```

Resumen del Análisis
Datos: HUMEDAD
Número de observaciones = 15647
Índice de comienzo (empezando en) = 1,0
Intervalo de muestra = 1,0

Resumen de Previsión
-----
Diferencia No estacional de orden: 1

Modelo de pronóstico seleccionado: ARIMA(0,1,1) con constante
Número de predicciones generadas: 12
Número de periodos retenidos para validación: 0

-----
Estadístico      Período      Período
                  Estimación    Validación
-----
RMSE              1,06765
MAE               0,179586
MAPE              0,223394
ME               -0,00000455351
MPE              -0,0355954

Resumen del Modelo ARIMA
-----
Parámetro      Estimación    Error Estd.    t          P-Valor
-----
MA(1)          -0,0523995    0,00795552     -6,58032    0,000000
Media          -0,00523646   0,00599488     -0,88216    0,360436
Constante      -0,00523646
-----

Predicción hacia atrás: xi
Varianza de ruido blanco estimada = 1,13937 con 15644 grados de libertad
Desviación típica del ruido blanco estimado = 1,06765
Número de iteraciones: 1

```

Cuadro 2.26 Salida del Modelo ARIMA(0,0,1)

La ecuación del Modelo ARIMA(0,0,1) estimado será la siguiente:

$$(1 - B)Y_t = (1 + 0.05239)a_t \Rightarrow Y_t - Y_{t-1} = a_t + 0.05239a_{t-1}$$

El gráfico 2.18 muestra el gráfico de la humedad relativa y sus ajuste del modelo obtenido ARIMA(0,0,1).

Los pronósticos para la variable humedad relativa a la altura de 18 a 20 Km se muestran en el cuadro 2.27.

Gráfico de Secuencia de tiempo para HUMEDAD

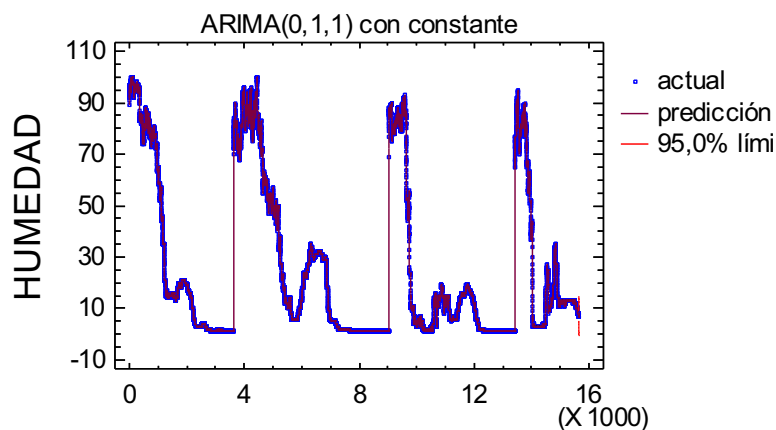


Gráfico 2.18 Humedad Relativa y el Ajuste Obtenido

↓	C1	C2	C3	C4-T	C5	C6	C7	C8	C9
	TIEMPO_S	ALTURA_M	PRESIÓN	TEMPERAT	HUMEDAD	TEMPERAT_1	RESI1	FITS1	
6808	6376	17990	81	191,2	30	-81,95	-0,0011	30,001	
6809	6378	17996	81	191,2	30	-81,95	-0,0003	30,000	
6810	6380	18001	81	191,3	30	-81,85	0,0004	30,000	
6811	6382	18006	81	191,5	30	-81,65	0,0009	29,999	
6812	6384	18011	81	191,5	30	-81,65	0,0014	29,999	
6813	6386	18015	81	191,6	30	-81,55	0,0018	29,998	
6814	6388	18020	81	191,7	30	-81,45	0,0022	29,998	
6815	6390	18025	81	191,7	30	-81,45	0,0025	29,998	
6816	6392	18029	81	191,7	30	-81,45	0,0027	29,997	
6817	6394	18034	80	191,8	30	-81,35	0,0029	29,997	
6818	6396	18039	80	191,9	30	-81,25	0,0031	29,997	
6819	6398	18044	80	192,1	30	-81,05	0,0032	29,997	
6820	6400	18048	80	192,2	30	-80,95	0,0034	29,997	
6821	6402	18052	80	192,3	30	-80,85	0,0035	29,997	
6822	6404	18057	80	192,3	30	-80,85	0,0036	29,996	
6823	6406	18061	80	192,3	30	-80,85	0,0036	29,996	
6824	6408	18066	80	192,4	30	-80,75	0,0037	29,996	
6825	6410	18070	80	192,4	30	-80,75	0,0037	29,996	
6826	6412	18075	80	192,4	30	-80,75	0,0038	29,996	
6827	6414	18081	80	192,5	30	-80,65	0,0038	29,996	
6828	6416	18085	80	192,5	30	-80,65	0,0039	29,996	
6829	6418	18089	80	192,6	30	-80,55	0,0039	29,996	
6830	6420	18094	80	192,7	30	-80,45	0,0039	29,996	
6831	6422	18098	80	192,9	30	-80,25	0,0039	29,996	
6832	6424	18103	79	193	30	-80,15	0,0039	29,996	
6833	6426	18108	79	193,1	30	-80,05	0,0040	29,996	
6834	6428	18112	79	193,2	30	-79,95	0,0040	29,996	
6835	6430	18117	79	193,3	29	-79,85	-0,9960	29,996	
6836	6432	18122	79	193,4	29	-79,75	0,0385	28,961	
6837	6434	18126	79	193,4	29	-79,75	0,0331	28,967	
6838	6436	18130	79	193,6	29	-79,55	0,0286	28,971	

Cuadro 2.27 Pronóstico obtenido para la Humedad Relativa

2.6 REGRESIÓN LINEAL CON MÚLTIPLES VARIABLES ATMOSFÉRICAS

A continuación vamos a crear un modelo de regresión con múltiples variables utilizando el método de mínimos cuadrados generales que nos permite: almacenar estadísticas de regresión, examinar diagnósticos de residuos, generar estimados de puntos, generar intervalos de predicción y de confianza y realizar pruebas de falta de ajuste.

Para construir el modelo de regresión con múltiples variables atmosféricas se debe analizar la correlación entre ellas y visualizar si existe una alta correlación entre las variables para poder construir el modelo.

Se calcula el coeficiente de correlación del momento del producto de Pearson entre cada par de variables atmosféricas. Se puede utilizar el coeficiente de correlación del momento del producto de Pearson para medir el grado de relación

lineal entre dos variables. El coeficiente de correlación presupone un valor entre -1 y +1. Si una variable tiende a incrementarse mientras las demás disminuyen, el coeficiente de correlación es negativo. En cambio, si las dos variables tienden a incrementarse al mismo tiempo, el coeficiente de correlación es positivo. Para una prueba de dos colas de la correlación se tiene:

$H_0: r = 0$ versus $H_1: r \neq 0$, donde r es la correlación entre un par de variables.

El cuadro 2.28 muestra el cálculo de los coeficientes de correlación para las variables atmosféricas y se observa la alta correlación entre las variables presión, humedad relativa y temperatura de igual manera al contrastar la prueba de hipótesis de correlación se observa que el valor $p < 0.01$ por lo que, se rechaza la hipótesis nula de que las variables no están correlacionadas por lo que existe una alta correlación significativa entre las variables para un nivel de confianza del 99%. El gráfico 2.19 nos muestra que al realizar un gráfico en componentes principales las tres variables atmosféricas están altamente correlacionadas por lo que podemos construir un modelo de regresión lineal múltiple.

Correlaciones

		PRESIÓN (hPa)	TEMPERATURA (°C)	HUMEDAD RELATIVA (%)	VELOCIDAD_ Y	VELOCIDAD_ X	ALTURA (m)
PRESIÓN (hPa)	Correlación de Pearson	1	,918**	,838**	,081**	-,077**	-,925**
	Sig. (bilateral)		,000	,000	,000	,000	,000
	N	15647	15647	15647	15647	15647	15647
TEMPERATURA (°C)	Correlación de Pearson	,918**	1	,722**	,155**	-,285**	-,759**
	Sig. (bilateral)	,000		,000	,000	,000	,000
	N	15647	15647	15647	15647	15647	15647
HUMEDAD RELATIVA (%)	Correlación de Pearson	,838**	,722**	1	-,072**	-,129**	-,749**
	Sig. (bilateral)	,000	,000		,000	,000	,000
	N	15647	15647	15647	15647	15647	15647
VELOCIDAD_Y	Correlación de Pearson	,081**	,155**	-,072**	1	-,054**	-,005
	Sig. (bilateral)	,000	,000	,000		,000	,523
	N	15647	15647	15647	15647	15647	15647
VELOCIDAD_X	Correlación de Pearson	-,077**	-,285**	-,129**	-,054**	1	-,101**
	Sig. (bilateral)	,000	,000	,000	,000		,000
	N	15647	15647	15647	15647	15647	15647
ALTURA (m)	Correlación de Pearson	-,925**	-,759**	-,749**	-,005	-,101**	1
	Sig. (bilateral)	,000	,000	,000	,523	,000	
	N	15647	15647	15647	15647	15647	15647

** La correlación es significativa al nivel 0,01 (bilateral).

Cuadro 2.28 Variables atmosféricas y sus correlaciones

Gráfico de componentes

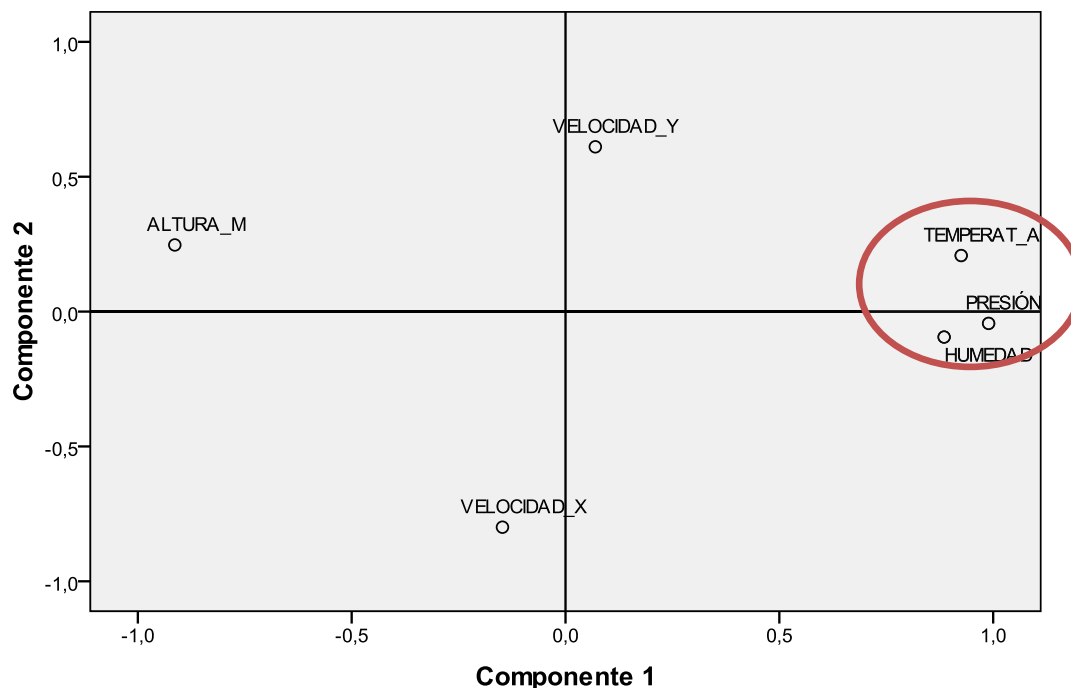


Gráfico 2.19 Análisis en componentes principales

Coeficientes^a

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
1 (Constante)	33694,352	112,493		299,524	,000
PRESIÓN (hPa)	-58,491	,266	-1,663	-219,757	,000
TEMPERATURA (°C)	158,659	1,498	,632	105,886	,000
HUMEDAD RELATIVA (%)	43,184	,996	,188	43,367	,000

a. Variable dependiente: ALTURA (m)

Cuadro 2.29 Modelo de Regresión lineal múltiple

$$\hat{y} = 33.694.352 - 58,491(\text{presión}(hPa)) + 158,659(\text{temperatura}(^{\circ}C)) + 43,18(\text{humedad}(\%))$$

El cuadro 2.29 muestra los resultados del ajuste a un modelo de regresión lineal múltiple para describir la relación entre ALTURA_M y 3 variables independientes. La ecuación del modelo ajustado es Dado que el p-valor en la tabla ANOVA cuadro 2.30 es inferior a 0,01, existe relación estadísticamente significativa entre las variables para un nivel de confianza del 99%.

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	7,490E11	3	2,497E11	57520,134	,000 ^a
	Residual	6,789E10	15643	4340228,187		
	Total	8,168E11	15646			

a. Variables predictoras: (Constante), HUMEDAD RELATIVA (%), TEMPERATURA (°C), PRESION (hPa)

b. Variable dependiente: ALTURA(m)

Cuadro 2.30 ANOVA de variables atmosféricas

El estadístico R-cuadrado indica que el modelo explica un 91,6882% de la variabilidad en ALTURA_M. El estadístico R-cuadrado ajustado, que es más conveniente para comparar modelos con diferentes números de variables independientes, es 91,6867%. El error estándar de la estimación muestra la desviación típica de los residuos que es 2083,32. Este valor puede usarse para construir los límites de predicción para las nuevas observaciones seleccionando la opción informes del menú del texto. El error absoluto medio (MAE) de 1630,23 es el valor medio de los residuos. El estadístico Durbin-Watson (DW) cuadro 2.31 examina los residuos para determinar si hay alguna correlación significativa basada en el orden en el que se han introducido los datos.

Resumen del modelo^b

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,958 ^a	,917	,917	2083,321	,001

a. Variables predictoras: (Constante), HUMEDAD RELATIVA (%), TEMPERATURA (°C), PRESION (hPa)

b. Variable dependiente: ALTURA(m)

Cuadro 2.31 Resumen de modelo con múltiples variables

Para decidir la simplificación del modelo, tenga en cuenta que el p-valor más alto en las variables independientes es 0,0000, perteneciendo a PRESIÓN. Puesto que el p-valor es inferior a 0,01, el término de orden superior es estadísticamente significativo para un nivel de confianza del 99%. Por tanto, probablemente no se puede quitar ninguna variable del modelo.

CAPÍTULO 3: MÉTODOS DE REGRESIÓN DIFUSA PARA VARIABLES DEPENDIENTES E INDEPENDIENTES

3.1 INTRODUCCIÓN

A continuación utilizaremos los conjuntos difusos (Apéndice A1) que fundamentan nuestro proyecto de investigación en el campo del análisis de regresión difusa. La justificación para la regresión difusa radica en el manejo de cantidades de datos atmosféricos afectados por imprecisión e incertidumbre⁶ que no está considerado en la regresión probabilística. Según, Jaime Mauricio Figueroa Quezada en su publicación Modelos de Regresión Difusa, Año 2001 menciona que *“El análisis de regresión es utilizado para buscar relaciones entre variables con el propósito de estimar o predecir fenómenos”*. La intención es describir cómo la variable dependiente para nuestro caso la altura debido a que no cambia en el tiempo está relacionada con las variables independientes con la presión, la temperatura y la humedad (en relación a estas 3 variables debido a que existe una correlación alta de estas variables que son apropiadas para construir un modelo, tal como se muestra en el cuadro 2.28 y gráfico 2.19). En esta relación debemos contar con los datos proporcionados por los radio sondeos. La regresión nos ayuda a encontrar un modelo que mejor ajuste el comportamiento de estos fenómenos, sin embargo, raramente describe la verdadera relación entre las variables. Se plantean otros modelos de medidas para la teoría de conjuntos difusos como son las medidas de posibilidad y de necesidad, que se traducen en nuevos retos para afrontar el análisis de regresión. La presencia de incertidumbre, en la regresión probabilística queda plasmada en los intervalos de confianza, generalmente con un nivel del 95%, lo que es atribuido a factores aleatorios. Esta es la única forma de incertidumbre que maneja la teoría de probabilidades.

Según Sergio Donoso Salgado en su publicación Análisis de Regresión Difusa Año 2006 manifiesta que *“La aleatoriedad puede considerarse uno de los componentes de la vaguedad, junto a la imprecisión. Sin embargo, otras formas*

⁶ Cuando no se conoce la probabilidad de ocurrencia de un evento

de incertidumbre se pueden encontrar en la ambigüedad, la incongruencia, problemas de especificación, factores no considerados por complejidad del sistema...". En resumen, podemos decir que hay diversas fuentes de incertidumbre en las variables atmosféricas, entre las cuales la aleatoriedad es sólo una de ellas.

La regresión lineal difusa presentada por Tanaka⁷ nos proporciona los medios para tratar problemas que no cumplen con las suposiciones mencionadas para la regresión estadística. El método se recomienda para situaciones prácticas donde las decisiones muchas veces tienen que ser hechas sobre la base de datos imprecisos o disponibles parcialmente como se considera en nuestro caso. En la regresión lineal difusa las relaciones entre la variable dependiente y las variables independientes no son tan precisas, es decir, se representa a un fenómeno impreciso por su razón de ser difuso. Este método utiliza las funciones difusas definidas por el principio de extensión (ver apéndice A.1), que proporciona un método general para extender conceptos matemáticos no difusos con el fin de tratar con cantidades difusas. Es un método no paramétrico, en el sentido de que las desviaciones de los valores observados de los valores estimados se deben a la imprecisión de los coeficientes del modelo.

3.2 MODELOS DE REGRESIÓN DIFUSA PARA OBSERVACIONES DEPENDIENTES E INDEPENDIENTES DIFUSAS

3.2.1 INTRODUCCIÓN

En este capítulo trataremos las observaciones difusas atmosféricas como la presión atmosférica (en hPa), la humedad relativa (en %) y la temperatura (en °C) vs la variable dependiente la altura (en metros). Al relacionar las variables atmosféricas serán modeladas como funciones difusas. Se supone que es lineal, cuadrática, exponencial o logarítmica según sea el caso. Los coeficientes de la regresión difusa son difusos y son estimados de acuerdo con ciertas condiciones.

⁷ Tanaka, H. Linear regression analysis with fuzzy model. IEEE Transactions on Systems, Man, and Cybernetics, 1982

Posteriormente, aparecen métodos difusos condicionados por ciertos índices que permiten comparar entre dos números difusos. Una condición necesaria y suficiente es: *La igualdad entre números difusos se define por medio de un índice denominado de posibilidad que es el número difuso estimado debe ser igual a la observación difusa correspondiente. Otra condición es que el número difuso estimado debe estar contenido en la observación difusa correspondiente, o viceversa.* La inclusión de dos números difusos se define por medio de un índice denominado de necesidad. Existen algunos conjuntos de coeficientes difusos que satisfacen las condiciones impuestas por cada modelo, se establecen ciertas medidas de vaguedad del modelo, con el objeto de hallar los coeficientes difusos que minimicen o maximicen estas medidas de vaguedad⁸.

