

Instituto Colombiano de Hidrología, Meteorología y  
Estudios Ambientales IDEAM-Subdirección de  
Meteorología

25 de octubre de 2011



Investigación del comportamiento de la precipitación bajo el fenómeno ENOS, mediante el análisis por terciles para su aplicación en modelos de distribución hipergeométrica para fines de predicción climática. (Contrato No. 070 de 2011)

# Jazikü

*El software para la aplicación de la Metodología de análisis compuesto entre los principales índices de la variabilidad climática y las principales variables meteorológicas en la escala puntual.*

*(Manual de usuario)*

Elaborado por: Inés Sánchez R.

-----

Revisado por : Gloria León Ariztizabal

Xavier Corredor

VoBo.-----

Bogotá D.C., Septiembre de 2011

# Contents

<b>Contents</b>	<b>4</b>
<b>1 Introducción</b>	<b>6</b>
<b>I Manual operativo</b>	<b>7</b>
<b>2 Como usar Jazikü?</b>	<b>8</b>
2.1. Entrada de datos. . . . .	10
<b>3 Acerca de Jazikü y sus resultados</b>	<b>14</b>
<b>II Manual teórico</b>	<b>20</b>
<b>4 Acerca de la información inicial</b>	<b>21</b>
4.1. Variable Dependiente . . . . .	21
4.2. Variable Independiente . . . . .	22
<b>5 Jazikü - Clima</b>	<b>31</b>
5.1. Información calculada por Jazikü . . . . .	31
5.2. Análisis compuesto y Tablas de contingencia . . . . .	31
5.3. Construcción e interpretación de la Tabla de contingencia . . . . .	32
5.3.1. Correlación lineal . . . . .	33
5.3.2. Prueba de bondad de ajuste $\chi^2$ . . . . .	34
5.3.3. Coeficiente V de Cramer . . . . .	35
5.3.4. Analisis de riesgos . . . . .	36
5.3.4.1. Distribución hipergeometrica . . . . .	36
5.3.4.2. Metodología del análisis de riesgos . . . . .	38
<b>6 Jazikü - Pronóstico</b>	<b>41</b>
6.1. Pronóstico operativo mensual de predicción climática de los posibles escenarios de impacto, en la variable dependiente debido al fenómeno ENOS . . . . .	41
<b>7 Reconocimientos y licencias de uso</b>	<b>43</b>

<i>CONTENTS</i>	5
<b>III Apéndices</b>	<b>44</b>
<b>8 Apéndice 1: Posibles errores arrojados por Jazikü</b>	<b>45</b>
8.1. Error por mala definición del período al correr Jazikü . . . . .	45
8.2. Errores en el modulo de entrada de datos. . . . .	45
8.2.1. Variable precipitación . . . . .	45
8.2.2. Variable numero de días con lluvia . . . . .	46
8.2.3. Variable temperatura media, máxima y mínima . . . . .	46
8.2.4. Variable presión . . . . .	46
8.2.5. Variable humedad relativa . . . . .	47
8.3. Errores en la variable independiente . . . . .	47
8.4. Error por inconsistencia de la tabla de contingencia. . . . .	47
8.5. Error en generación de pronóstico . . . . .	48
8.6. Acerca del contenido de las carpetas “maps” generadas por el software . . . . .	48
8.7. Ejemplo de aplicación del software . . . . .	49
<b>Bibliography</b>	<b>56</b>

# Capítulo 1

## Introducción

De acuerdo con el compromiso IDEAM por promover y orientar la investigación científica sobre el pronóstico del tiempo y el clima "Jazikü" es un programa diseñado para evaluar las teleconexiones entre las variables meteorológicas con los principales índices de la variabilidad climática que afectan el clima en Colombia.

Jazikü, sigue la metodología de análisis de compuesto propuesto por la Universidad - Corporación para la Investigación Atmosférica (UCAR)), Administración Nacional Oceánica y Atmosférica (NOAA) y EE.UU. Departamento de Comercio (DOC) [1][1, 2][1, 2, 21][1, 2, 21] y puede producir escenarios de probabilidad bajo los cuales se espera que la precipitación o cualquier otra variable en los sitios de mayor afectación o áreas específicas interpoladas se comporten, en función de los valores de probabilidad estimados para cada uno de los fenómenos de variabilidad del clima y de la historia historia climatológica en relación al valor promedio trimestral. Tales escenarios se han convertido en una herramienta poderosa para la toma de decisiones por los servicios meteorológicos nacionales.

Parte I

Manual operativo

## Capítulo 2

# Como usar Jazikü?

Una vez Jazikü ha sido instalado correctamente en el sistema operativo, existen varias opciones de corrida, que se corresponden directamente con los cálculos que pueden ser elaborados por el programa (ver sección 3, al final ver tabla de referencia).

A continuación se presenta el comando estándar y básico de corrida del programa:

```
jaziku -station stations.csv --climate
```

Phyton es el compilador y lenguaje bajo el cual el programa ha sido elaborado<sup>1</sup>; "jaziku" es el código fuente del programa. El comando "-station" indica que seguido a este y separado por un espacio el programa encontrara el archivo .csv con información básica de corrida, en dirección absoluta (ver sección 2.1); "-climate" le indica al programa que solo se correrá en modo clima, opcionalmente puede usarse -c.

Para correr el programa en modo pronóstico (ver sección 3), se utiliza adicionalmente el comando "-forecasting" u opcionalmente -f. Una opción avanzada incluye el análisis de riesgos para cada entrada de la tabla de contingencia, utilizando el comando "-ra". Este análisis en efecto proporciona mayor confiabilidad, pero puede reducir el tamaño de la muestra utilizada para el pronóstico (ver sección 5.3.4, 8.5), así

```
jaziku -station stations.csv --climate --forecasting
```

```
jaziku -station stations.csv --climate --forecasting -ra
```

El usuario puede seleccionar el período para el cual desea sea aplicado la metodología de análisis compuesto (ver sección 5.2), habilitando la opción "-period" separado por un espacio en formato yyyy-yyyy (año-año), así

```
jaziku -station stations.csv --climate --forecasting -period  
1990-2000
```

---

<sup>1</sup>Mayor información en [www.python.org/](http://www.python.org/)



El usuario puede hacer que Jazikü etiquete las categorías por encima, por debajo y dentro de lo normal de la variable independiente, para que en los resultados sea impreso bajo estos nombres (ver sección 3). Para ello incluye en el comando de inicio las etiquetas “-p\_below”, “-p\_normal” y “-p\_above”, donde “-p” hace referencia al fenómeno (phenomenon, ver sección 2.1) que representa la variable independiente, así

```
jaziku -station stations.csv --climate --forecasting -period
1990-2000 -ra -p_below "LaNiña" -p_normal "Neutro" -p_above
"ElNiño"
```

Del mismo modo adicionando el comando “-maps” se logra que Jazikü interpole las salidas de la climatología y el pronóstico asociado. Finalmente se espera una vez Jazikü ha sido corrido de manera correcta, el usuario visualice<sup>2</sup>

```
##### JAZIKU #####
## Jaziku is a software for the implementation of composite ##
## analysis methodology between the major indices of climate ##
## variability and major meteorological variables in ##
## puntual scale. ##
## ##
## Version 1.0 ##
## @IDEAM Colombia, 2011 ##
#####
Setting period in: 1990-2000
##### Station: 4801501-01 (4801501)
Processing climate (1990-2000).....done
Processing forecasting (1990-2000).....done
```

---

<sup>2</sup>Un ejemplo para el cual 4801501 es el código de la estación, corrido en clima y pronóstico, en el periodo 1990-2000

Cuadro 2.1: Tabla de referencia para correr Jazikü

*Usage: python Jazikü.py [OPTION]... [-T] SOURCE DEST*  
*RUN Jazikü a software for the implementation of composite*  
*analysis methodology between the major indices of climate*  
*variability and major meteorological variables in puntual scale.*

*Los argumentos obligatorios para las opciones largas son también obligatorios para las opciones cortas. -station : seguido de la dirección absoluta del archivo inicial de corrida del software, información de ubicación de archivos*

*-c, --climate :Corre la climatología de las estaciones asociadas bajo el fenómeno de variabilidad climática seleccionado*