3.2.2 MODELOS DE REGRESIÓN LINEAL DIFUSA

La característica fundamental de estos modelos es que su variable dependiente es difusa y las observaciones de las variables independientes se consideran números reales y a la relación de estas observaciones es modelada por una función lineal difusa de la siguiente manera:

$$y_i \approx A_0x_{i0} + A_1x_{i1} + \dots + A_px_{ip} \quad (x_{i0} = 1; i = 1, \dots, n) \quad 3.1$$

Donde,

- A_j , $j = 1, \dots, p$, es el j-ésimo coeficiente desconocido. Es un número difuso LL notado simbólicamente por $A_j = (m_j, \sigma_j)_L$; m_j es su valor medio y σ_j su amplitud y L es la forma de su función de pertenencia.
- x_{ij} , $i = 1, \dots, n$, $j = 0, \dots, p$ es la i-ésima observación de la j-ésima variable, es un número real.
- y_i , $i = 1, \dots, n$, es la i-ésima observación de la variable dependiente, es un número difuso LL notado simbólicamente por $y_i = (y_i, e_i)_L$; y_i es su valor medio y e_i su amplitud.
- El símbolo \approx significa que y_i es aproximado por $A_0x_{i0} + A_1x_{i1} + \dots + A_px_{ip}$

⁸ JAIME FIGUEROA, Modelos de Regresión Difusa con Aplicaciones. Año. 2001.

El tipo de número difuso que se obtiene por medio de la función difusa es del tipo LL (ver apéndice A1)

$$\hat{y}_i = \left(\sum_{j=0}^p m_j x_{ij}, \sum_{j=0}^p \sigma_j |x_{ij}| \right)_L$$

Con la siguiente función de pertenencia

$$\mu_{\hat{y}_i}(y) = L \left(\frac{\left(\sum_{j=0}^p m_j x_{ij} - y \right)}{\sum_{j=0}^p \sigma_j |x_{ij}|} \right)$$

La desviación de la observación difusa de la variable dependiente de su correspondiente número difuso estimado por el modelo 3.1, se nota simbólicamente por ε_i , y es

$$\varepsilon_i = y_i - \hat{y}_i$$

Esta desviación es debido a la imprecisión de la relación imperante entre las variables y se refleja en el modelo por sus coeficientes difusos.

A continuación se presenta los modelos usados en nuestro proyecto de investigación de variables atmosféricas.

3.2.3 EL MODELO CLÁSICO

En este modelo de regresión lineal difusa se deben determinar los coeficientes difusos A_j $j = 1, \dots, p$, tal que el conjunto H-nivel del número difuso \hat{y}_i obtenido a través del modelo 3.1 incluya al conjunto H-nivel de la observación difusa. Para cumplir con esta condición se deben satisfacer las siguientes restricciones:

$$\sum_{j=0}^p m_j x_{ij} - L^{-1}(H) \sum_{j=0}^p \sigma_j |x_{ij}| \leq y_i - L^{-1}(H) e_i \quad i = 1, \dots, n. \quad (3.2)$$

$$-\sum_{j=0}^p m_j x_{ij} - L^{-1}(H) \sum_{j=0}^p \sigma_j |x_{ij}| \leq -y_i - L^{-1}(H) e_i \quad i = 1, \dots, n \quad (3.3)$$

$$\text{Con } H \in [0,1]$$

El parámetro H es establecido por el usuario, da el grado de ajuste del modelo lineal difuso estimado al conjunto de datos disponibles, y $L^{-1}(H)$ es la inversa de la función de pertenencia.

La idea es hallar el vector de coeficientes $\hat{A} = (\hat{A}_0, \dots, \hat{A}_p)$ que producen la “menor vaguedad” del modelo de regresión lineal difusa.

Definición 3.1 Se denomina vaguedad total, S , a la suma de amplitudes individuales de los coeficientes difusos del modelo $A_j, j = 0, \dots, p$, esto es,

$$S(\sigma) = \sum_{i=0}^p \sigma_i$$

donde $\sigma = (\sigma_0, \dots, \sigma_p)$ y σ_j es la amplitud correspondiente al j -ésimo coeficiente difuso.

El problema de encontrar los coeficientes difusos es un problema de programación lineal, y puede ser resuelto utilizando métodos tradicionales; por ejemplo, por el método del Simplex.

3.2.4 EL MODELO DE DOS PASOS

En este modelo, $m = (m_0, \dots, m_p)$ no forma parte de la función objetivo en el problema de programación lineal, y solo el incremento en amplitud dificulta el criterio de elección. Podría explicarse debido a que una observación de las variables independientes, que satisface el valor de pertenencia umbral del modelo identificado sin esta observación en particular, no causa un cambio en el modelo; es decir, el modelo no es sensitivo con respecto a nuevas observaciones de las variables independientes que satisfacen las restricciones del modelo previamente establecido. Entonces, se plantea la formulación de regresión lineal difusa para observaciones reales de las variables dependientes e independientes que pueden resolver el problema de infinitas soluciones bajo observaciones repetidas⁹.

El método propuesto consiste de dos fases:

Se busca la línea de regresión por medio del criterio de mínimos cuadrados utilizando los datos previamente disponibles. El resultado de esta fase, el vector $m^* = (X'X)^{-1} X'Y$, es utilizado como un dato en la siguiente fase. X es la matriz $n \times (p+1)$ de las observaciones de las variables independientes y Y es el vector $n \times 1$ de las observaciones de las variables dependientes.

⁹ Savic D. and Pedrycz W. Evaluation of fuzzy linear regression analysis for fuzzy input-output.

1. Se determina la vaguedad mínima, minimizando S , pero sin considerar al vector m como un vector de variables de decisión. Esto es:

$$S(\sigma) = \sum_{i=0}^p \sigma_i$$

Sujeto a

$$\sum_{j=0}^p m_j x_{ij} - L^{-1}(H) \sum_{j=0}^p \sigma_j |x_{ij}| \leq y_i \quad i = 1, \dots, n$$

$$-\sum_{j=0}^p m_j x_{ij} - L^{-1}(H) \sum_{j=0}^p \sigma_j |x_{ij}| \leq -y_i \quad i = 1, \dots, n$$

$$\sigma_j \geq 0, \quad j = 0, \dots, p$$

Donde σ es el vector de las variables desconocidas.

Si X es una matriz de rango total, entonces m^* es definida en forma única y resuelve la posibilidad de un número infinito de soluciones. Se debe notar que la utilización de la expresión $m^* = (X'X)^{-1} X'Y$ no tiene suposiciones estadísticas implícitas.

3.3 MODELOS BASADOS EN ÍNDICES DE COMPARACIÓN

Se presentarán dos modelos de regresión lineal difusa, el problema es determinar los coeficientes desconocidos $A_j, j = 0, \dots, p$, sujetos a ciertas condiciones basadas en la comparación de la i -ésima observación difusa \hat{y}_i obtenida por medio del modelo (3.1).

3.3.1 MODELOS BASADOS EN LOS ÍNDICES DE NECESIDAD

En estos modelos la i -ésima observación difusa debe estar contenida en su i -ésima estimación por medio del modelo 3.1, o viceversa. Para que se cumpla tal condición mencionamos al siguiente.

Teorema 3.3. $Nes(y_i \subset \hat{y}_i) \geq H$ si y solo sí:

$$y_i + L^{-1}(1-H)e_i \leq \hat{Y}_{iH}^R$$

$$y_i - L^{-1}(1-H)e_i \leq \hat{Y}_{iH}^L$$

La intención es hallar el vector de coeficientes difusos $\hat{A} = (\hat{A}_0, \dots, \hat{A}_p)$ del modelo de regresión difusa (3.1).

Podemos formular el siguiente problema de programación lineal:

$$\begin{aligned} \min \quad & Z_0(m, \sigma) = \sum_{i=1}^n (\hat{Y}_{i0}^R - \hat{Y}_{i0}^L) \\ \text{sujeto a} \quad & \hat{Y}_{iH}^L \leq y_i - L^{-1}(1-H)e_i \quad i = 1, \dots, n \\ & -\hat{Y}_{iH}^R \leq -y_i - L^{-1}(1-H)e_i \quad i = 1, \dots, n \\ & (m, \sigma)_L \in L(J_1, J_2, J_3), \end{aligned}$$

Teorema 4.4 $Nes(y_i \supset \hat{y}_i) \geq H$ si y solo si:

$$\begin{aligned} y_i + L^{-1}(1-H)e_i &\leq \hat{Y}_{i,1-H}^R \\ y_i - L^{-1}(1-H)e_i &\leq \hat{Y}_{i,1-H}^L \end{aligned}$$

Entonces, se puede formular el siguiente problema de programación lineal:

$$\begin{aligned} \max \quad & Z_0(m, \sigma) = \sum_{i=1}^n (\hat{Y}_{i0}^R - \hat{Y}_{i0}^L) \\ \text{sujeto a} \quad & \hat{Y}_{i,1-H}^L \leq y_i - L^{-1}(1-H)e_i \quad i = 1, \dots, n \\ & -\hat{Y}_{i,1-H}^R \leq -y_i - L^{-1}(1-H)e_i \quad i = 1, \dots, n \\ & (m, \sigma)_L \in L(J_1, J_2, J_3), \end{aligned}$$

3.4 PARÁMETROS Y MEDIDAS DE BONDAD DE AJUSTE DE LA REGRESIÓN DIFUSA

Otro aspecto fundamental de la regresión difusa es el estudio de los parámetros y medidas de bondad del ajuste obtenido. Para ello, veremos algunos índices de bondad que se pueden utilizar para determinar la calidad de la estimación hecha gracias al proceso de regresión. No son muchos las aportaciones de regresión difusa que incorporan un estudio de bondad de ajuste de los métodos

propuestos¹⁰. Se tiene como objetivo construir índices de bondad de ajuste. En esta propuesta se define un coeficiente de correlación híbrido, como

$$(HR)^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Este indicador es idéntico al coeficiente de determinación de la regresión probabilística (R^2), con la diferencia que aquí se construye con la aritmética difusa ponderada. Sin embargo, el índice puede tomar valores mayores que 1, lo que hace menos atractivo que otros índices que fluctúan entre 0 y 1.

Chang¹⁰, también define las siguientes medidas, usando la aritmética difusa ponderada:

Definición 3.1 (Media híbrida).

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

Definición 3.2 (Desviación estándar híbrida).

$$S_Y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Definición 3.3 (Error estándar de la estimación híbrido).

$$HS_e = \sqrt{\frac{1}{n-m-1} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}$$

Por otro lado, Chang define un interesante *coeficiente de regresión parcial estandarizado* t_i , mediante la aritmética difusa ponderada y como un número preciso. Este autor sugiere que dicho valor representa el nivel de importancia que la variable que acompaña a cada coeficiente tiene en el modelo. Es decir, si $t_i > t_j$, entonces, la variable X_i es más importante en el modelo que la variable X_j .

Definición 3.4 (Coeficiente de regresión parcial estandarizado).

$$t_i = \hat{A}_i \frac{S_x}{S_y}$$

¹⁰ Chang Y. Fuzzy regression methods –a comparative assesment. Fuzzy and Systems 1998.

D,Urso¹¹ utiliza estas dos definiciones de R^2 para un proceso iterativo de ajuste polinomial, de manera que cuando R^2 no aumenta significativamente, se termina el proceso de incorporar nuevos términos polinomiales

Una propuesta interesante de Wang H.¹² que tiene en cuenta que la suma de cuadrados totales (SCT) se puede descomponer en suma de cuadrados de la regresión (SCR) más la suma de cuadrados del error (SCE), en forma similar al caso probabilístico. Para este fin se define

$$SCT = \sum_{i=1}^n (y_i - \widehat{y_i - p_i})^2 + \sum_{i=1}^n (\widehat{y_i + q_i} - y_i)^2$$

$$SCR = \sum_{i=1}^n (\hat{y}_i - y_i - p_i)^2 + \sum_{i=1}^n (\widehat{y_i + q_i} - \hat{y}_i)^2$$

$$SCE = 2 \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

De forma que se puede definir un índice de confianza, IC, como

$$IC = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT}$$

Este resultado, que es similar al de la regresión probabilística, solo es válido cuando la función de pertenencia es simétrica. Podemos indicar que los trabajos sobre regresión difusa no han profundizado sobre el estudio de la calidad de la estimación, y la mayoría de las propuestas corresponden a caso específicos de desdifusificación o funciones de pertenencia simétricas.

3.5 APLICACIÓN DE LOS MODELOS DE REGRESIÓN DIFUSA A LAS VARIABLES METEOROLÓGICAS

Una vez conocidos los modelos de regresión difusa implementamos un algoritmo de programación que permita obtener los modelos de regresión difusa por el Método de Dos Pasos para cada una de las variables meteorológicas. Este algoritmo fue diseñado y construido en el paquete computacional MATLAB a través de una Interfaz que permite al usuario obtener y visualizar los diferentes escenarios en que las variables meteorológicas se comportan en el perfil vertical.

¹¹ D,Urso P. An orderview polynomial regression procedure for fuzzy data. Año 2000.

¹² Wang H. Insight of a fuzzy regression model. Año 2002.

El objetivo es construir un intervalo de confianza para cada variable atmosférica en el perfil vertical, gráfico 3.1. Además el programa permite automatizar los archivos de los datos de las radio sondas entregados por el INAMHI (cuadro 2.1). Los modelos se ajustan de acuerdo con las restricciones establecidas por el Método de Dos pasos. A su vez, este algoritmo nos permite visualizar en cada una de las variables meteorológicas, dentro de un intervalo de confianza, el posible comportamiento que tendría cada variable en los diferentes niveles de la altura. Debajo de cada gráfico se muestra un resumen de los datos obtenidos para los diferentes niveles de la altura y sus promedios.

Los pasos considerados en la implementación del algoritmo en MATLAB son:

1. Ingreso de datos.
2. Elaboración de la matriz de datos.
3. Algoritmos de regresión:
 - Regresión Lineal Difusa, según sea el caso de las variables.
 - a. Cambio de escala a variables que puedan ser ajustas por un línea recta.
 - b. Mínimos, máximos de datos. (Escalas).
 - c. Función a minimizar, límites y restricciones.
 - Matrices de regresión y limites: A,b,f,lb
 - Simplex, que permite encontrar los regresores del modelo
 - Salida de los regresores: $x=[\sigma_0, \sigma_1, \mu_0, \mu_0]$ esto en el caso de la regresión lineal
 - Compilación numérica para elaborar las graficas
4. Salida de resultados: Numéricos y gráficas.

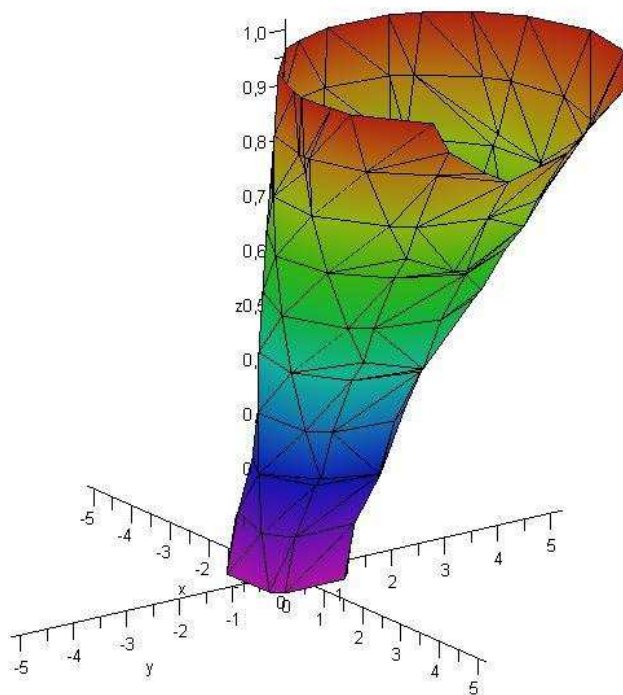


Gráfico 3.1 Región de confianza para los lanzamientos de radio sondas

3.5.1 RESULTADOS OBTENIDOS PARA UNA RADIO SONDA

El mejor lanzamiento para el análisis de estas variables fue tomado del 27 de mayo del 2009 alrededor de las 13H30 horas.

Se obtuvieron 5404 datos de la radiosonda alcanzando hasta una altura de 29038 metros.

i. PRESIÓN ATMOSFÉRICA (hPa)

Para el análisis de la Altura vs Presión debido al comportamiento estable de estas variables utilizamos el modelo de regresión lineal difusa (3.1) para observaciones dependientes difusas. El problema es determinar encontrar el modelo de regresión lineal difusa que mejor se ajuste a los datos, se soluciona resolviendo el problema de programación lineal asociado (3.2).

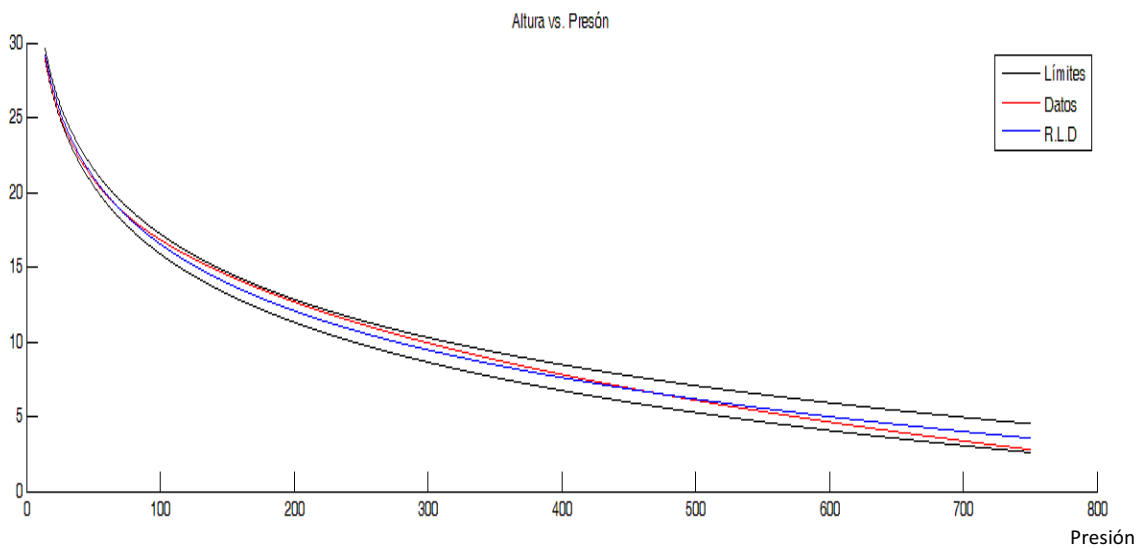
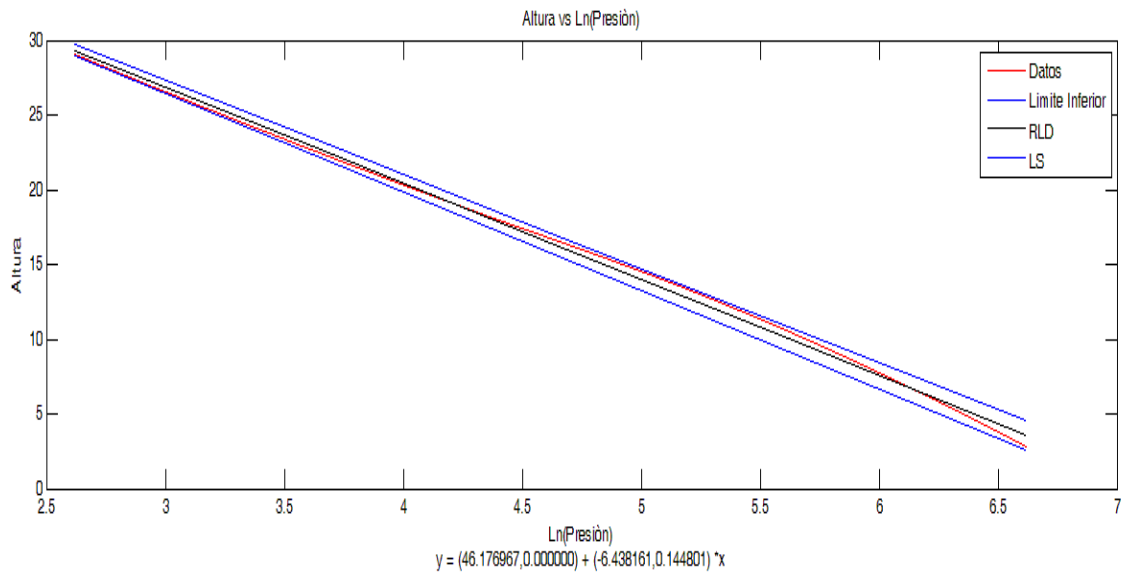


Gráfico 3.2 Regresión Difusa Presión Atmosférica (hPa)

Utilizando el modelo de dos pasos se obtiene el modelo de la regresión difusa.