*-f , --forecasting :Corre el pronóstico asociado*

*-ra :Para utilizar análisis de riesgos*

*-period :Para indicar el período bajo el cual se desea correr la climatología*

*-p\_below :Label del fenómeno asociado a la var\_Independiente en la categoría por encima de lo normal*

*-p\_normal :Label del fenómeno asociado a la var\_Independiente en la categoría dentro de lo normal*

*-p\_above :Label del fenómeno asociado a la var\_Independiente en la categoría por debajo de lo normal*

*Jazikü por defecto crea dos carpetas asociadas a los procedimientos -c y -f*

## 2.1. Entrada de datos.

Jazikü a sido diseñado para que el usuario incluya en dos archivos los datos correspondientes a las series de tiempo de los valores promedio mensual de la variable independiente y dependiente. Dichos archivos corresponden a extensiones .txt (plain text document), la Fig.2.1 presenta dos ejemplos de series vistas desde el lector Kwrite de Opensuse, no olvide que el valor estándar para datos perdidos o no medidos es 99999.

Para correr Jazikü existen convenciones asociadas a las series de las variables dependientes e independientes, a continuación son listados:

- PPT(Precipitación)
- NDPPT (número de días con lluvia)
- TMIN (temperatura mínima)

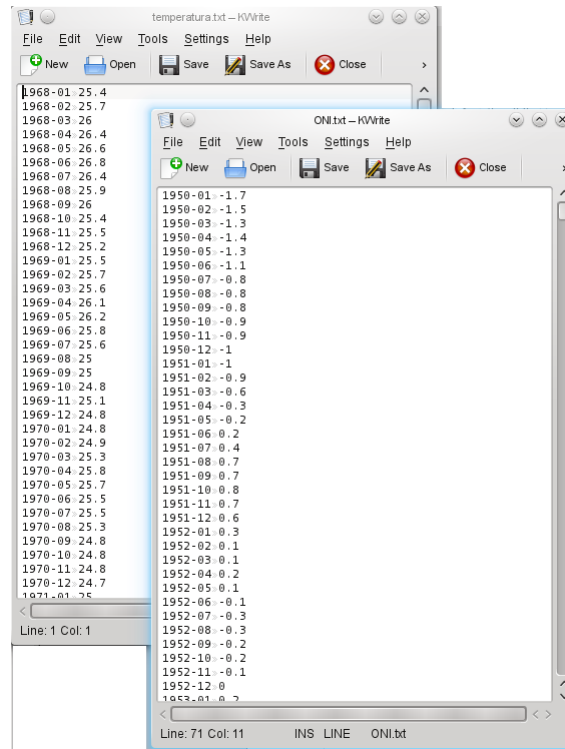


Figura 2.1: Ejemplo de formato de datos de entrada

- TMAX (temperatura máxima)
- TEMP (temperatura media)
- PATM (presión atmosférica)
- AHR (humedad relativa del dipolo Amazónico)
- SST (Temperatura superficial del Mar)
- ONI ( Índice Oceánico El Niño)
- SOI ( Índice de la Oscilación del Sur)
- MEI ( Índice Multivariado del ENSO)
- OLR ( Índice de radiación de longitud de onda larga)
- W200 ( Índice de la anomalía del viento a 200 hPa)
- ARH (Índice de humedad del Amazonas)
- NAO (Índice del Atlántico Norte)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	P	Q	R	S	T	V
Codigo	Nombre St	Lat	Long	Nombre archivo	Variable	Nombre archivo	Variable	Threshold d_below	Threshold d_above	F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag	TRIM
3506006	Paraguasito	24.45	74.3	3506006_ppt.txt	PPT	TSM34.txt	SST	26.287	27.095	23	32	65	23	92	41	23	92,4	41	8
3506009	Pte.Corto	22.5	74.2	3506009_ppt.txt	PPT	TSM34.txt	SST	26.906	27.441	23	32	65	23	92	41	23	92	41,2	5
3506012	Ha.Verdura	20.55	74.1	3506012_ppt.txt	PPT	TSM34.txt	SST	27.33	27.826	23	32	65	23	92	41	23	92,5	42	8
3506015	Pte.Medio	18.6	74	3506015_ppt.txt	PPT	TSM34.txt	SST	27.565	27.948	23	32	65	23	92	41	23	92	41,3	5
3506018	El Muelle	16.65	73.9	3506018_ppt.txt	PPT	TSM34.txt	SST			23	32	65	23	92	41	23	92,6	43	9

**Callout Box A:**

A	B
Codigo	Nombre St
3506006	Paraguasito
3506009	Pte.Corto
3506012	Ha.Verdura
3506015	Pte.Medio
3506018	El Muelle

**Callout Box E:**

E	F
Nombre archivo	Tipo
Var dep	Dep
3506006_ppt.txt	PPT
3506009_ppt.txt	PPT
3506012_ppt.txt	PPT
3506015_ppt.txt	PPT
3506018_ppt.txt	PPT

**Callout Box G:**

G	H
Nombre archivo	Tipo
Variable	Indp
TSM34.txt	SST
TSM34.txt	SST
TSM34.txt	SST
TSM34.txt	SST
TSM34.txt	SST

**Callout Box N:**

N	P	Q
F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag
1 Below	1 Normal	1 Above
23	92	41
23	92	41
23	92	41
23	92	41
23	92	41

**Callout Box C:**

C	D
Lat	Long
24.45	74.3
22.5	74.2
20.55	74.1
18.6	74
16.65	73.9

**Callout Box I:**

I	J
Threshold d_below	Threshold d_above
_var_I	_var_I
26.287	27.095
26.906	27.441
27.33	27.826
27.565	27.948

**Callout Box K:**

K	L	M
F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag
0 Below	0 Normal	0 Above
23	32	65
23	32	65
23	32	65
23	32	65
23	32	65

**Callout Box R:**

R	S	T
F_Var Indp Lag	F_Var Indp Lag	F_Var Indp Lag
2 Below	2 Normal	2 Above
23	92,4	41
23	92	41,2
23	92,5	42
23	92	41,3
23	92,6	43

**Callout Box V:**

V
TRIM
8
5
8
5
9

Figura 2.2: Ejemplo de formato de la lista inicial

### ■ QBO(Índice Oscilación Cuasi-bienal)

Dado que Jazikü evalúa las consistencia de las series de tiempo al inicio de su ejecución (ver sección4), es posible que se encuentren errores que no permitan su inicio, para ello consultar la sección 8de este documento.

Como Jazikü es capaz de analizar mas de una variable dependiente en función de mas de una variable independiente, esto es, múltiples archivos de estación en función de múltiples archivos de índices de variabilidad climática, requiere de una lista de entrada que sea capaz de especificarle el tipo de proceso que debe ejecutar.La Fig.2.2 presenta un ejemplo de esta lista de inicio, que debe darse en formato .csv .

### 2.1

En la Fig.2.2 aparecen 19 columnas , cuyos campos son especificados así:

A y B: Lista de códigos y nombres de las estaciones de las cuales fueron extraídas las series de tiempo de la variable dependiente

C y D: Coordenadas latitud y longitud de las estaciones de las cuales fueron extraídas las series de tiempo de la variable dependiente

E :Ubicación absoluta del archivo asociado a cada una de las estaciones de las cuales fueron extraídas las series de tiempo de la variable dependiente. Eje: /home/Usuario/Documentos/Climatologia/2011/series/precipitacion/3506006\_ppt.txt

F:Convención de la variable dependiente

G:Ubicación absoluta del archivo asociado a la serie de tiempo de la variable independiente. Eje: /home/Usuario/Documentos/Climatologia/2011/series/indice\_tsm/TSM34.txt

H:Convención de la variable independiente

V:Trimestre de pronostico (un numero de 1 a 12, donde 1 corresponde al trimestre Enero-Febrero-Marzo)

I y J:Umbrales de la variable independiente, que dan lugar a las categorías por encima, por debajo y dentro de lo normal (donde el valor de la columna I debe ser siempre menor al valor de la columna J), de acuerdo al siguiente formato:

*default;default* Hace que el programa utilice las reglas de la Tabla 4.2(ultimas dos columnas)

*P##;P##* Hace que el programa utilice los percentiles seleccionados, por ejemplo al colocar “P33;P66”, es utilizado el percentil 33 para categorizar los valores por debajo de lo normal y el percentil 66 para los valores por encima de lo normal.