Gráfico 3.2

$$\begin{aligned}
 A_0 &= (46,1770,0) & Y &= 46,177 - 6,4382x_1 \\
 A_1 &= (-6,4382,0.2317)
 \end{aligned}$$

ALTURA	L.INFERIOR	MEDIO	L.SUPERIOR
16.3	93.5738 hPa	103.607 hPa	115.254 hPa
16.7	88.057 hPa	97.3659 hPa	108.157 hPa
17.1	82.8655 hPa	91.5007 hPa	101.497 hPa
17.5	77.9801 hPa	85.9888 hPa	95.2466 hPa
17.9	73.3826 hPa	80.8089 hPa	89.3815 hPa
18.3	69.0562 hPa	75.9411 hPa	83.8776 hPa
18.7	64.9849 hPa	71.3665 hPa	78.7125 hPa

El gráfico 3.2 muestra el comportamiento de la variable presión con respecto a la altura, y el comportamiento de la misma variable implementado el modelo de regresión difusa. Se puede observar claramente que el modelo difuso ajusta de una buena forma las estimaciones de los datos a obtenerse por el modelo. Siendo las observaciones reales las siguientes:

ALTURA (m)	PRESIÓN (hPa)
16300	109.50
16700	102.00
17100	95.00
17504	88.40
17900	82.40
18300	76.80
18700	71.70

ii. TEMPERATURA (°C)

Para analizar el comportamiento de la temperatura, gráfico 3.3, es necesario analizar los datos por tramos, debido al descenso que comienza a producirse alrededor de los 17 a 18 Km. Utilizando el método de regresión difusa a dos pasos para los datos que van desde la altura de 2789 metros hasta 17835 metros y de 17835 a 29038 metros.

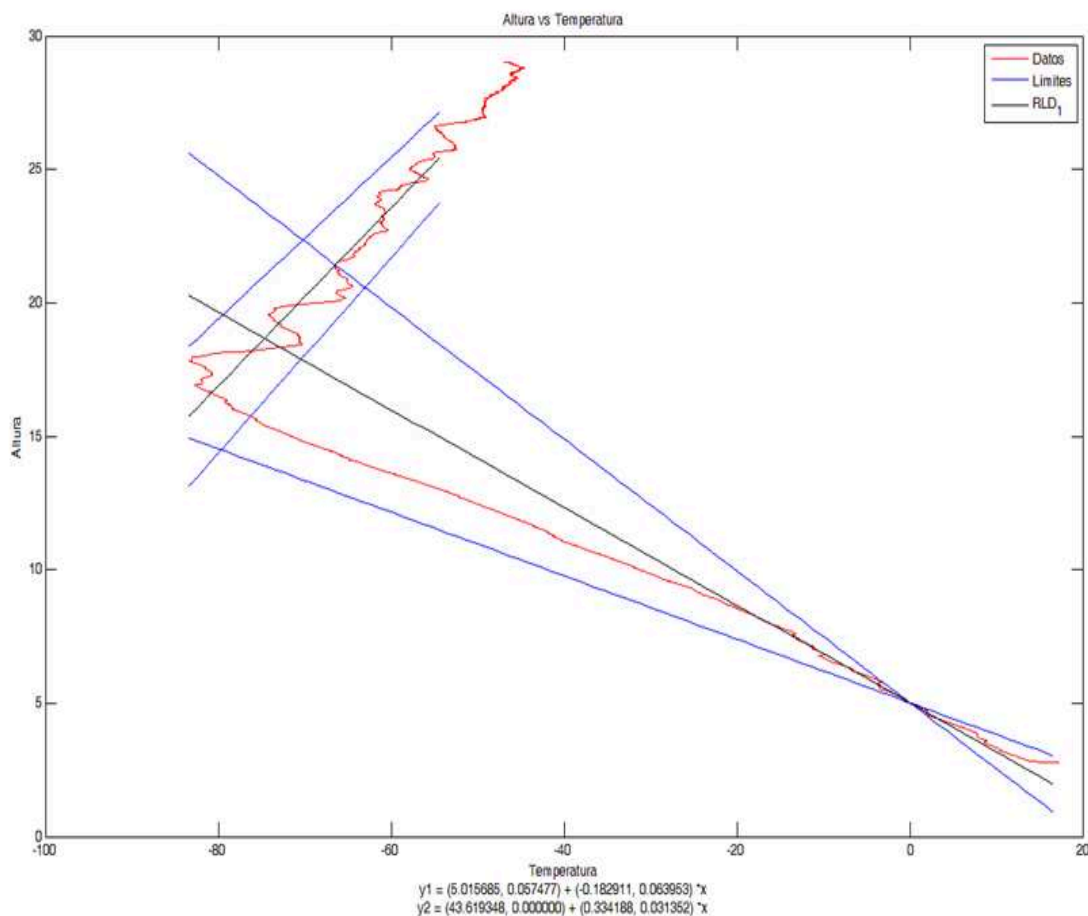


Gráfico 3.3 Regresión Difusa Temperatura (°C)

1.) El modelo obtenido para los datos del descenso de la temperatura (-83,35°C) hasta los 17821 metros es:

$$A_0 = (5,0157; 0,057) \quad Y = 5,0157 - 0,1829x_1 .$$

$$A_1 = (-0,1829; 0,064)$$

2.) El modelo obtenido para los datos siguientes de la temperatura hasta los 29038 metros es:

$$A_0 = (43,6193, 0) \quad Y = 43,6193 + 0,334x_1 .$$

$$A_1 = (0,334, 0,0313)$$

ALTURA	L.INFERIOR	MEDIO	L.SUPERIOR
15.62	-92.4571 °C	-83.7832 °C	-76.5972 °C
16.36	-90.0136 °C	-81.5689 °C	-74.5728 °C
17.1	-87.57 °C	-79.3546 °C	-72.5484 °C
17.84	-85.1264 °C	-77.1403 °C	-70.524 °C
18.58	-82.6829 °C	-74.9259 °C	-68.4996 °C
19.32	-80.2393 °C	-72.7116 °C	-66.4752 °C

Las observaciones reales para estos niveles de altura son:

ALTURA (m)	TEMPERATURA (°C)
15620	-81.45
16360	-81.85
17100	-81.85
17840	-83.25
18580	-73.65
19320	-70.55

iii. HUMEDAD RELATIVA (%)

Debido a la irregularidad en los datos, se implementa una regresión polinomial. La eficiencia en las corridas del algoritmo es una prueba que debe superarse con supresión de líneas de código o improvisación de funciones programadas en el software MATLAB. Aquellos valores que son menores a cero no tienen sentido en la medida de la humedad relativa, es por esto que en la regresión se los toma como 1 ya que este es el mínimo valor que alcanza la variable en estudio.

Los valores observados para esta variable son:

ALTURA (m)	HUMEDAD (%)
15620	31.00
16360	28.00
17100	24.00
17840	20.00
18580	17.00
19320	8.00

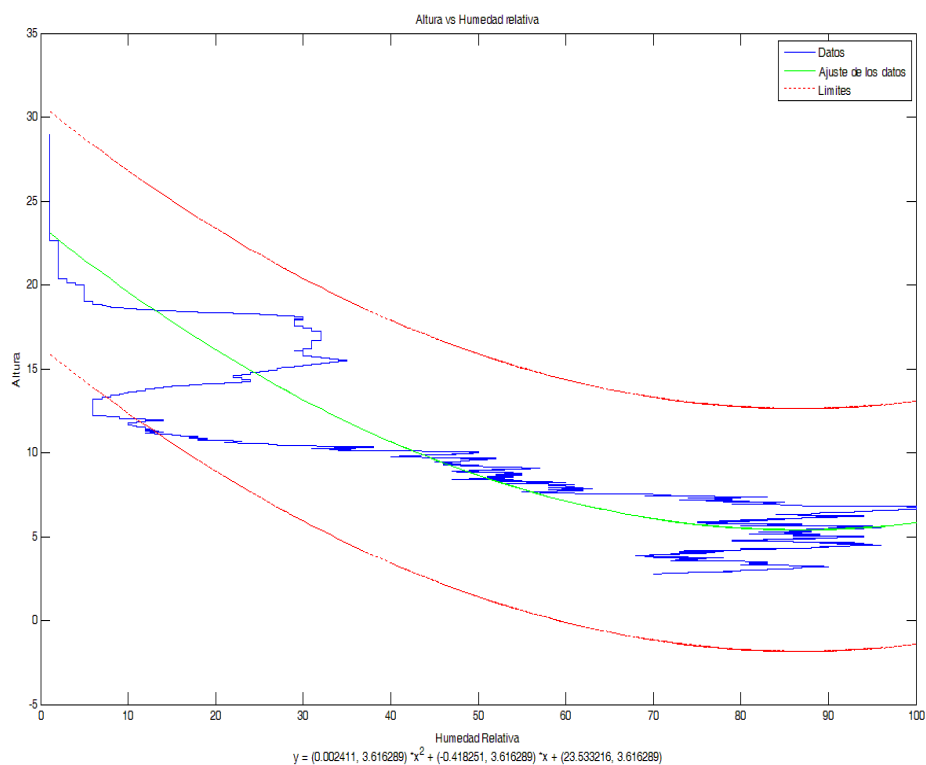


Gráfico 3.4 Regresión Difusa Humedad Relativa (%)

ALTURA	L.INFERIOR	MEDIO	L.SUPERIOR
15.62	10.96687 %	21.61232 %	34.37921 %
16.36	8.967885 %	19.29706 %	31.52596 %
17.1	7.019012 %	17.05869 %	28.81309 %
17.84	5.116658 %	14.89002 %	26.22171 %
18.58	3.257645 %	12.78492 %	23.73683 %
19.32	1.439137 %	10.7381 %	21.3463 %
20.06	-0.341403 %	8.744979 %	19.04014 %

3.5.2 RESULTADOS OBTENIDOS PARA LOS PROMEDIOS DE RADIO SONDAS EN AMBATO

I. PRESIÓN ATMOSFÉRICA (hPa)

ALTURA (m)	PRESIÓN (hPa)
16300	121.13
16700	117.25
17100	113.85
17500	109.32
17900	102.69
18300	97.81
18700	90.48

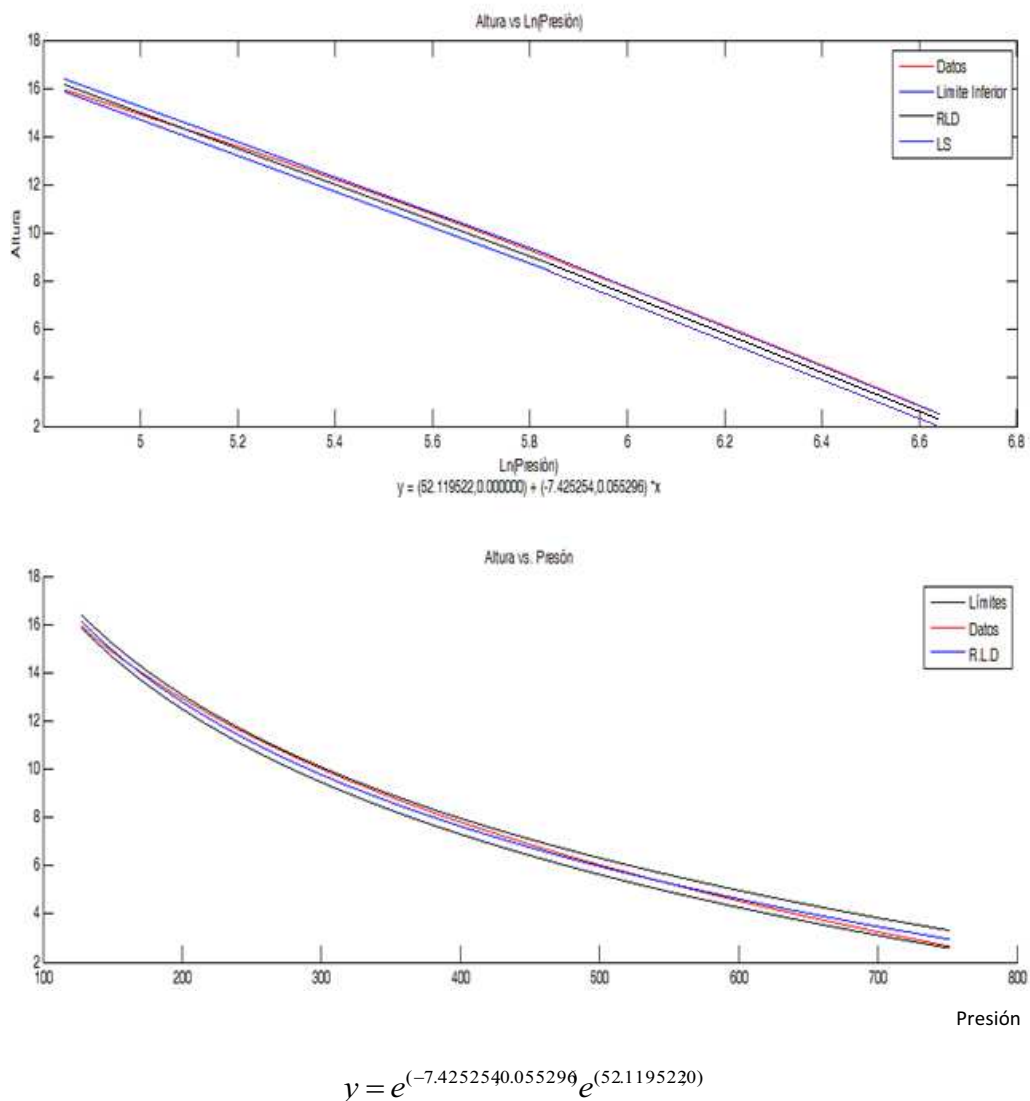


Gráfico 3.5 Regresión Difusa Promedios Presión (hPa)

ALTURA	L.INFERIOR	MEDIO	L.SUPERIOR
16.3	120.104 hPa	124.464 hPa	129.051 hPa
16.7	113.85 hPa	117.937 hPa	122.234 hPa
17.1	107.923 hPa	111.751 hPa	115.777 hPa
17.5	102.303 hPa	105.891 hPa	109.66 hPa
17.9	96.9766 hPa	100.337 hPa	103.867 hPa
18.3	91.9273 hPa	95.075 hPa	98.3802 hPa
18.7	87.1409 hPa	90.0888 hPa	93.183 hPa

II. TEMPERATURA (°C)

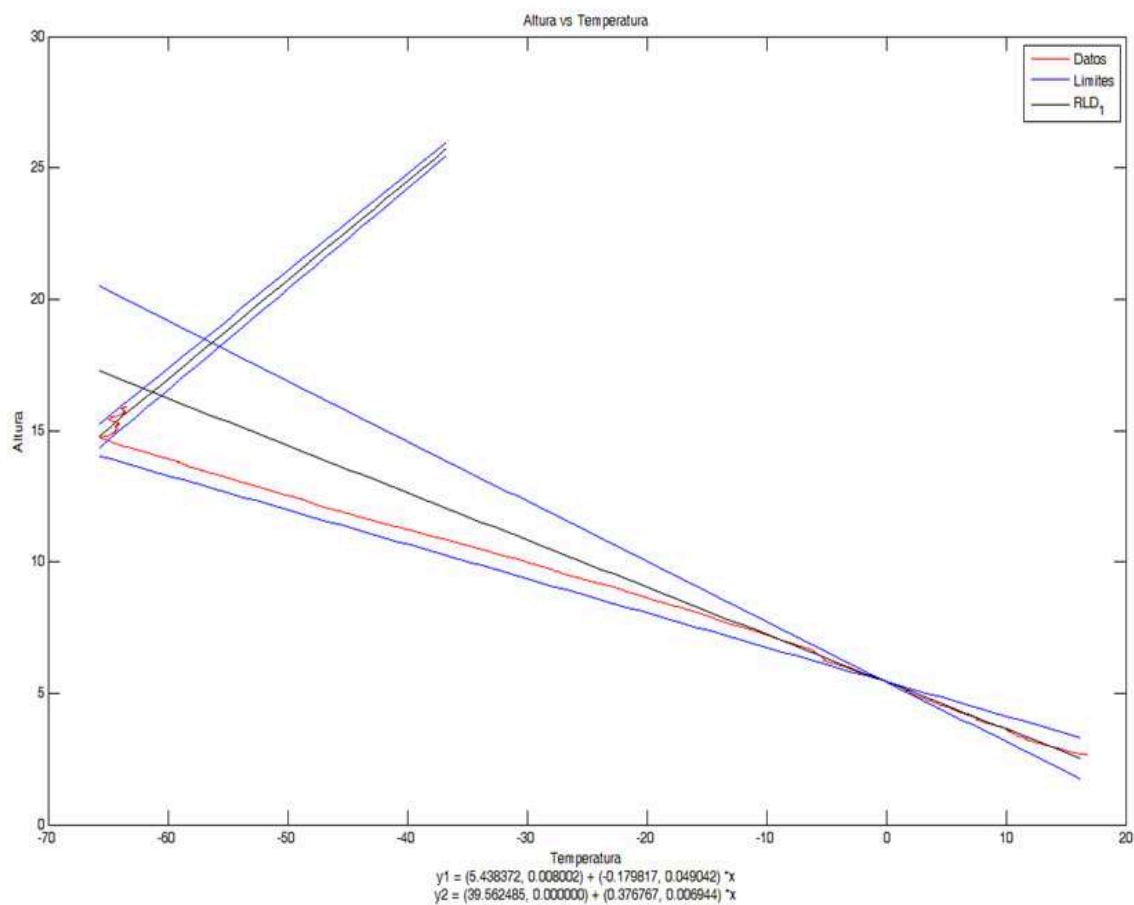


Gráfico 3.6 Regresión Difusa Promedios Presión (hPa)