*sdn;sdn* Hace que el programa utilice n veces (de uno hasta 3) la desviación estándar de la serie

*##;##* Valores numéricos, enteros o decimales (separados por ,)ingresados por el usuario bajo sus propios criterios.

K,L y M : Valores de probabilidad de que la variable independiente a CERO rezagos del trimestre de pronóstico, este por debajo de lo normal, dentro de lo normal y por encima de lo normal.

N,P y Q : Valores de probabilidad de que la variable independiente a UN rezagos del trimestre de pronóstico, este por debajo de lo normal, dentro de lo normal y por encima de lo normal.

R,S y T : Valores de probabilidad de que la variable independiente a DOS rezagos del trimestre de pronóstico, este por debajo de lo normal, dentro de lo normal y por encima de lo normal.

## Capítulo 3

# Acerca de Jazikü y sus resultados

El software Jazikü ha sido diseñado para encontrar las relaciones de la metodología de análisis compuesto entre los valores de las series de tiempo de las variables dependientes con las series de tiempo de las variables independientes. Jazikü tiene 3 divisiones fundamentales en su proceso de datos. Inicia con un archivo que le da a conocer el listado de las estaciones en análisis, con su respectiva ubicación geográfica, así como información correspondiente a la variable dependiente e independiente, adicional solicita al operador los valores de probabilidad de comportamiento de la variable independiente (Fig.3.1,también ver sección2.1).

Se espera que en dos fases siguientes Jazikü genere dos carpetas, una asociada al esquema “Jazikü - clima” y otra asociada al esquema “Jazikü - pronóstico”, que son guardadas en el mismo directorio de donde se corre el programa y en donde se espera se encuentre ubicado el archivo de lista de inicio (ver Fig.3.2).

En la segunda fase (esquema Jazikü - clima), Jazikü elabora la climatología de las series de la variable dependiente bajo las fases asociadas al fenómeno representado por la variable independiente. En esta etapa Jazikü inicialmente genera las series de valores promedio trimestrales para cada uno de los trimestres del año (12 en total<sup>1</sup>) tanto de la variable independiente como la dependiente. Estas series son la materia prima para el desarrollo de la técnica de análisis compuesto, cuyo principal resultado es la tabla de contingencia; esta tabla de contingencia es evaluada mediante el procedimiento de análisis de riesgos (ver Fig.3.1) .

---

<sup>1</sup>Trimestre 1(EFM): Enero-Febrero-marzo ;Trimestre 2 (FMA): Febrero-marzo-abril ; Trimestre 3 (MAM): Marzo-abril-mayo ; Trimestre 4 (AMJ): Abril-mayo-junio Trimestre 5 (MJJ): Mayo-junio-julio;Trimestre 6 (JJA): Junio-julio-agosto;Trimestre 7 (JAS): Julio-agosto-septiembre;Trimestre 8 (ASO): Agosto-septiembre-octubre;Trimestre 9 (SON): Septiembre-octubre-noviembre;Trimestre 10 (OND): Octubre-noviembre-diciembre;Trimestre 11 (NDE): Noviembre-diciembre-enero;Trimestre 12 (DEF): Diciembre-enero-febrero

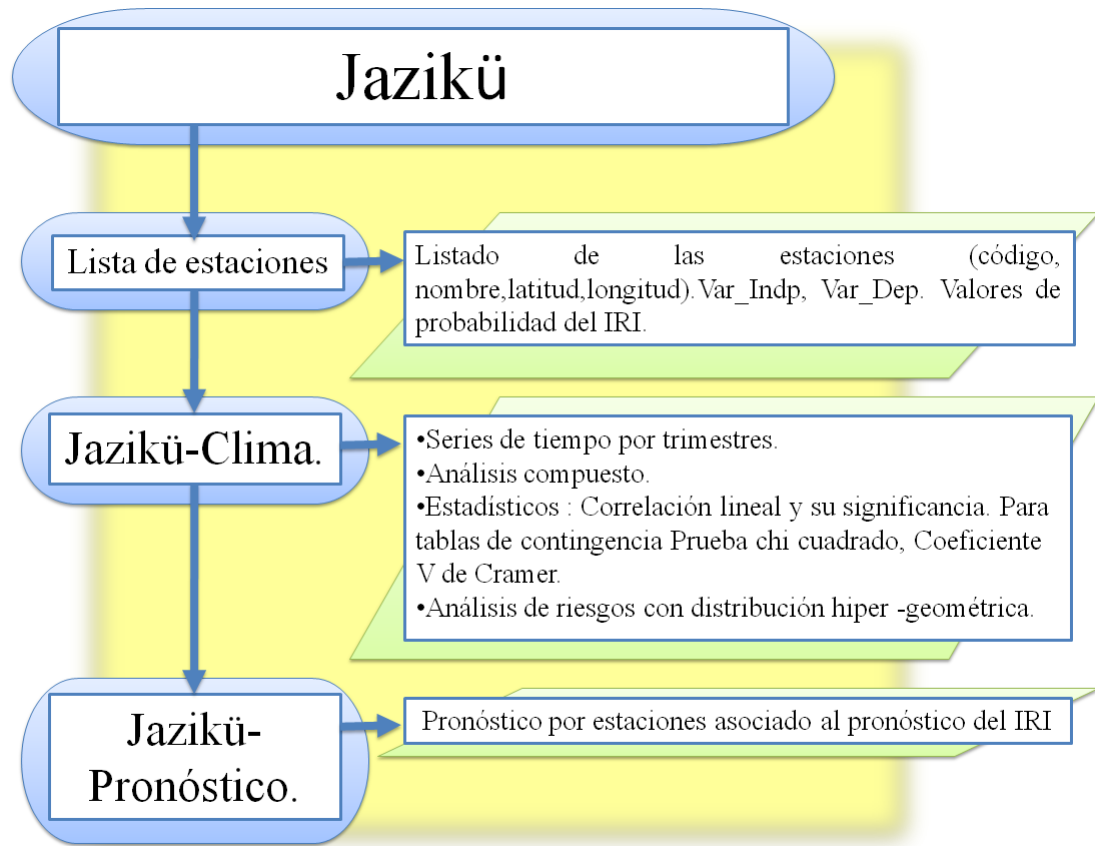


Figura 3.1: Esquema general del software Jazikü

Otros estadísticos de interés son calculados por Jazikü en esta etapa, como la correlación lineal de las series por trimestres, la correlación y el test chi cuadrado de la tabla de contingencia. Finalmente y por medio de la distribución hipergeométrica, un análisis de riesgos es elaborado.

Posterior al desarrollo de la climatología, Jazikü utiliza los valores de probabilidad del fenómeno representado por la variable independiente, y por medio de una combinación lineal de los valores de la tabla de contingencia, genera un pronóstico de afectación de la variable dependiente, en función de los valores pronosticados por la variable independiente(ver Fig.3.1).

Los resultados arrojados por el software, son guardados en la carpeta `data_out` (ver Fig.3.3), esta carpeta guarda a su vez una carpeta llamada “maps”



Figura 3.2: Esquema de las carpetas de resultados de Jazikü y ubicación del archivo station\_list (lista de estaciones ver sección 2.1)

y otras carpetas cuyo nombre coincide con el de las estaciones incluidas en la lista de estaciones. La carpeta “maps”, por rezagos guarda los archivos .csv, que deben ser utilizados para la generación de mapas interpolados en función de las categorías: por encima, por debajo y dentro de lo normal de la variable independiente (ver apéndice 8.6). Cada una de las carpetas asociadas a cada estación, guarda a su vez las series de tiempo construidas por trimestres para los valores promedio de la variable dependiente e independiente, las tablas de contingencia por rezagos son guardadas en la carpeta “contingency\_table”, las gráficas asociadas a estas tablas de contingencia en la carpeta “graphics”; estas gráficas son diagramas de barras en mosaico por rezagos o de manera independiente por trimestre(ver Fig.3.3).