ALTURA	L.INFERIOR	MEDIO	L.SUPERIOR
14.77	-67.0388 °C	-65.8032 °C	-64.6124 °C
14.83	-66.8765 °C	-65.644 °C	-64.456 °C
14.89	-66.7143 °C	-65.4847 °C	-64.2997 °C
14.95	-66.5521 °C	-65.3255 °C	-64.1433 °C
15.01	-66.3898 °C	-65.1662 °C	-63.9869 °C
15.07	-66.2276 °C	-65.007 °C	-63.8306 °C
15.13	-66.0653 °C	-64.8477 °C	-63.6742 °C

Las observaciones de datos promedios para esta variable son:

ALTURA (m)	TEMPERATURA (°C)
14770	-63.53
14830	-65.08
14890	-64.98
14950	-64.89
15010	-63.79
15070	-63.42
15130	-63.03

III. HUMEDAD RELATIVA (%)

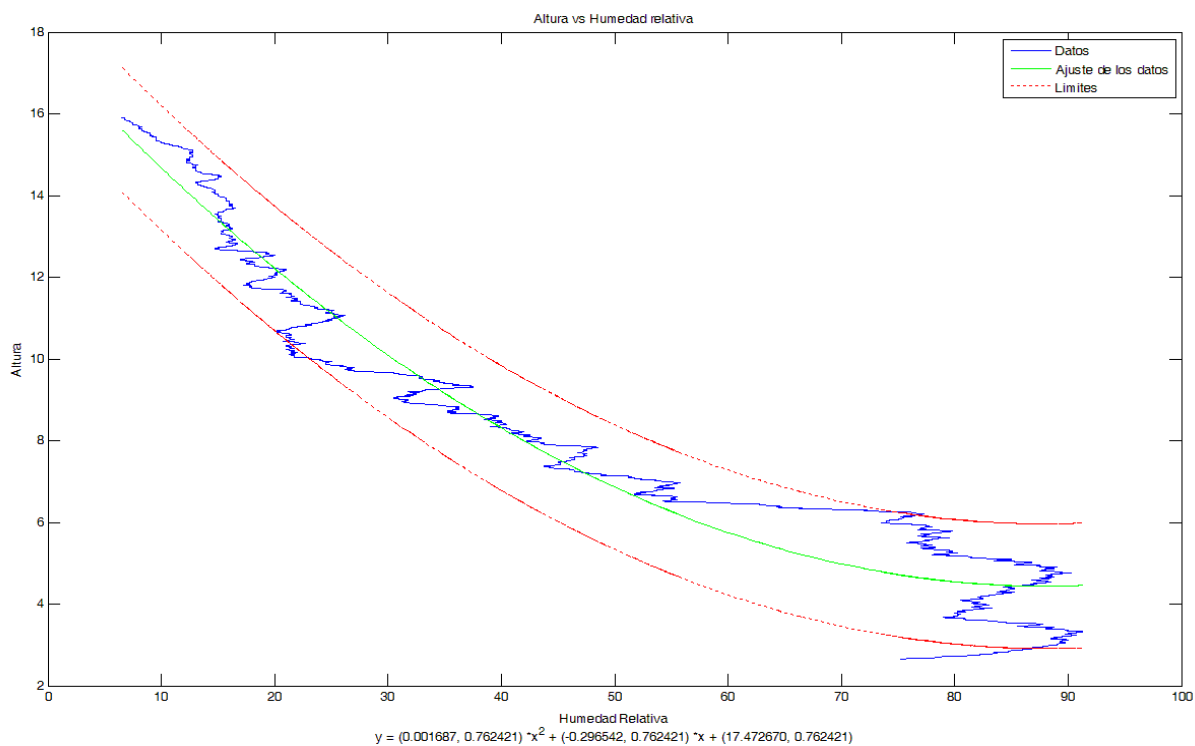


Gráfico 3.7 Regresión Difusa Promedios Humedad Relativa (%)

ALTURA	L.INFERIOR	MEDIO	L.SUPERIOR
14.3602	8.3185017 %	11.21089 %	14.21674 %
14.53	7.6885352 %	10.55736 %	13.53678 %
14.6998	7.0634785 %	9.9093 %	12.86299 %
14.72	6.989441 %	9.832563 %	12.78323 %
14.7402	6.9154712 %	9.755902 %	12.70356 %
14.7604	6.8415689 %	9.679316 %	12.62398 %
14.7806	6.7677339 %	9.602804 %	12.54448 %

Las observaciones de datos promedios para esta variable son:

ALTURA (m)	HUMEDAD (%)
14360	12.0
14530	12.0
14699	10.0
14720	9.0
14740	9.0
14760	8.0
14780	7.0

CAPÍTULO 4. APLICACIÓN DE TÉCNICAS DIFUSAS A LAS VARIABLES ATMOSFÉRICAS.

4.1 INTRODUCCIÓN.-

Hoy en día, una de las grandes ventajas es el desarrollo alcanzado por los microprocesadores que permiten almacenar ciento de miles de datos, como puntos en espacios de una alta dimensionalidad, en tiempos relativamente cortos, información que se presenta, en diferentes espacios de la vida real, en base de datos de instituciones financieras, telecomunicaciones, biología, meteorología entre otros.

Varias son las aplicaciones en las que se requiere el manejo de las bases de datos espaciales que permiten conocer la identificación de grupos, para descubrir importantes distribuciones del espacio en estudio, en nuestro caso las variables atmosféricas tienen un comportamiento inestable y caótico, razón por la cual se utilizarán técnicas de agrupamiento difusas.

Para el reconocimiento de formas existe un amplio conjunto de técnicas multivariantes para el tratamiento de datos en la selección y extracción de características, la división de los datos en grupos (agrupamiento) y la clasificación de un objeto en un grupo dado.

Dentro de este enfoque, se distinguen entre los métodos paramétricos y no paramétricos. Para el primer caso, se conoce a priori sobre la forma funcional de las distribuciones de probabilidad de cada clase sobre el espacio de representación, determinadas por un conjunto finito y normalmente fijo de parámetros, las fronteras de decisión estarán definidas por dichas distribuciones. En cambio, para el segundo caso no supone ninguna forma de las distribuciones de probabilidad sobre el espacio de representación, de modo que el único reconocimiento a priori será el correspondiente a la información inducida a partir de un conjunto de muestras, las fronteras de decisión estarán determinadas por las muestras del conjunto de entrenamiento.

El problema de formar grupos en un conjunto de datos es muy importante ya que se determina el comportamiento de una población de la cual solo se tiene una parte de sus elementos. La solución de estos problemas se realiza mediante la creación de algoritmos de agrupamiento, que pueden definirse como sigue: *dado n puntos en un espacio n -dimensional particionar los mismos en k grupos tales que los puntos dentro de un grupo sean más similares y diferentes a los de los otros grupos, dicha similaridad se mide atendiendo a alguna función distancia (función de disimilaridad) o alguna función de similaridad.*

Los métodos de agrupamiento no paramétricos pueden dividirse en tres grupos fundamentales: jerárquicos, particionales y basados en densidad¹³.

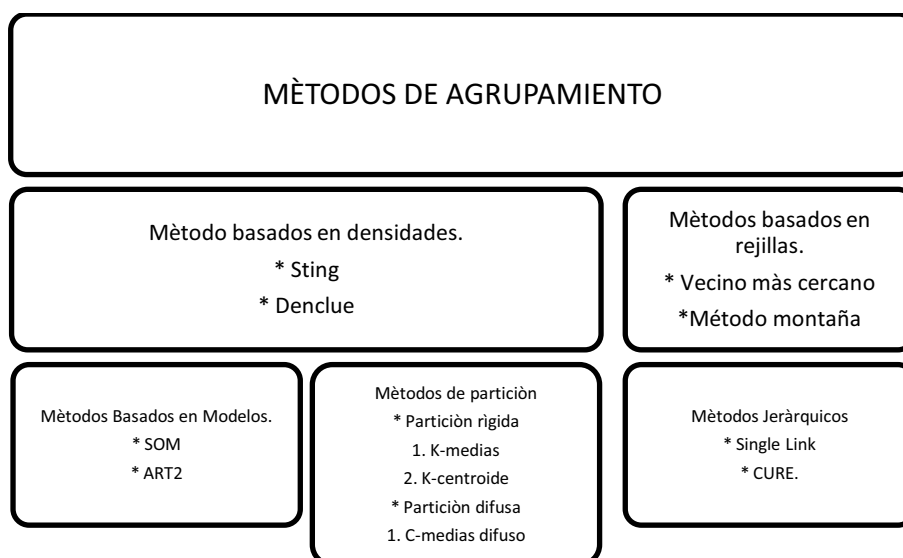


Gráfico 4.1 Métodos de Clasificación

Para el desarrollo de nuestro estudio nos enfocaremos en el método de agrupamiento difuso C-Medias.

4.2 AGRUPAMIENTO DIFUSO CON FUNCIÓN OBJETIVO

Como es de conocimiento en el agrupamiento difuso los objetos pueden pertenecer a más de un conjunto, por ende tendrán un grado de pertenencia a cada partición representando normalmente por un número real en el intervalo

¹³ "Algoritmos de agrupamiento". D. Pascual, F. Pla, S. Sánchez pág. 164

[0,1]. De esta forma, la representación de datos inciertos (o de ruido) se puede hacer de forma más adecuada.

La base del agrupamiento difuso es la partición difusa; de esta forma, una partición difusa U se representa por una relación difusa¹⁴

Sea $X = \{x_1, x_2, \dots, x_m\}$ en el cual x_k ($k = 1, 2, \dots, m$) es una variable u objeto dentro de los datos; x_1, x_2, \dots, x_m son vectores en el espacio \mathcal{R}^n y sea c un entero con valores $2 \leq c < m$. La matriz de pertenencias $U = [\mu_{ik}]$, representa la partición rígida de los datos en X si y solo si sus elementos satisfacen las siguientes propiedades:

$$u_{ik} \in [0,1], \quad 1 \leq k \leq m, \quad 1 \leq i \leq c$$

$$\sum_{i=1}^c u_{ik} = 1, \quad 1 \leq k \leq m$$

$$0 < \sum_{k=1}^m u_{ik} < m, \quad 1 \leq i \leq c$$

La i -ésima columna de U contiene los valores de la función de pertenencia del i -ésimo subconjunto difuso de X . La información de pertenencia actual, al proporcionar información acerca de la pertenencia de un objeto a cada conglomerado puede permitir un mejor manejo de datos atípicos. En efecto, si a un punto se le asignan valores bajos de pertenencia a cada grupo, esto es un indicativo de que no está suficientemente relacionado a ninguno de los grupos y por tanto puede ser un valor atípico. Basado en las anteriores definiciones de partición, se han desarrollado diferentes métodos de conglomeración.

La segunda restricción no es imprescindible para formar una partición borrosa.

Para algunos autores de técnicas de agrupamiento difuso se presentan tres categorías:

¹⁴ Algoritmo de Agrupamiento Genético Borroso Basado en el algoritmo de las C-Medias Borroso. Sergio López García. Año 2001.

1. Agrupamiento difuso basado en función objetivo, que son los más extendidos.
2. Agrupamiento difuso basado en relación difusa.
3. Generalización del algoritmo del vecino más cercano.

En este capítulo utilizaremos el algoritmo basado en función objetivo y el más conocido, el FCM o algoritmo de las c-medias difuso.

4.3 TÉCNICA DE AGRUPAMIENTO DIFUSO C-MEDIAS

Podemos definir qué agrupación es el proceso de organización de objetos en grupos cuyos miembros son similares de alguna manera. Una agrupación (clúster) es tanto una colección de objetos que son “similares” entre ellos y son “diferentes” a los objetos pertenecientes a otros grupos.

En el gráfico 4.2 se identifica fácilmente los 4 grupos en los que los datos se pueden dividir, el criterio de similitud es la distancia. Dos o más objetos pertenecen al mismo grupo si están “cerca” de acuerdo a una determinada distancia. Esto se conoce como agrupación basada en la distancia.

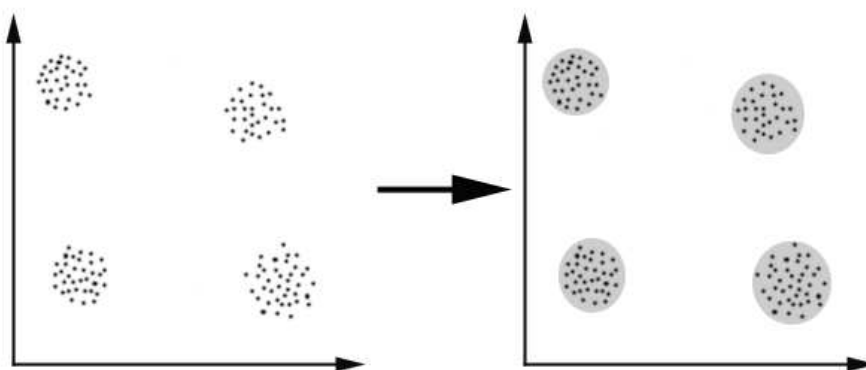


Gráfico 4.2. Grupos de división de datos

Otro tipo de agrupación es el agrupamiento conceptual, dos o más objetos pertenecen al mismo grupo si se define un concepto común a todos los objetos. Es decir, los objetos son agrupados de acuerdo a los conceptos descriptivos, no de acuerdo con las medidas de similitud simple.

Los principales requerimientos que un algoritmo de agrupamiento debe cumplir son:

1. Escalabilidad;
2. Abordar los diferentes tipos de atributos;
3. Descubrir los grupos con forma arbitraria;
4. Requisitos mínimos para el conocimiento del dominio que permitan determinar los parámetros de entrada;
5. Capacidad para tratar con el ruido y los valores extremos;
6. Alta dimensionalidad;
7. Interpretación y usos.

4.3.1 ALGORITMO DE LAS C-MEDIAS DIFUSO (FCM)

Para la técnica de agrupamiento difuso C-Medias (FCM) se debe determinar la partición de la muestra de los datos para cada variable exploratoria (entrada) en un número determinado de conglomerados.

Dado un conjunto de objetos X de la forma:

$$X = \{x_1, x_2, \dots, x_n\}$$

Donde cada x_i es un objeto descrito por p características (es decir, $x_i \in \mathbb{R}^p$). Se define una partición c-difusa como una clase de c conjuntos difusos, siendo c un entero en el rango $[2, n]$. Tal como se ha definido previamente, una c-partición difusa para X es el conjunto:

$$M_{fcn} = \left\{ U \in \mathbb{R}^{c \times n} \mid u_{ik} \in [0, 1], \forall i, \forall k; \sum_{i=1}^c u_{ik} = 1, \forall k; 0 < \sum_{k=1}^n u_{ik} < n, \forall i \right\}$$

A pesar de que el algoritmo FCM está siendo utilizado con éxito en muchos problemas reales, tiene varios inconvenientes para la resolución de ciertos problemas de agrupamiento¹⁵. Los principales son:

¹⁵ Algoritmo de Agrupamiento Genético Borroso Basado en el algoritmo de las C-Medias Borroso. Sergio López García. Año 2001.

1. Los resultados del algoritmo dependen de la partición inicial, la cual es totalmente arbitraria. Una mala selección de la partición inicial puede llevar a un mínimo local de la función objetivo, en vez de llegar al mínimo global.
2. Otro problema importante es que el FCM no calcula el número c de conjuntos, ya que dicho número se fija en el primer paso del algoritmo. En algunos casos este dato puede ser conocido, pero en otros puede no estar disponible.
3. Por último, el uso de la distancia euclídea lleva a que el FCM sólo detecte conjuntos hiperesféricos (esferas de n dimensiones). La detección de otras formas de conjuntos obligaría a utilizar definiciones distintas de distancias.

El objetivo del algoritmo es encontrar la mejor partición U en M_{fcn} . Para ello se minimiza la siguiente función:

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m d_{ik}^2(v_i, x_k), U \in M_{fcn}, 1 < m < \infty$$

donde los v_i son los prototipos de cada conjunto o centroides, m es una función de peso y d es la distancia euclídea. Los centroides se obtienen mediante la formula:

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m}$$

El algoritmo FCM es iterativo. Sus pasos son:

1. Elegir una partición u_0 .
2. Calcular los centroides v_i
3. Actualizar las funciones de pertenencia difusas

$$u_{ik} = \frac{\left(\frac{1}{d_{ik}^2(v_i, x_k)} \right)^{\frac{1}{m-1}}}{\sum_{j=1}^c \left(\frac{1}{d_{jk}^2(v_j, x_k)} \right)^{\frac{1}{m-1}}}$$

4. Se repiten los pasos (2) y (3) hasta que el valor de J_m no decrezca más o esta disminución sea menor que un umbral. Esta es la condición de final de la iteración.

En algunos autores podemos encontrar algunas consideraciones con respecto al peso m . Aunque se ha demostrado que el algoritmo converge para cualquier valor de $m \in (1, \infty)$, resulta que las particiones son cada vez más difusas según se incrementa m . Esto implica que el valor de m se debe elegir en función del problema particular bajo consideración.

4.4 TÉCNICA DE AGRUPAMIENTO DE LA C- MEDIAS DIFUSO MEJORADO

Un componente importante de un algoritmo de agrupamiento es el medir la distancia entre puntos de datos. Si los componentes de los vectores de datos se encuentran con las mismas unidades físicas entonces, es suficiente utilizar la distancia métrica euclídea. El gráfico 4.3 muestra el ancho y la altura de un objeto. A pesar de que ambas medidas adoptan las mismas unidades físicas, la decisión tiene que tomarse con conocimiento de la escala relativa, ya que, diferentes escalas puede llevar a agrupamientos diferentes.

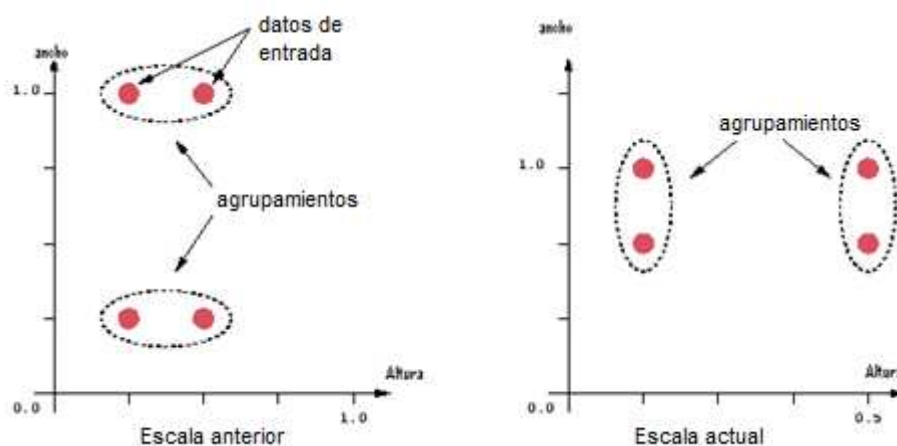


Gráfico 4.3. Datos Antes y Después del Escalamiento

No es solamente un problema de gráficos: el problema surge de la fórmula matemática utilizada para combinar las distancias entre los componentes

individuales de los datos de vectores en función de la distancia única que se puede utilizar para agrupar los propósitos, las fórmulas diferentes conduce a diferentes agrupamientos. Por ello, podemos encontrar mejoras a la técnica de la c-medias difuso (FCM) que es donde nos enfocaremos.

Para el análisis de nuestro estudio recordemos que el método del FCM requiere como entrada el número de agrupaciones c en que se dividen los datos. En ciertas ocasiones podemos determinar este dato. Por ejemplo en un proceso que se fabrican n productos y que requiere llevar a cabo un control de calidad; se sabe que cada grado de producto corresponde a una producción con condiciones de operación distintas a la del resto; por lo tanto, se espera que al aplicar el método, con el dato de entrada $n=c$, la información de las operaciones puntuales a lo largo de los históricos se clasifique en n grupos o categorías.

Se plantea una forma más general planteada en “Algorithms for Fuzzy Clustering Methods in c-Means Clustering with Applications. Autor: Sadaaki Miyamoto, Hidetomo Ichihashi. Editorial: Springer, Berlin 2008” a partir del cual plantearemos la mejora al algoritmo, en el cual se debe determinar la partición de la muestra de los datos para cada variable exploratoria (entrada) en un número determinado de conglomerados. Estos grupos tienen bondades “difusas”, en el sentido de que cada valor de datos pertenece a cada grupo en un cierto grado. La pertenencia no es “única”. Después de haber decidido el número de agrupaciones a seleccionar, es necesario utilizar un procedimiento que permita localizar sus puntos medios (o de manera más general, sus centroides), determinar las funciones de pertenencia y los grados de pertenencia de los datos a cada grupo.

Este método se basa en el concepto de particiones difusas. Se define la matriz de datos X de dimensión $m \times n$, y lo que se desea es dividir los “ n ” datos de cada variable en “ c ” grupos difusos $G_i = (i = 1, \dots, c)$ donde $c < n$, y al mismo tiempo determina la agrupación de estas agrupaciones en el espacio correspondiente.

Los datos pueden ser multidimensionales, y las métricas que constituye la base para el FCM utilizan el error cuadrático de la distancia. La base matemática para este procedimiento es el siguiente. Sea x_k el k -ésimo vector de puntos ($k = 1, \dots, n$).

Sea v_i el centro del i -ésimo grupo difuso ($i = 1, \dots, c$). Sea $d_{ik} = \|x_k - v_i\|$ la distancia entre x_k y v_i , y u_{ik} el grado de adhesión del vector " k " en el grupo " i ", donde:

$$\sum_{i=1}^c (u_{ik}) = 1$$

El objetivo es la partición de los datos en " c " agrupaciones, y simultáneamente localizar las agrupaciones y determinar el correspondiente grado de pertenencia a fin de minimizar el funcional:

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m (d_{ik})^2$$

No hay forma prescrita para la elección del parámetro exponente " m ", este debe cumplir con la siguiente condición $1 < m < \infty$; y es muy importante que si este toma valores cercanos a 1 la partición de los datos resultantes será rígida, mientras que valores más lejanos a 1 harán que la partición sea más difusa. En la práctica, $m = 2$ es una común elección. En términos generales, el algoritmo FCM implica los siguientes pasos:

1. Seleccionar la ubicación inicial para la agrupación de centros.
2. Generar una (nueva) partición de los datos mediante la asignación de cada punto de datos a un grupo cuyo centro sea más cercano.
3. Calcular nuevo grupo de centros como los centroides de los grupos.
4. Si el grupo de partición es estable entonces se detiene. De lo contrario ir al paso 2.

En el caso de membresías difusas, el multiplicador de Lagrange genera la siguiente expresión que se utilizarán en el paso 2:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left[\left(\frac{d_{ik}^2}{d_{ij}^2} \right)^{\frac{1}{m-1}} \right]}$$

En efecto, de una manera más general se desea resolver el problema:

$$\min J = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|x_k - v_i\|^2 \quad (1)$$

con respecto a $v_i \in V$ y $u_{ik} \in U$.

La solución a este problema se la realiza en dos pasos, primero hay que tomar en cuenta que la norma está asociada a una distancia, luego utilizando los multiplicadores de Lagrange se calcula el gradiente de J con respecto a $u_{ik} \in U$ y $v_i \in V$.

Tomando d_{ik} como la distancia asociada a $\|x_k - v_i\|$ en la ecuación anterior.

Se tiene:

$$J(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2$$

sujeto a:

$$\sum_{i=1}^c (u_{ik}) = 1$$

Utilizando los multiplicadores de Lagrange se obtiene la siguiente expresión:

$$L = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m (d_{ik})^2 - \gamma \left(\sum_{i=1}^c (u_{ik}) - 1 \right)$$

Ahora, en la ecuación anterior se calcula la derivada de L con respecto a $u_{ik} \in U$ y se obtiene:

$$\frac{\partial L}{\partial u_{ik}} = m(u_{ik})^{m-1} d_{ik}^2 - \gamma = 0$$

Por consiguiente:

$$u_{ik} = \left(\frac{\gamma}{m} \right)^{\frac{1}{m-1}} \frac{1}{(d_{ik})^{\frac{2}{m-1}}}$$

Como $\sum_{i=1}^c \mu_{ik} = 1$ se obtiene:

$$\left(\frac{\gamma}{m}\right)^{\frac{1}{m-1}} \sum_{i=1}^c \frac{1}{(d_{ik})^{\frac{2}{m-1}}} = 1$$

Entonces:

$$\left(\frac{\gamma}{m}\right)^{\frac{1}{m-1}} = \frac{1}{\sum_{i=1}^c \frac{1}{(d_{ik})^{\frac{2}{m-1}}}}$$

Reemplazando las dos ecuaciones anteriores se obtiene:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left[\left(\frac{d_{ik}^2}{d_{ij}^2} \right)^{\frac{1}{m-1}} \right]}$$

Nótese que una singularidad se producirá si $d_{ij} = 0$ en la expresión anterior, esto ocurre si, algún punto de los centros de los grupos coincide exactamente con un punto de los datos. Esto se puede evitar al inicio del algoritmo y, en general, no se produciría en la práctica debido a la precisión de la máquina.

Si el conjunto de datos para las agrupaciones fuera del tipo clásico entonces,

$$u_{ik} = 0: \forall i \neq j$$

$$u_{ik} = 1: \forall i = j$$

La actualización de los centros del conglomerado en el paso anterior se obtiene a través de la expresión

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m}; \quad i = 1, 2, \dots, c$$

En efecto, de la ecuación (1) se obtiene el gradiente con respecto a v_i como sigue:

$$J = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|X_k - v_i\|^2$$

Sabiendo que $\|x_k - v_i\|^2$ es igual a $(x_k - v_i)^T(x_k - v_i)$ y este es igual a $x_k^T x_k - 2x_k^T v_i + v_i^T v_i$ entonces se tiene que:

$$\nabla_u J = 2 \sum_{k=1}^n (u_{ik})^m (x_k - v_i) = 0$$

La función de distancia d_{ik} utilizada en este caso es la distancia euclidiana; debido a esto, el FCM solo detecta conglomerados con la misma forma (básicamente esféricos).

Pero, no siempre es conocido el número de conglomerados a elegirse, como en nuestro caso para las variables temperatura, presión, humedad relativa, por lo que recurriremos a la mejora del método del FCM conocido como el método de conglomerados subtractivo (MSCI), de las múltiples mejoras que pueden realizarse al FCM mediante el uso de algoritmos evolutivos¹⁶.

4.4.1. EL MÉTODO DE CONGLOMERADOS SUBTRACTIVO (MSCI)

Este método se propone como una mejora al método del FCM. Se plantea para determinar c en un conjunto de datos, así como un estimado inicial de los centros de cada conglomerado de un conjunto de datos.

Este método se basa en considerar inicialmente cada punto (dato en el conjunto Y) como un centro del conglomerado en potencia. Posteriormente, se define una función que mida para cada punto su potencial P_i como centro del conglomerado; con estos P_i y mediante un proceso iterativo se logra una estimación del número c de conglomerados en X .

A continuación describiremos el algoritmo:

¹⁶ "Algoritmo de agrupamiento genético borroso basado en el algoritmo de las C-Medias difuso". Sergio López. Pág. 38

Se normalizan los datos en X con lo que se restringe la dimensión original de los datos a un hipercubo de igual dimensión, pero teniendo como rango máximo de valores el intervalo $[0,1]$ en cada dimensión.

1. Se calcula el potencial de cada punto x_i

$$P_i = \sum_{i=1}^m e^{-\alpha \|x_i - x_1\|^2}$$

$$\alpha = \left(\frac{2}{r_a}\right)^2$$

donde P_i representa el potencial del i -ésimo punto y r_a es el radio de una vecindad. P_i es una función de las distancias del punto i al resto de los $(m-1)$ puntos en X .

2. Una vez calculado los P_i para cada punto, se selecciona el punto con el mayor P_i como el primer centro del conglomerado, y sus coordenadas como las coordenadas del centro del conglomerado:

$$P_1^* = \max (P_i)$$

3. Se corrige el potencial de cada punto x_i . Para ello, se resta a cada uno una cantidad que es función de su distancia al primer potencial de punto detectado:

$$P_{i-corr} = P_i - P_1^* e^{-\beta \|x_i - x_1\|^2}$$

$$\beta = \left(\frac{2}{r_b}\right)^2$$

Al aplicar la ecuación anterior, los potenciales de los puntos cercanos al primer centro serán los que se vean más fuertemente reducidos.

4. Se detecta el siguiente centro del conglomerado como el punto con el máximo potencial corregido:

$$P_1^* = \max (P_{i-corr})$$

6. Se repiten 3 y 4 de forma continua y según el resultado del siguiente criterio de terminación:

- Si $P_i^* > \bar{\varepsilon}P_i^*$ Se acepta x_i^* y como un centro del conglomerado y se vuelve a 3.

$\bar{\varepsilon}$ representa un umbral sobre el cual se acepta (sin lugar a dudas) un punto con centro del conglomerado. Por el contrario, $\underline{\varepsilon}$ representa un umbral bajo el cual se rechaza (sin lugar a dudas) un punto como centro. La región entre estos 2 valores se evalúa según la parte final del criterio.

- Caso contrario, si $P_i^* < \underline{\varepsilon}P_i^*$ se rechaza x_i^* y se termina el proceso de agrupamiento.
- Caso contrario, sea d_{min} la distancia más corta entre x_i^* y todos los centros de los conglomerados previamente encontrados:

$$\text{Si, } \frac{d_{min}}{r_a} + \frac{P_i^*}{P_1^*} \geq 1$$

Se acepta x_i^* como un centro del conglomerado y se vuelve a 3.

- Caso contrario, se rechaza x_i^* y se fija su potencial en 0. Se selecciona el siguiente punto con P_{i-corr} más alto y se vuelve a evaluar todo el criterio de terminación.
- Fin

A continuación utilizando el modelo difuso propuesto por Brunson (1998) describiremos la metodología que se aplicará a nuestro problema de variables atmosféricas.

Empezamos con el análisis de un modelo simple en el cual solo se tiene una sola variable independiente; entonces, la relación difusa es de la forma:

$$y = f(x) + \varepsilon$$

La relación funcional típicamente conlleva al desconocimiento de los parámetros y el error aleatorio ε . Si el error tiene una media cero, la función difusa representa la media condicional de la variable de salida y ; y el grado de pertenencia que se

obtienen a partir de conglomerados difusos será utilizado como pesos para la evaluación de la estimación de la variable independiente.

Por lo tanto, para obtener nuestro modelo difuso seguimos los siguientes pasos:

1. Particionamos la muestra de observaciones x en c conglomerados difusos, usando el algoritmo FCM mejorado. Este genera valores de membresía para cada valor de x con respecto a cada conglomerado, e implícitamente define una partición para los datos de la variable y .
2. Usando los datos separados de cada conglomerado, ajustamos el modelo:

$$y_{ij} = f_i(x_{ij}) + \varepsilon_{ij} \quad i = 1, \dots, n_i; \quad j = 1, \dots, c$$

3. En particular, si la estimación se realiza mediante mínimos cuadrados ordinarios, entonces:

$$y_{ij} = \beta_{i0} + \beta_{i1}x_{ij} + \varepsilon_{ij} \quad i = 1, \dots, n_i; \quad j = 1, \dots, c$$

4. Utilizamos la media condicional para el modelo y predicción:

$$\hat{y}_k = \frac{\sum_{i=1}^c (b_{i0} + b_{i1}x_k)u_{ik}}{\sum_{i=1}^c u_{ik}}; \quad k = 1, \dots, n$$

Donde u_{ik} es el grado de membresía del k -ésimo valor de x en el i -ésimo conglomerado difuso, y los b_{im} estimaciones de β_{im} ($m = 0,1$) mediante mínimos cuadrados, utilizando la i -ésima partición difusa de la muestra.

5. Se calcula la relación predictiva entre la variable independiente y la derivada de la media condicional de y :

$$\left(\frac{\delta \hat{y}_k}{\delta x_k} \right) = \frac{\sum_{i=1}^c (b_{i1}u_{ik})}{\sum_{i=1}^c u_{ik}}; \quad k = 1, \dots, n$$

La derivada de la media condicional de y , con respecto a la variable independiente x , también tiene una estructura de ponderación promedio difusa.

Como se observa en la ecuación anterior el predictor difuso de la media condicional de y es un promedio ponderado del predictor lineal basado en la partición difusa de la variable exploratoria, con los pesos variando continuamente por muestra, esta última característica permite que el modelado no lineal sea

efectivo. Se puede ver que en cada conglomerado difuso incluye la lógica difusa de la forma: SI los datos de entrada se encuentra en esta región ENTONCES este es probable el predictor de la variable de salida.

4.4.2 GENERALIZACIÓN DEL MODELO DIFUSO

En general, el modelo puede tener más de una variable explicativa y se muestra en la siguiente ecuación:

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon$$

Los pasos para el modelo difuso generalizado son:

1. Se particiona la muestra de n observaciones para cada x_r en c_r conglomerados difusos (donde $r = 1, 2, \dots, p$) usando el algoritmo del FCM mejorado. Este genera valores de membresía para cada valor de x con respecto a cada conglomerado.

2. Se consideran todas las c posibles combinaciones de los conglomerados difusos con las p variables de entrada, donde:

$$c = \prod_{r=1}^p c_r$$

Se descarta cualquier combinación cuya matriz X no sea de rango completo, entonces, el número de posibles conglomerados será: \hat{c}

3. Usando los datos de cada uno de estos \hat{c} conglomerados difusos, se ajusta el modelo:

$$y_{ij} = \beta_{i0} + \beta_{i1}x_{1ij} + \dots + \beta_{ip}x_{pij} \quad j = 1, \dots, n; \quad i = 1, \dots, \hat{c}$$

4. Utilizamos la media condicional para el modelo y predicción:

$$\hat{y}_k = \frac{\sum_{i=1}^{\hat{c}} (b_{i0} + b_{i1}x_{1k} + \dots + b_{ip}x_{pk})w_{ik}}{\sum_{i=1}^{\hat{c}} w_{ik}}; \quad k = 1, \dots, n$$

$$\text{Donde, } w_{ik} = \prod_{r=1}^p \delta_{ij} \mu_{rjk}; \quad 1, \dots, \hat{c}$$

δ_{ij} es un selector que cambia el valor de membresía para el j-ésimo conglomerado difuso (para la r-ésima variable independiente). Si ese grupo está relacionado con i-ésima combinación del conglomerado ($i = 1, 2, \dots, c$) y los b_{im} son estimadores de β_{im} ($m = 0, 1$) mediante mínimos cuadrados utilizando la i-ésima participación difusa de la muestra.

1. Se calcula la relación predictiva entre la variable independiente y la derivada de la media condicional de y

$$\left(\frac{\delta y_k}{\delta x_{rk}} \right) = \frac{\sum_{i=1}^c (b_{ir} u_{rik})}{\sum_{i=1}^c u_{ik}}; r = 1, \dots, p; k = 1, \dots, n$$

4.4.3 IMPLEMENTACIÓN DEL MÉTODO DE LAS C-MEDIAS DIFUSO MEJORADO A LA METEOROLOGÍA

Al implementar el método de las c-medias difuso mejorado con aplicaciones en la Meteorología en el proyecto PGA, se utilizó los software Matlab 9.3 y Spss 17.0 (Apéndice A9) para el análisis de los conglomerados (El análisis de conglomerados se encuentra implementado en el cuadro 4.1 y en el cuadro 4.2). En el lenguaje de Matlab el entorno informático de análisis numérico y programación nos permitió implementar el método de las c-medias incluso utilizando las herramientas de Matlab para simulación y la inferencia estadística; Gráfico 4.4 y Gráfico 4.5. y a su vez su código de implementación en este programa se encuentra en la página 128.

Al obtener el número de conglomerados en SPSS 17.0 nos presenta como salida los centros de los conglomerados y para obtener el número de conglomerados a estimar y el grado de pertenencia de cada observación a cada conglomerado se realizó de la siguiente manera:

1. Para la obtención del número de conglomerados por cada variable: la presión atmosférica, temperatura y humedad relativa, se utilizó la función SUBCLUST, aquí está implementado el método Substractivo. Sección 4.4.1.

Número de casos en cada conglomerado

Conglomerado	1	2074.000
	2	7430.000
	3	2581.000
	4	3562.000
Válidos		15647.000
Perdidos		.000

Cuadro 4.1 Total de datos para las variables meteorológicas

- Al tener el número de conglomerados se aplica la función FCM, la cual utiliza el algoritmo c-medias difuso (sección 4.3.1) para los grados de pertenencia, que tiene cada observación, en cada uno de los conglomerados. En el cuadro 4.1 se muestra el total de 15.647 datos para cada una de las variables meteorológicas.
- Al tener los conglomerados y su grado de pertenencia se procede a la estimación del modelo de regresión difusa (sección 4.4.2) y se lo programa en MATLAB.

Posteriormente realizamos las comparaciones con el modelo de regresión lineal múltiple que se toma como variable dependiente a la altura y la presión atmosférica, temperatura y humedad relativa como variables independientes (sección 2.7). Se obtuvo el mejor modelo siguiente:

Coefficientes^a

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
1 (Constante)	33694,352	112,493		299,524	,000
PRESIÓN (hPa)	-58,491	,266	-1,663	-219,757	,000
TEMPERATURA (°C)	158,659	1,498	,632	105,886	,000
HUMEDAD RELATIVA (%)	43,184	,996	,188	43,367	,000

a. Variable dependiente: ALTURA (m)

Cuadro 2.29 Modelo obtenido de regresión lineal múltiple

$$\hat{y} = 33.694.352 - 58,491(\text{presión}(hPa)) + 158,659(\text{temperatura}(^{\circ}C)) + 43,18(\text{humedad}(\%))$$

La suma de los errores al cuadrado es 7.49×10^{11} (cuadro 2.30)

ANOVA^b

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
1 Regresión	7,490E11	3	2,497E11	57520,134	,000 ^a
Residual	6,789E10	15643	4340228,187		
Total	8,168E11	15646			

a. Variables predictoras: (Constante), HUMEDAD RELATIVA (%), TEMPERATURA (°C), PRESION (hPa)

b. Variable dependiente: ALTURA (m)

Cuadro 2.30 ANOVA obtenido de variables atmosféricas

Ahora, se estimará la relación de las variables meteorológicas mediante el uso de la regresión difusa y se estima el número de conglomerados mediante el uso del algoritmo Subtractivo (sección 4.4.1) y se obtuvo los siguientes resultados Cuadro 4.2:

Centros iniciales de los conglomerados

	Conglomerado			
	1	2	3	4
PRESIÓN (hPa)	753.00	13.70	494.60	257.00
TEMPERATURA (°C)	12.75	-47.05	2.75	-32.35
HUMEDAD RELATIVA (%)	89.00	1.00	60.00	15.00

Centros de los conglomerados finales

	Conglomerado			
	1	2	3	4
PRESIÓN (hPa)	624.68	60.91	414.43	224.86
TEMPERATURA (°C)	6.18	-65.76	-12.30	-46.09
HUMEDAD RELATIVA (%)	85.19	7.74	43.56	19.07

Cuadro 4.2 Estimación de los conglomerados para las variables meteorológicas

Luego de obtener el número de conglomerados mediante el algoritmo subtractivo se procede a calcular el grado de membresía para las distintas variables meteorológicas, mediante el algoritmo C- medias difuso el cual toma como entrada el número de grupos estimados en el cuadro 4.1, obteniendo como salida los grados de membresía de cada observación a cada conglomerado. Gráfico 4.4.

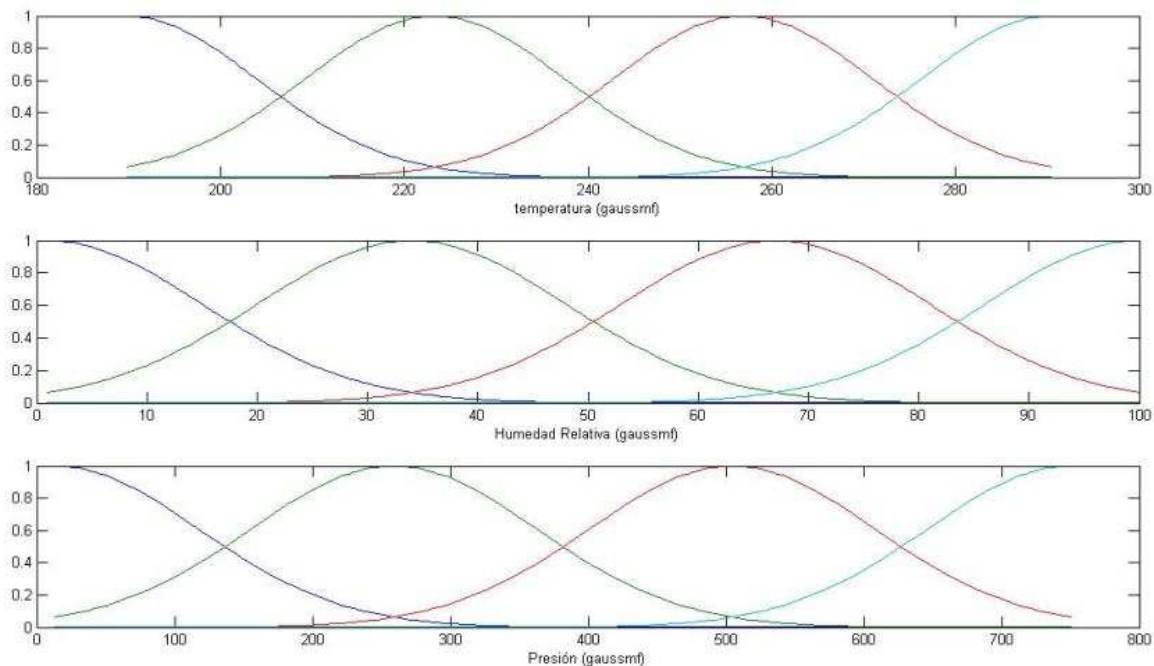


Gráfico 4.4. Grados de Membresía de las variables atmosféricas

Siguiendo los pasos de acuerdo al algoritmo del modelo difuso se combinan cada conglomerado de las variables, temperatura y humedad con cada conglomerado de la variable presión y así se obtiene la matriz X, la que será utilizada para estimar los parámetros β mediante la regresión lineal múltiple. Posteriormente, se eliminarán aquellas matrices que no son de rango completo.

Para esta estimación se generaron 6 matrices X con los datos de las variables temperatura, presión, humedad relativa y la variable dependiente la altura el cuadro 4.3 muestra los coeficientes obtenidos por la combinación 1:

Coefficientes^a

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	22067.390	62.227		354.629	.000
	PRESIÓN (hPa)	-32.913	.223	-.926	-147.599	.000
2	(Constante)	-8435.749	604.373		-13.958	.000
	PRESIÓN (hPa)	-51.340	.402	-1.445	-127.722	.000
	TEMPERATURA (°K)	152.137	3.005	.573	50.628	.000
3	(Constante)	-13811.644	537.001		-25.720	.000
	PRESIÓN (hPa)	-37.155	.518	-1.045	-71.698	.000
	TEMPERATURA (°K)	177.321	2.659	.667	66.693	.000
	HUMEDAD RELATIVA (%)	-98.798	2.702	-.509	-36.559	.000

a. Variable dependiente: ALTURA (m)

El modelo es:

$$\hat{y} = -13811.644 - 37.155(\text{presión}(hPa)) + 177.321(\text{temperatura}(^{\circ}C)) - 98.798(\text{humedad}(\%))$$

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.925 ^a	.855	.855	2751.164
2	.952 ^b	.907	.907	2204.933
3	.958 ^c	.917	.917	2083.321

a. Variables predictoras: (Constante), PRESIÓN (hPa)

b. Variables predictoras: (Constante), PRESIÓN (hPa), TEMPERATURA (°C)

c. Variables predictoras: (Constante), PRESIÓN (hPa), TEMPERATURA (°C), HUMEDAD RELATIVA (%)

Cuadro 4.3. Combinación 1, obtención del modelo difuso

A continuación, el cuadro 4.4 muestra las 6 posibles combinaciones obtenidas con las variables meteorológicas:

MATRIZ X	CONSTANTE	PRESIÓN	TEMPERATURA	HUMEDAD	R²
Combinación 1	-13811,644	-37,155	177,321	-98,798	0,917
Combinación 2	34623,499	-58,491	158,659	43,184	0,939
Combinación 3	36268,003	-58,898	189,44	13,436	0,926
Combinación 4	35334,946	-65,57	178,067	100,213	0,956
Combinación 5	10028,556	-100,663	-8,025	4,883	0,983
Combinación 6	9859,187	-92,77	-7,596	0,9566	0,997

Cuadro 4.4. Estimaciones de los coeficientes por cada combinación de variables

Las combinaciones anteriores nos genera 6 modelos difusos múltiples y el coeficiente de determinación es alto por lo que en cada combinación se puede afirmar que los datos observados pueden explicarse por el modelo difuso con una probabilidad del 99% de confiabilidad. Los modelos son:

$$\hat{y} = -13811.644 - 37.155(\text{presión}(hPa)) + 177.321(\text{temperatura}(^{\circ}C)) - 98.798(\text{humedad}(\%))$$

$$\hat{y} = 34623.499 - 58.491(\text{presión}(hPa)) + 158.659(\text{temperatura}(^{\circ}C)) + 43.184(\text{humedad}(\%))$$

$$\hat{y} = 36268.003 - 58.898(\text{presión}(hPa)) + 189.44(\text{temperatura}(^{\circ}C)) + 13.436(\text{humedad}(\%))$$

$$\hat{y} = 35334.946 - 65.57(\text{presión}(hPa)) + 178.067(\text{temperatura}(^{\circ}C)) + 100.213(\text{humedad}(\%))$$

$$\hat{y} = 10028.556 - 100.663(\text{presión}(hPa)) - 8.025(\text{temperatura}(^{\circ}C)) + 4.883(\text{humedad}(\%))$$

$$\hat{y} = 9859.187 - 92.77(\text{presión}(hPa)) - 7.596(\text{temperatura}(^{\circ}C)) + 0.9566(\text{humedad}(\%))$$

Finalmente, con cada uno de los coeficientes estimados y utilizando el grado de membresía o pertenencia de cada individuo con respecto a cada conglomerados se procede a calcular la media condicional de la variable dependiente, la altura. Y, a continuación graficamos el modelo de regresión lineal múltiple obtenido en el cuadro 2.29 y el modelo difuso obtenido en el cuadro 4.4.

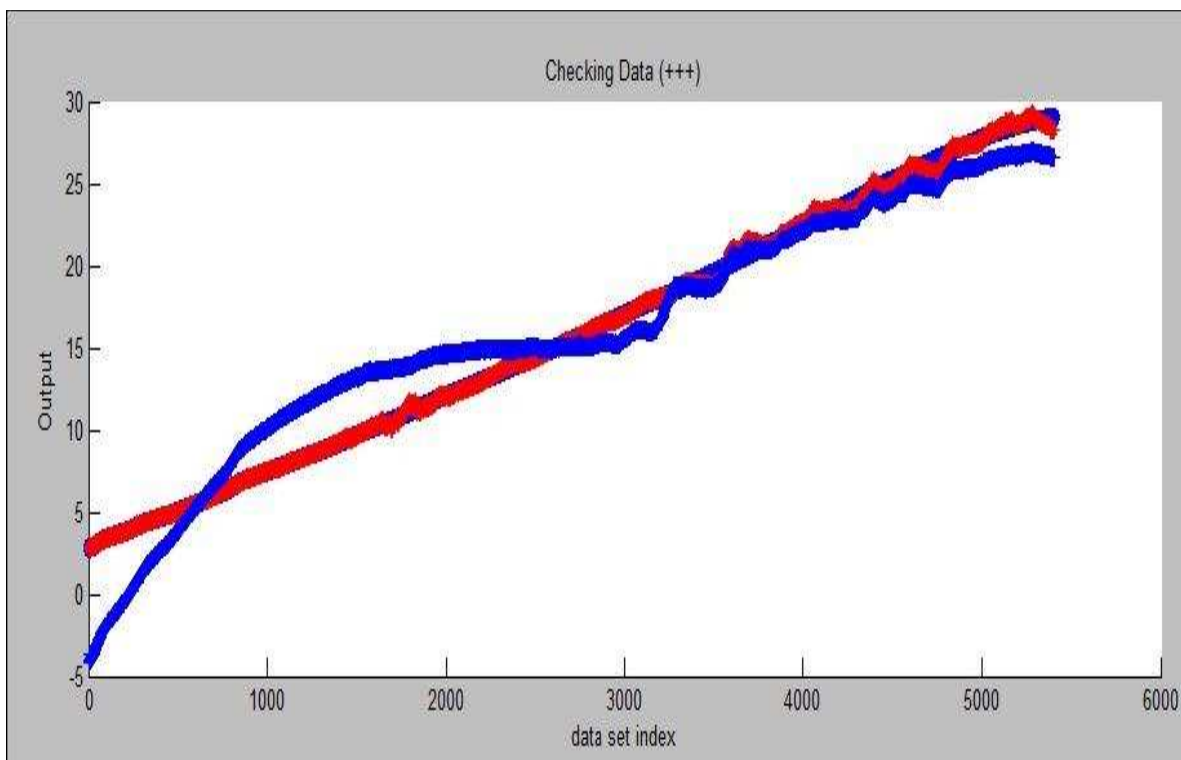


Gráfico 4.5. Regresión múltiple vs Regresión Difusa (C-Medias Mejorado)

El gráfico 4.5 muestra, en el eje x, el total de datos utilizados para el análisis y el eje y (output) muestra la altura. La gráfica de color azul, muestra la curva obtenida por el modelo de regresión lineal múltiple de las variables presión atmosférica, temperatura y humedad relativa obtenido en el cuadro 2.29; pág. 80 y la curva de color roja muestra el ajuste de la curva para las mismas variables atmosféricas pero, utilizando el método de las c-medias difuso mejorado.

A continuación presentamos un cuadro comparativo de análisis de resultados:

Total de datos reales para el análisis	Técnica de aproximación	Suma de residuos al cuadrado
15647 datos	regresión lineal múltiple	6.789×10^{10}
15647 datos	método econométrico difuso de las c-medias	6.459×10^9
15647 datos	método econométrico difuso de las c-medias mejorado	6.093×10^9

Cuadro 4.5 Cuadro comparativo de resultados

4.4.4 CONCLUSIONES

Al realizar la estimación por medio de la regresión lineal múltiple, se obtiene una suma de residuos al cuadrado de 6.789×10^{10} , mientras que con el modelo difuso se obtiene una suma de residuos al cuadrado de 6.459×10^9 por lo que el ajuste del modelo difuso es mejor que el modelo de regresión lineal múltiple. Sin embargo, se pudo mejorar el ajuste del modelo utilizando el método de la c-medias difuso mejorado y nos da un ajuste de 6.093×10^9 y este ajuste es mucho mejor que los anteriores.

A continuación se presenta un resumen de las 15647 observaciones y los datos de las variables meteorológicas obtenidas por el modelo difuso:

OBSERVACIÓN	ALTURA (m)	PRESIÓN ATMOSFÉRI CA (hPa)	TEMPE RATURA (°C)	HUMEDAD RELATIVA (%)	R ²
315	3224	693,5	18,25	92	0,734
630	3669	655,2	7,95	81	0,897
945	4117	620,4	5,05	88	0,944
1260	4568	592,1	10,85	88	0,967
1575	4984	563,2	8,65	78	0,978
1890	5445	526,2	-2,45	88	0,984
2205	5885	513,8	-4,85	83	0,987
2520	6300	479,6	0,95	50	0,989
2835	6769	444,9	-7,65	8	0,991
3150	7217	427,4	-3,55	3	0,992
3465	7645	408,5	-13,45	55	0,992
3780	8092	385,6	-17,65	85	0,992
4095	8532	362,8	-19,95	54	0,992
4410	8993	341	-23,15	52	0,992
4725	9461	310,5	-27,55	1	0,992
5040	9926	290,8	-31,45	6	0,992
5355	10399	280,2	-34,45	31	0,992
5670	10855	254,7	-37,85	12	0,993
5985	11312	238,2	-42,15	18	0,993
6300	11774	222,3	-46,05	13	0,993
6615	12243	213,5	-48,25	6	0,994
6930	12692	199,3	-51,85	6	0,994
7245	13153	179,8	-56,55	6	0,994
7560	13630	166,7	-60,25	6	0,994
7875	14115	159	-64,65	20	0,995
8190	14590	149,2	-63,05	12	0,995
8505	15069	135,6	-71,95	28	0,995
8820	15552	120,9	-76,15	15	0,995
9135	16035	111,1	-80,35	16	0,995
9450	16523	105,3	-80,15	31	0,995

Cuadro 4.6 Resumen de resultados obtenido por el modelo difuso

El cuadro 4.7 muestra un resumen de los valores reales obtenido por las radio sondas.

OBSERVACIÓN	ALTURA (m)	PRESIÓN ATMOSFÉRI CA (hPa)	TEMPE RATURA (°C)	HUMEDAD RELATIVA (%)
315	3224	691,7	18,35	92
630	3669	676,5	5,75	100
945	4117	623,9	13,45	81
1260	4568	586,5	2,75	81
1575	4984	575,3	-1,05	96
1890	5445	531,4	6,65	67
2205	5885	512,3	-3,95	81
2520	6300	486,1	-6,85	84
2835	6769	458,6	-9,65	79
3150	7217	427	-3,65	3
3465	7645	408,5	-13,45	55
3780	8092	385,1	-17,75	85
4095	8532	363,1	-20,55	77
4410	8993	362,7	-20,55	77
4725	9461	341,1	-23,85	80
5040	9926	319,8	-27,35	81
5355	10399	299,4	-30,65	49
5670	10855	280	-34,45	30
5985	11312	262,1	-38,35	18
6300	11774	245,1	-41,35	12
6615	12243	228,9	-44,45	11
6930	12692	213,1	-49,85	16
7245	13153	199,2	-51,95	6
7560	13630	185,3	-55,75	6
7875	14115	171,5	-60,75	15
8190	14590	158,8	-64,75	20
8505	15069	146,8	-68,25	22
8820	15552	135,5	-72,05	28
9135	16035	124,5	-74,85	20
9450	16523	114,7	-78,35	30

Cuadro 4.7. Datos reales obtenidos por las radiosondas

CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

1. Se realizó un estudio de prefactibilidad de uso de la lógica difusa y sus aplicaciones, para esto se comenzó observando que una de sus principales propiedades es el poder trabajar con incertidumbre en los datos atmosféricos y cuando las variables no se pueden ajustar usando métodos clásicos. Se utilizó la regresión difusa analizando cada uno de los métodos y seleccionando el método de dos pasos difuso que proporcionó un alto grado de confiabilidad de ajuste R^2 para cada una de la modelización de las variables. Sin embargo, el análisis estadístico nos permitió dar un mejor alcance de cómo trabajar con las variables atmosféricas en el lugar de interés para el PGA.
2. Se centró luego el análisis en el estudio de las funciones de pertenencia para conjuntos difusos, los diferentes tipos y cualidades de cada uno de ellos. Se analizó la forma en la que se las utiliza para tratar de determinar la más adecuada para nuestro estudio, que fue la utilización de las funciones triangulares.
3. Fue necesario implementar algoritmos que se utilizan para predecir eventos cuando se dispone de datos inciertos. Se analizó la forma de trabajar de los algoritmos clásicos en predicción de variables atmosféricas, encontrando la mejor estimación curvilínea utilizando funciones lineales, exponenciales, logarítmicas y cuadráticas y la utilización de los modelos ARIMA con un 95% de confianza en el eje vertical. Se encontró el mejor modelo de ajuste con un alto coeficiente de determinación R^2 que indica el porcentaje de ajuste de los datos ajustados con los datos observados. Posteriormente, se decidió iniciar el trabajo con el Algoritmo de Mínimos Cuadrados para la regresión difusa que proporciona un buen ajuste a los datos y al final utilizamos el algoritmo de conglomerados difuso c medias para minimizar los problemas con el tratamiento de grandes cantidades de información.

4. Los programas computacionales (software) SPSS17, MINITAB, VAISALA y MATLAB fueron los más adecuados para visualizar el uso de los algoritmos de predicción y análisis estadístico así como la factibilidad de trabajar en una simulación a gran escala obtenido por los datos de radio sondeos lanzados en la ciudad de Ambato.
5. Para realizar un análisis estadístico de variables atmosféricas fue indispensable trabajar con los datos obtenidos en el lugar y a la hora adecuada de interés del proyecto debido a que nos muestran un análisis más minucioso y detallado de la zona para poder visualizar el comportamiento y poder implementar pruebas de hipótesis para las variables atmosféricas.
6. Los pronósticos de los principales parámetros meteorológicos se presentan generalmente en forma de intervalos con un 95% de confianza, es decir un valor mínimo y un valor máximo, usando técnicas regresivas y autorregresivas.
7. El ensayo de la regresión difusa lineal en problemas meteorológicos es algo novedoso en nuestro país, dando resultados satisfactorios, principalmente en los perfiles verticales de la presión y de la temperatura, que explican un comportamiento correlacionado con la altura. Los modelos, en cierta forma, corroboran lo que la teoría meteorológica afirma.
8. El comportamiento tanto de la velocidad y dirección del viento es bastante caótico. Fundamentalmente para estos parámetros se aprecia el comportamiento incierto de la atmósfera, lo que obligó a estudiar modelos dinámicos como el WRF e implementarlo a la velocidad y direcci[on del viento. Algo similar ocurre con la humedad, que luego se estabiliza pasados los 20 kilómetros sobre el nivel del mar, sin embargo pudo encontrarse a través de un método autorregresivo ARIMA el mejor ajuste con un 95% de confianza.

9. Los modelos de los perfiles verticales expuestos únicamente pueden ser utilizados en la zona local estudiada y sus pronósticos confiables para horizontes de tiempo de 24 horas. Los modelos dinámicos, a su vez, servirán para hacer pronósticos en regiones más grandes y para horizontes de tiempo más largos.

10. La elaboración de estos modelos estadísticos, que son limitados, ha ayudado para entender en mejor el comportamiento multidimensional de la atmósfera y encontrando la correlación que existe entre estas variables se pudo obtener un modelo de regresión lineal múltiple del perfil atmosférico, con relación a la presión atmosférica, temperatura y humedad relativa siguiendo una metodología eficiente.

11. El carácter caótico del comportamiento del tiempo se puede apreciar con las permanentes informaciones recibidas sobre el apareamiento de ciclones y su posterior cambio de rumbo, cambio de velocidad y aumento o disminución de su severidad.

12. El algoritmo de la c-medias difuso mejorado fue oportuno para el análisis del comportamiento no lineal de los datos atmosféricos y los agrupe dentro de conglomerados dentro de los cuales el comportamiento de los datos sean lineales.

13. El método de las c-medias difuso mejorado puede ser utilizado como una técnica de suavización de datos y ajuste de relación multivariante de datos. En contraste la regresión difusa puede ser utilizada para modelos complejos y como se analizó para una carga computacional no tan desarrollada dependiendo de la alta cantidad de datos.

5.2. RECOMENDACIONES

1. Para elaborar modelos estadísticos, la premisa fundamental es disponer de abundantes datos de la misma especie, capturados bajo las mismas condiciones, con el fin de realizar operaciones con elementos comparables, de lo contrario arrojan resultados, muchos de ellos, sin sentido.
2. Para la modelización de otras variables atmosféricas como la radiación solar no se ha podido ejecutar por la ausencia de datos de las radio sondas. Sin embargo, se recomienda que se disponga de una estación meteorológica a nivel de superficie que sirva para modelizar y reparametrizar las condiciones para un modelo atmosférico global con múltiples variables atmosféricas.
3. El permanente cambio climático que ocurre en todas las regiones del planeta obliga a estudiar la atmósfera que nos rodea, con el fin de mitigar los efectos catastróficos que producen los eventos extremos producidos por los fenómenos meteorológicos.
4. El estudio la aplicación de algunos modelos en base a redes neuronales, permitirán realizar pronósticos del tiempo mejores que los estadísticos propuestos, pero a nivel de la superficie terrestre y que se puedan también comparar con los resultados que se provea un modelo global.
5. Desde el inicio de todas las actividades investigativas es sumamente conveniente contar con asesoría experimentada, en especial en tópicos sobre los cuales ya trabajan en otros centros del mundo y que cuentan con mayor tradición y cultura de investigación, así como las aplicaciones respectivas; esta colaboración, como se puede apreciar, ahorra recursos de toda índole.

6. Es imprescindible disponer de jóvenes estudiantes de los últimos semestres de diversas carreras para la construcción de un modelo global climatológico, lo que sin lugar a dudas, garantizará la conformación de grupos de investigadores ecuatorianos altamente capacitados, que podrán colaborar en amplios equipos multidisciplinarios de investigación, ya sea a nivel nacional o internacional.

7. En relación a la interpretación de los coeficientes difusos estimados para las variables atmosféricas queda pendiente para el futuro determinar la relación de cada variable de entrada con el número difuso estimado con restricciones posibilísticas.

8. En forma permanente discutir el planteamiento de proyectos paralelos que permitan diversificar las aplicaciones de los resultados esperados del proyecto principal.

9. Realizar alianzas estratégicas con otros grupos de investigación nacionales o internacionales con el fin de intercambiar resultados y experiencias en la ejecución de los mismos.

BIBLIOGRAFÍA

1. Asli Celikyilmaz y I. Burhan Turksen, "Modeling Uncertainty with fuzzy logic". Pais: Estados Unidos. Año 2009
2. Timothy J. Ross, "Fuzzy Logic with Engineering Application", Published by John Wiley & Sons. Ltd. Pais: Estados Unidos. Año 2004
3. Buckley J. J., Eslami E., Feuring T. Fuzzy Mathematics in Economics and Engineering, Heidelberg: Phisica-Verlag, Pais: Estados Unidos. Año 2002.
4. Modelos de Regresión Difusa con Aplicaciones. Proyecto Previo a la obtención de Ingeniero Matemático. Jaime Mauricio Figueroa. Abril 2001.
5. Dr. S.N. Sivanandam, Dr. S. Sumathi "Introduction to Fuzzy Logic using MATLAB." Editorial Springer. Pais: India. Año: 2007.
6. Tanaka H, and Asaid K. Linear regression model. Fuzzy Sets and System. Pais: Estados Unidos. Año 1991.
7. Clustering Methods in c-Means Clustering with Applications, Springer, Berlin. Pais: Estados Unidos. Año 2008
8. Cesar Pérez. "Econometría de las Series Temporales. Editorial Pearson Prentice Hall. Pais: España. Año 2006.
9. Econometric Modelling based on Pattern Recognition with the Fuzzy C-Means Clustering Algoriting. David Giles y Robert Draeseke.
<http://web.uvic.ca/econ/research/papers/ewp0101.pdf>
10. Análisis y Diseño de Experimentos. Humberto Guitierrez – Román De la Vara Salazar. Editorial Mc Graw Hill. . Pais: España. Año 2006
11. Técnicas Estadística con SPSS. Cesar Pérez. Editorial Prentice Hall. Pais: España. Año 2003.
12. Algoritmo de Agrupamiento Genético Borroso Basado en el algoritmo de las C-Medias Borroso. Sergio López García. Pais: España. Año 2001.

ANEXOS A : CONJUNTOS DIFUSOS VERSUS CONJUNTOS REALES

A.1 DEFINICIÓN DE CONJUNTO DIFUSO Y REALES

Un conjunto difuso puede definirse completamente por su función de pertenencia:

$$\mu_A : U \longrightarrow [0,1]$$

$$x \longrightarrow \mu_A(x)$$

$$A = \{(x, \mu_A(x)) : x \in U, \mu_A(x) \in [0,1]\}$$

Donde U es el conjunto universo. Como se puede observar para un conjunto difuso, la pertenencia de un elemento al conjunto no es cuestión de todo o nada, sino que son posibles diferentes grados de pertenencia. La función de pertenencia puede tomar cualquier valor en el intervalo real $[0,1]$. Lofti A. Zadeh, fue el primer investigador de la lógica difusa a mediados de los años sesenta en la Universidad de Berkeley(California) para ilustrar el concepto de conjunto difuso, tomó el conjunto representado por "hombres altos". En la lógica clásica el conjunto anterior es un conjunto al que pertenecerían los hombres con una estatura mayor a un cierto valor, por ejemplo 1.80 metros y cualquier individuo que sea menor a este valor quedarían fuera del conjunto. En este ejemplo se ilustra la definición de conjunto difuso ya que no parece lógico decir que un individuo es alto y otro no lo es cuando su altura difiere en pocos centímetros para aquello se asigna a cada valor de la altura un grado de pertenencia al conjunto, entre 0 y 1. Así por ejemplo, un hombre que mida 1.79 metros podría pertenecer al conjunto difuso "hombres altos" con un grado de pertenencia 0.8, una persona que mida 1.81 metros con un grado 0.85, y uno que mida 1.50 metros con un grado 0.1. Ver gráfico A.2.

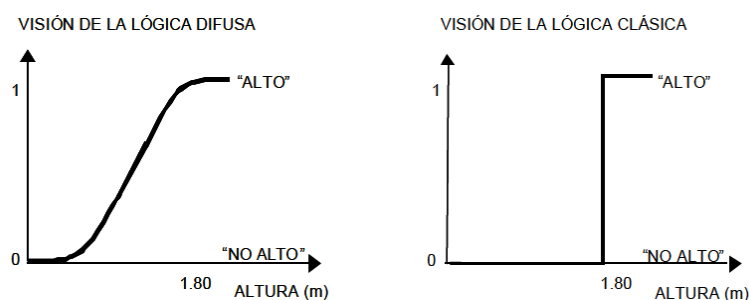


Gráfico A.1. Lógica clásica y lógica difusa

Es así, que a los conjuntos difusos pueden ser considerados como una generalización de los conjuntos clásicos: la teoría clásica de conjuntos solo contempla la pertenencia o no pertenencia de un elemento a un conjunto, sin embargo, la teoría de conjuntos difusos establece la pertenencia parcial de un elemento a un conjunto, es decir, se asigna a cada elemento un grado de pertenencia a un conjunto difuso que puede tomar cualquier valor entre 0 y 1. Este grado de pertenencia se define mediante la función característica asociada al conjunto difuso: para cada valor que pueda tomar un elemento o variable de entrada x la función característica $\mu(x)$ proporciona el grado de pertenencia de este valor de x al conjunto difuso A . Zadeh también considera que las etiquetas lingüísticas pueden clasificarse en dos categorías que informalmente se definen como sigue:

- **Tipo I:** las que pueden representarse como operadores que actúan en un conjunto difuso: "muy", "más o menos", "mucho", "ligeramente", "altamente", "bastante", etc. y,
- **Tipo II:** las que requieren una descripción de cómo actúan en los componentes del conjunto difuso (operando): "esencialmente", "técnicamente", "estrictamente", "prácticamente", "virtualmente", etc...

En otras palabras, las etiquetas lingüísticas pueden ser caracterizadas como operadores más que construcciones complicadas sobre las operaciones primitivas de conjuntos difusos.

La aplicación a nuestro estudio se basará en las funciones de pertenencia de tipo-1.

Cuando se escucha que el estado del cielo, con respecto a la presencia de nubes, tiene los "valores": *despejado, parcialmente nublado, nublado o totalmente nublado*, no nos imaginamos en función de qué referencias se pueden determinar estos valores. Más o menos se entiende en términos generales el razonamiento humano subyacente. Sin embargo, los meteorólogos, con el fin de que el común de las personas pueda interpretar esa información, han establecido una escala de niveles, de la variable lingüística "estado del cielo". Con este fin han creado la unidad **octa**, que resulta de catalogar a la cobertura del cielo por nubes en ocho

octavos, por lo que se tiene como resultado los valores: 0, 1, 2, 3, 4, 5, 6, 7, y 8 octas. Significando lo que contiene el siguiente cuadro:

	NIVEL	MEDIDA EN OCTAS	DESCRIPCIÓN
[Barra azul]	DESPEJADO	0	Totalmente despejado
		1	Cubierto menos de 1/8
[Barra azul claro]	PARCIALMENTE NUBLADO	2	
		3	
		4	
[Barra gris claro]	NUBLADO	5	
		6	
[Barra gris medio]	NUBLADO	7	Cubierto menos de 8/8
		8	Totalmente cubierto

Cuadro A.1 COBERTURA DEL CIELO

Como se puede apreciar, éste es un ejemplo típico de lo que es un conjunto difuso, borroso o nebuloso, tal como se indica en el siguiente gráfico, donde están representadas las funciones de pertenencia de cada uno de los niveles de cobertura del cielo por nubes:

En el eje x está representada la cobertura del cielo por nubes en octas.

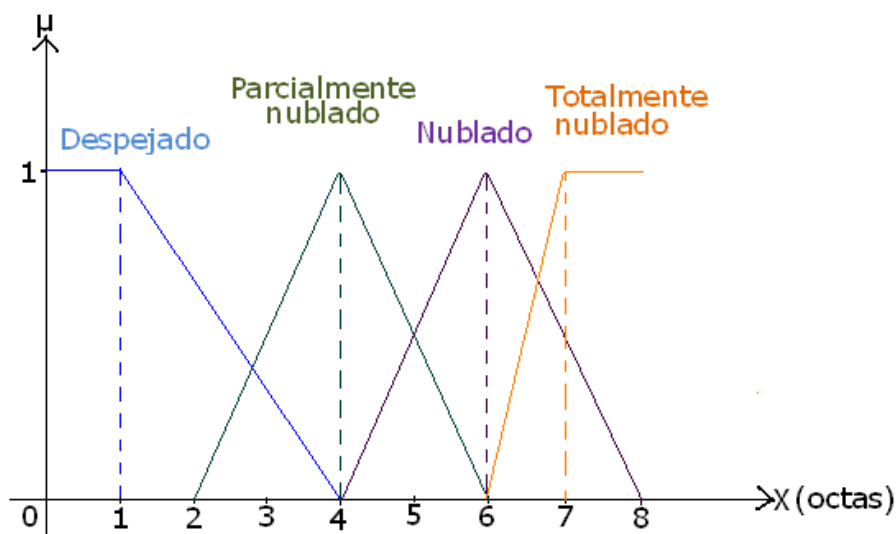


Gráfico.A.2. Funciones de pertenencia de conjuntos difusos en meteorología

En los conjuntos difusos como se vio anteriormente suavizamos la restricción de que la función de pertenencia valga 0 ó 1, y dejamos que tome valores en el intervalo [0,1]. La necesidad de trabajar con conjuntos difusos surge de un hecho: Hay conceptos que no tienen límites claros como en las variables atmosféricas: ¿La presión de la atmósfera a 350hPa, es baja? ¿Una temperatura de 15 °C es media?

La notación habitual para los conjuntos difusos es la definida por Lofti Zadeh, que es la siguiente:

Sea A un conjunto difuso definido sobre el universo U:

$$A = \{(x, \mu_A(x)) / x \in U\}$$

Que indica que A está formado por todos los pares ordenados x y es el resultado de la función de pertenencia para todo elemento u, dentro del universo de discurso U. Para denotar el conjunto difuso A:

Si el universo es discreto: $\sum_U u_A(x) / x$

Si el universo es continuo: $F = \int_U u_A(x) / x$

Observación: El sumatorio o la integral pierden su significado habitual. En lógica difusa quieren simbolizar una mera enumeración de tuplas. La barra tampoco indica una fracción sino simplemente separa los dos elementos de la tupla.

Ejemplo: Como se definiría el conjunto difuso discreto “Tirada alta del dado”.

$$F = \{0/1 + 0/2 + 0.3/3 + 0.6/4 + 0.9/5 + 1/6\}$$

La parte derecha de la tupla indica el elemento y la parte izquierda el grado de pertenencia.

A.3 PRINCIPIO DE EXTENSION

El principio de extensión es una de las principales herramientas en la teoría de conjuntos difusos, ya que hace posible la aplicación de principios matemáticos no difusos al tratamiento de cantidades difusas; la aplicación del álgebra clásica a números difusos lo hace esencialmente útil en los cálculos difusos.

Sea $y = f(x)$, donde $x \in X$ e $y \in Y$, si tenemos una aplicación lineal de X en Y , fácilmente la imagen de esta aplicación puede ser encontrada; en cambio si tenemos un conjunto difuso $A \in X$ donde la pregunta viene a ser ¿Cuál es la imagen difusa de A por medio de f ?

Si llamamos B a la imagen, es evidente que el soporte de B debería ser la imagen del soporte de A .

Si fuese una relación biunívoca no se tendría ningún problema, el hecho es que los valores de $y \in B$ es posible que provengan de más de un x original, por esta razón Zadeh propuso el siguiente valor para la función de pertenencia de B .

$$\mu_B(y) = \begin{cases} \sup_{x \in X: f^{-1}(y)=x} \mu_A(x), & \text{si: } f^{-1}(y) \neq \Phi \\ 0 & \text{caso contrario} \end{cases}$$

Este resultado puede ser interpretado como la unión a través de todos los x que constituyen el conjunto difuso cuyo soporte está formado por un único punto, los cuales $f(x) = y$, con grado de pertenencia $\mu_A(x)$. Ejemplo:

Sea la función $y = x^2 + 3$, y el conjunto difuso S con la siguiente función de pertenencia:

$$\mu_S(x) = \begin{cases} \frac{x+1}{3} & -1 \leq x \leq 2 \\ 3-x & 2 \leq x \leq 3 \\ 0, & \text{caso contrario} \end{cases}$$

El procedimiento de obtención del conjunto imagen, para algunos valores concretos del conjunto difuso S , se presenta en el siguiente cuadro A.3.

x	$\mu_S(x)$	y	$\mu_B(x)$
-1	0	4	0,67
-0,75	0,083	3,56	0,58
-0,5	0,17	3,25	0,5
-0,25	0,25	3,06	0,41
0	0,33	3	0,33
0,25	0,42	3,06	0,41
0,5	0,5	3,25	0,5
0,75	0,58	3,56	0,58
1	0,67	4	0,67

Cuadro A.3 Algunos valores del conjunto imagen del conjunto difuso S

En conclusión el conjunto imagen $B = f(A)$, podría definirse de la siguiente manera:

$$\mu_B(y) = \begin{cases} \frac{(y-3)^{1/2}}{3} + 1/3 & 3 \leq y \leq 7 \\ 3 - (y-3)^{1/2}, & 7 \leq y \leq 12 \\ 0. & \text{caso contrario} \end{cases}$$

A.4 DISTANCIA DIFUSA ENTRE CONJUNTOS DIFUSOS

Sea U un espacio métrico, con una pseudo-métrica $d: U \times U \rightarrow \mathfrak{R}^+$. La distancia difusa, d , entre conjuntos difusos A y B , puede definirse de la siguiente manera:

$$\mu_{d(A,B)}(y) = \sup_{\substack{(x_1, x_2) \in U \times U \\ d(x_1, x_2) = y}} \min \{ \mu_A(x_1), \mu_B(x_2) \}$$

Ejemplo

Se considera $U = \mathfrak{R}^1$ y la distancia $d(x_1, x_2) = |x_1 - x_2|$ y nuestros conjuntos difusos vienen dados por:

$$\mu_A(x_1) = \begin{cases} \frac{x_1 - 8}{2}, & \text{si, } 8 \leq x_1 \leq 10 \\ \frac{14 - x_1}{4} & \text{si, } 10 \leq x_1 \leq 14 \\ 0, & \text{caso contrario} \end{cases}$$

$$\mu_B(x_2) = \begin{cases} \frac{x_2 - 12}{2}, & \text{si, } 12 \leq x_2 \leq 15 \\ \frac{19 - x_2}{4} & \text{si, } 15 \leq x_2 \leq 19 \\ 0, & \text{caso contrario} \end{cases}$$

La distancia máxima sería $|19 - 8| = 11$, con valor cero para ambas funciones de pertenencia. Como no hay otra combinación posible de valores de A y B que de esta cifra, el supremo tomará, también, el valor de cero. Continuando los dos puntos extremos hacia arriba, tendríamos distancias decrecientes con funciones de pertenencia crecientes, hasta tomar el valor uno en $|15 - 10| = 5$,

Descendiendo, ahora, en ambos conjuntos, la distancia sigue decreciendo hasta la intersección de ambos, con $d=0$, y también lo hace la función de pertenencia, hasta llegar al punto $x_1 = x_2 = 12.86$ ($d=0$), en que toma valor 0.285. A partir de este punto tendríamos distancias ya calculadas, pero con nivel de pertenencia siempre menor que las anteriores.

La solución viene dada por el conjunto difuso cuya función de pertenencia es:

$$\mu_{d(A,B)}(y) = \begin{cases} \frac{0.285 + 0.715 \frac{y}{5}}{3} + 1/3 & 0 \leq y \leq 5 \\ \frac{11-y}{6}, & 5 \leq y \leq 11 \\ 0. & \text{caso contrario} \end{cases}$$

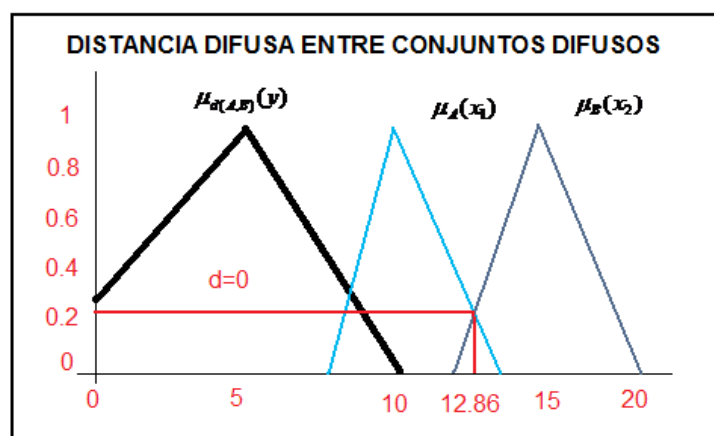


Gráfico.A.3 Distancia Difusa entre conjuntos difusos

A.6 CONVEXIDAD DE CONJUNTOS DIFUSOS

La convexidad juega un papel importante en la teoría de conjuntos difusos. Las condiciones de convexidad se definen en referencia a la función de pertenencia.

Definición:

Se dice que un conjunto difuso A , es convexo si:

$$\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \min\{\mu_A(x_1), \mu_A(x_2)\},$$

Para todo, $x_1, x_2 \in U$ y

Para todo $\lambda \in [0,1]$

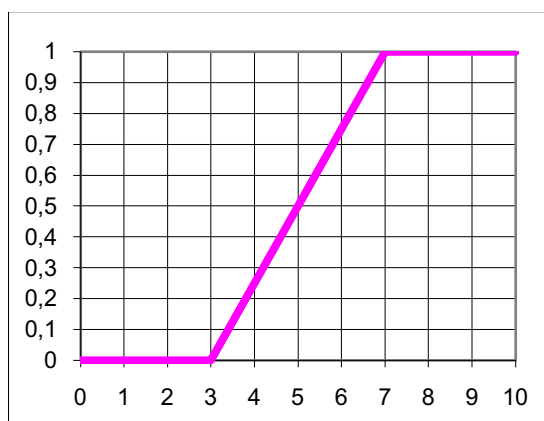


Gráfico A.4 Conjunto difuso convexo

A.7 CLASIFICACIÓN DE NUMEROS DIFUSOS

Un número difuso se define en \mathfrak{R} como un conjunto convexo normalizado. Para una mejor manipulación, suelen definirse los números difusos $L-R$ (left-right) como sigue:

$$\mu_A(x) = \begin{cases} L((M-x)/l) & \text{si } x \leq M, \quad l > 0 \\ R((x-M)/r) & \text{si } x \geq M, \quad r > 0 \end{cases}$$

Donde L y R son funciones crecientes y decrecientes en \mathfrak{R}^+ , con $L(0) = R(0) = 1$. M es llamado el valor central del número difuso. L y R son, respectivamente, las funciones de la forma a *izquierda* y a *derecha*, mientras que l y r son, respectivamente, la extensión, amplitud o dispersión a *izquierda* y a *derecha*.

Un intervalo difuso de tipo L-R es una generalización del concepto de número difuso que acabamos de ver:

$$\mu_A(x) = \begin{cases} L((M_1 - x)/l) & \text{si } x \leq M_1, & l > 0 \\ 1, & \text{si } x \in [M_1, M_2], \\ R((x - M_2)/r) & \text{si } x \geq M_2, & r > 0 \end{cases}$$

En el gráfico A.6 se observa una interpretación gráfica de los conceptos de

número Difuso de tipo L – R. Para:

$$L(x) = R(x) = \frac{1}{1+x^2} \text{ con } l=2, r=4 \text{ y } M=5,$$

Y de intervalo difuso, para:

$$L(x) = R(x) = \frac{1}{1+x^2} \text{ con } l=2, r=4 \text{ y } M_1=10 \text{ y } M_2=15.$$

Los respectivos conjuntos difusos quedan definidos de la siguiente

Número Difuso:

$$\mu(x) = \begin{cases} L((5-x)/2) = \frac{1}{1+((5-x)/2)^2} & \text{para } x \leq 5 \\ R((x-5)/4) = \frac{1}{1+((x-5)/4)^2} & \text{para } x \geq 5 \end{cases}$$

Intervalo Difuso:

$$\mu_a(x) = \begin{cases} 1, & \text{para } x = a \\ 0, & \text{para } x \neq a \end{cases}$$

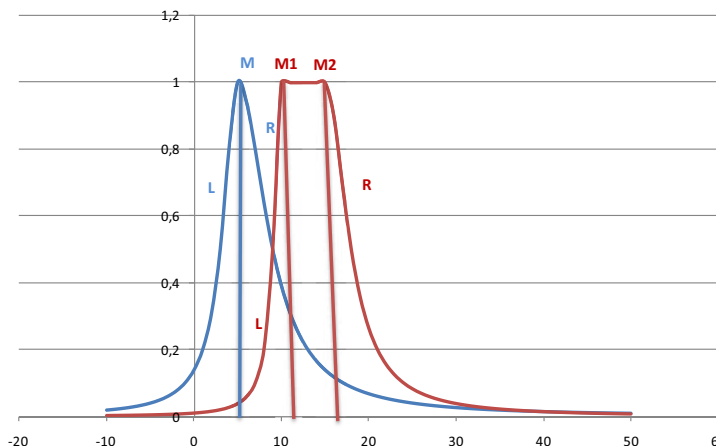


Gráfico.A.6 Interpretación gráfica de los conceptos de números difusos L – R.

A.8 OPERACIONES DEFINIDAS ENTRE CONJUNTOS DIFUSOS DEFINIDOS COMO INTERVALOS

Se presenta las operaciones más importantes entre los números difusos definidos con intervalos. Cuando se aplica una función a un conjunto de intervalos, el límite inferior (superior) del intervalo resultante será el valor mínimo (máximo) calculado aplicando esa función a todas las posibles combinaciones de valores pertenecientes a los intervalos considerados.

Suma: $[a_1, a_2](+)[b_1, b_2] = [a_1 + b_1, a_2 + b_2]$

Resta: $[a_1, a_2](-)[b_1, b_2] = [a_1 - b_1, a_2 - b_2]$

Multiplicación:

$$[a_1, a_2](\cdot)[b_1, b_2] = [\min(a_1.b_1, a_1.b_2, a_2.b_1, a_2.b_2); \max(a_1.b_1, a_1.b_2, a_2.b_1, a_2.b_2)]$$

Si el conjunto difuso está definido sobre \mathbb{R}^+ , sería $[a_1, a_2](\cdot)[b_1, b_2] = [a_1.b_1, a_2.b_2]$

Inverso:

$$[a_1, a_2]^{-1} = [\min(1/a_1, 1/a_2); \max(1/a_1, 1/a_2)] \text{ excepto para } a_1 \leq 0 \leq a_2$$

Si el conjunto difuso está definido sobre \mathbb{R}^+ , sería $[a_1, a_2]^{-1} = [1/a_2; 1/a_1]$

División:

$$[a_1, a_2](\cdot)[b_1, b_2] = [\min(a_1/b_1, a_1/b_2, a_2/b_1, a_2/b_2); \max(a_1/b_1, a_1/b_2, a_2/b_1, a_2/b_2)]$$

excepto para $a_1 \leq 0 \leq a_2$

Si el conjunto difuso está definido sobre \mathfrak{R}^+ , sería $[a_1, a_2](\cdot)[b_1, b_2] = [a_1/b_2, a_2/b_1]$

Mínimo: $\min([a_1, a_2]; [b_1, b_2]) = [\min(a_1, b_1); \min(a_2, b_2)]$

Máximo: $\max([a_1, a_2]; [b_1, b_2]) = [\max(a_1, b_1); \max(a_2, b_2)]$

Multiplicación por un número real

Un número real puede interpretarse como $a = [a, a]$, $a \in \mathfrak{R}$, por tanto,

$$a[b_1, b_2] = [\min(a.b_1, a.b_2); \max(a.b_1, a.b_2)]$$

$$[a_1^\alpha, a_2^\alpha](+)[b_1^\alpha, b_2^\alpha] = [a_1^\alpha + b_1^\alpha, a_2^\alpha + b_2^\alpha]$$

Todos estos resultados pueden aplicarse a números difusos, expresándolos en función de cada uno de sus α -cortes. Por ejemplo, en el caso de la suma:

$$\forall \alpha \in [0, 1], a_1^\alpha, a_2^\alpha, b_1^\alpha, b_2^\alpha \in \mathfrak{R}$$

A.9. OPERACIONES CON NÚMERO DIFUSOS TRIANGULARES

La suma, la resta, el complemento y el producto por un escalar de números difusos triangulares, da como resultado otro número difuso triangular; operaciones como la multiplicación, inverso, división, máximo y mínimo, entre otras, no arrojan necesariamente como resultado un número difuso triangular.

Definimos las siguientes operaciones:

Sean $A = (a_1, a_2, a_3)$ y $B = (b_1, b_2, b_3)$ dos números difusos triangulares.

- Suma: $A(+)B = (a_1, a_2, a_3)(+) (b_1, b_2, b_3) = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$

$$A = (-3, 2, 4) \text{ y } B = (-1, 0, 5)$$

$$A(+)B = (-4, 2, 9)$$

- Diferencia: $A(-)B = (a_1, a_2, a_3)(-)(b_1, b_2, b_3) = (a_1 - b_3, a_2 - b_2, a_3 - b_1)$

$$A = (-3, 2, 4) \text{ y } B = (-1, 0, 5)$$

$$A(-)B = (-8, 2, 5)$$

- Complemento: $A^- = (-a_1, -a_2, -a_3)$

$$A = (-3, 2, 4)$$

$$A^- = (-4, -2, 3)$$

- Producto de un escalar por un número difuso triangular

$$k.A = (\min(ka_1, ka_2), ka_M, \max(ka_1, ka_2)), \text{ para todo } k \in \mathfrak{R} / \{0\}$$

A.10 IMPLEMENTACIÓN EN CÓDIGO MATLAB DE LAS REGRESIONES DIFUSAS

```
disp('Ingrese la ubicación del archivo a ser analizado')
```

```
S=input("");
```

```
Data=importdata(S, '\t', 38);
```

```
clear S
```

```
Data.data(:,7)=Data.data(:,7)/1000; %km a m
```

```
Data.data(:,3)=Data.data(:,3)-273.15; %°K a °C
```

```
Y=Data.data(:,7);
```

```
l=size(Data.data,1);
```

```
f=[1; 1; 0; 0];
```

```
lb=[0; 0; -inf; -inf];
```

```
r=1;
```

```
while r==1
```

```
    K=menu('¿Qué análisis desea realizar? Altura  
vs','Presión','Temperatura','Humedad Relativa');
```

```
    H=input('Ingrese el grado de ajuste del modelo: ');
```

```
    invL=1-H; %funcion
```

```
%-----
```

```
if K==1 %Análisis de la Presión
```

```
    b=[Y;-Y];
```

```
    ejex=log(Data.data(:,8));
```

```
    X=[ones(l,1) ejex];
```



```

A=[-invL*abs(X), X;-invL*abs(X), -X];
clear invL
[x1,fval,exitflag,output,lambda] = linprog(f,A,b,[],[],lb);
%Cálculo del R^2
ey=polyval([x1(4) x1(3)],ejex);
Rsq=1-((norm(Y-ey))^2)/(l*var(Y,1));

z=min(ejex):1:max(ejex); %tomamos una malla en x para graficar la aprox

subplot(2,1,1);
plot(ejex,Y,'-r')
hold on
plot(z,x1(3)+(x1(4)-x1(2))*z) %límite inferior
plot(z,x1(3)+x1(4)*z,'k')
plot(z,x1(3)+(x1(4)+x1(2))*z) %límite superior

legend('Datos','Limite Inferior','RLD','LS')
title('Altura vs Ln(Presión)')
xlabel(['Ln(Presión)';[ 'y = (' ,num2str(x1(3),'%f'),',',num2str(x1(1),'%f'),')', ' + (' ,num2str(x1(4),'%f'),',',num2str(x1(2),'%f'),') *x' ]}]')
ylabel('Altura')
text(min(ejex)+1,min(Y)+1,['R^2 =',num2str(Rsq,'%f')])

%*****

subplot(2,1,2)
hold on
si=exp(x1(3))*(Data.data(:,8)).^(x1(4)-x1(2));
s=exp(x1(3))*(Data.data(:,8)).^(x1(4));
ss=exp(x1(3))*(Data.data(:,8)).^(x1(4)+x1(2));
plot(Data.data(:,8),log(si),'k')
plot(Data.data(:,8),Y,'r')
plot(Data.data(:,8),log(s),'b')
plot(Data.data(:,8),log(ss),'k')
legend('Límites','Datos','R.L.D')
title('Altura vs. Presión')
%*****

end
%-----
if K==2
disp('Se realizará una regresión lineal difusa en dos partes')
[m,c]=min(Data.data(:,3));
X1=[ones(c,1) Data.data(1:c,3)];
X2=[ones(l-c,1) Data.data(c+1:l,3)];
A1=[-invL*abs(X1), X1;-invL*abs(X1), -X1];
A2=[-invL*abs(X2), X2;-invL*abs(X2), -X2];
clear invL
b1=[Y(1:c);-Y(1:c)];

```

```

b2=[Y(c+1:l);-Y(c+1:l)];
[x1,fval1,exitflag1,output1,lambda1] = linprog(f,A1,b1,[],[],lb);
[x2,fval2,exitflag2,output2,lambda2] = linprog(f,A2,b2,[],[],lb);

ey1=polyval([x1(4) x1(3)],X1(:,2));
Rsq1=1-((norm(Y(1:c)-ey1))^2)/(l*var(Y(1:c)));
ey2=polyval([x2(4) x2(3)],X2(:,2));
Rsq2=1-((norm(Y(c+1:l)-ey2))^2)/(l*var(Y(c+1:l)));

z=min(Data.data(:,3)):(max(Data.data(:,3))-
min(Data.data(:,3)))/l:max(Data.data(:,3));
plot(Data.data(:,3),Y,'-r')
hold on

plot(z,x1(3)+(x1(4)-x1(2))*z,'-b') %limite inferior
plot(z,x1(3)+x1(4)*z,'k')
plot(z,x1(3)+(x1(4)+x1(2))*z,'-b') %limite superior

plot(z(1:l-c),x2(3)+(x2(4)-x2(2))*z(1:l-c),'-b') %limite inferior
plot(z(1:l-c),x2(3)+x2(4)*z(1:l-c),'k')
plot(z(1:l-c),x2(3)+(x2(4)+x2(2))*z(1:l-c),'-b') %limite superior

legend('Datos','Limites','RLD')
title('Altura vs Temperatura')
%xlabel({'Temperatura';['y_1 = (' ,num2str(x1(3),'%f'),' , ',num2str(x1(1),'%f'),' )' +
(' ,num2str(x1(4),'%f'),' , ',num2str(x1(2),'%f'),' ) *x'];['y_2 = (' ,num2str(x2(3),'%f'),' ,
',num2str(x2(1),'%f'),' )' + (' ,num2str(x2(4),'%f'),' , ',num2str(x2(2),'%f'),' ) *x']})
%ylabel('Altura')
%text(Data.data(c,3),Y(l),{'R_1^2 =',num2str(Rsq1,'%f')};[],['R_2^2
=',num2str(Rsq2,'%f')'],'HorizontalAlignment','right')
end
%-----
if K==3
% Calculate fit parameters
[p,ErrorEst] = polyfit(Data.data(:,4),Y,2);
% Evaluate the fit and the prediction error estimate (delta)
[pop_fit,delta] = polyval(p,Data.data(:,4),ErrorEst);

Rsq=1-((norm(Y-pop_fit))^2)/(l*var(Y));

% Plot the data, the fit, and the confidence bounds
plot(Data.data(:,4),Y,'-',Data.data(:,4),pop_fit,'g-
',Data.data(:,4),pop_fit+2*delta,'r:',Data.data(:,4),pop_fit-2*delta,'r:');

legend('Datos','Ajuste de los datos','Limites')
title('Altura vs Humedad relativa')
ylabel('Altura')

```

```

    xlabel({'Humedad Relativa';['y = (' ,num2str(p(1),'%f'),', ' ,num2str(delta(1),'%f'),')
*x^2,' + (' ,num2str(p(2),'%f'),', ' ,num2str(delta(1),'%f'),') *x,' + (' ,num2str(p(3),'%f'),',
',num2str(delta(1),'%f'),')']])
    text(mean(Data.data(:,4)),max(Y)-1,['R^2 =',num2str(Rsq,'%f')])
    clear delta
end
r=menu('¿Desea realizar otro análisis con estos datos?','Si','No');
end

disp('End')

```

A.11 IMPLEMENTACIÓN EN CÓDIGO MATLAB MÉTODO DE LAS C-MEDIAS DIFUSO MEJORADO

```

numMFs = [4 4 4];
mfType = str2mat('gaussfm','gaussfm','gaussfm');
fismat = genfis1(datos,numMFs,mfType);
[x,mf] = plotmf(fismat,'input',1);
subplot(3,1,1), plot(x,mf);
xlabel('temperatura (gaussfm)');
[x,mf] = plotmf(fismat,'input',2);
subplot(3,1,2), plot(x,mf);
xlabel('Humedad Relativa (gaussfm)');
[x,mf] = plotmf(fismat,'input',3);
subplot(3,1,3), plot(x,mf);
xlabel('Presión (gaussfm)');

```

genfis1

Generate Fuzzy Inference System structure from data using grid partition

plotmf

Plot all membership functions for given variable

Syntax

```
plotmf(fismat,varType,varIndex)
```

Description

This function plots all of the membership functions in the FIS called fismat associated with a given variable whose type and index are respectively given

by varType (must be 'input' or 'output'), and varIndex. This function can also be used with the MATLAB function, subplot.

genfis1

Generate Fuzzy Inference System structure from data using grid partition

Syntax

```
fismat = genfis1(data)
```

```
fismat = genfis1(data,numMFs,inmftype,outmftype)
```

Description

genfis1 generates a Sugeno-type FIS structure used as initial conditions (initialization of the membership function parameters) for anfis training.