Sin lugar a dudas el resultado mas importante que arroja el software es la tabla de resultados. En esta tabla se presentan por rezagos los valores de los estadísticos calculados por el software, así como la tabla de contingencia. Esta tabla se considera de gran utilidad porque en una sola fila es posible encontrar la respuesta climática de la estación a los cambios de la variable independiente, bajo la técnica de análisis compuesto. La salida de esta tabla se da en formato .csv, que es de fácil lectura bajo cualquier editor de texto

La Fig. 3.4 presenta un esquema de la tabla. Nótese que las columnas han sido agrupadas por letras así:

- A y B : columnas de los trimestres asociados a las variables dependiente e independiente
- C,D y E: correlación de Pearson, significancia de la correlación mediante la prueba t y selección de los valores significativos con un 95 % de confianza (“yes” para correlaciones significativas, “no” para las no significativas)



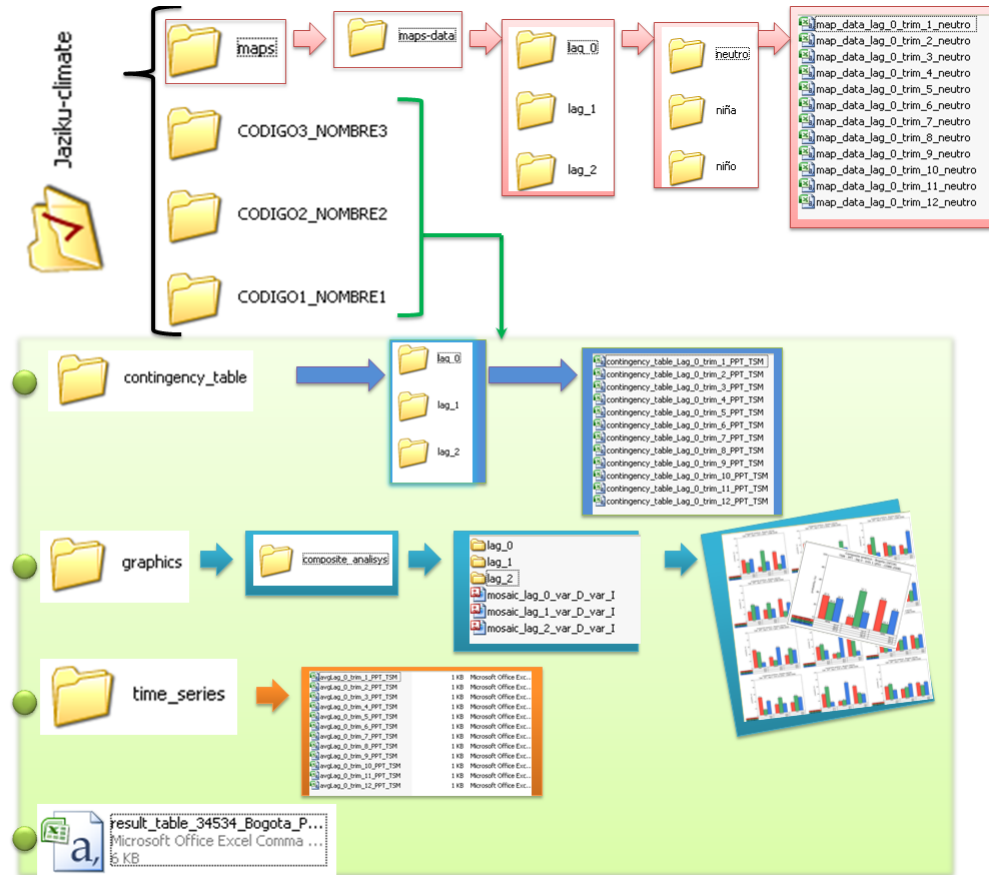


Figura 3.3: Esquema de los resultados de Jazikü - clima

- F y G: Umbrales usados para calcular las tablas de contingencia de acuerdo a las categorías por encima, por debajo y dentro de lo normal, de la variable dependiente.
- H y I: Umbrales usados para calcular las tablas de contingencia de acuerdo a las categorías por encima, por debajo y dentro de lo normal, de la variable independiente.
- J: Valores de la tabla de contingencia en frecuencias empíricas, para cuando se analiza la categoría por debajo de lo normal de la variable independiente. El primer valor corresponde a la categoría por debajo de lo normal de la variable dependiente, el segundo a la categoría dentro de lo normal de la variable dependiente y el tercero para la categoría por encima de lo normal.

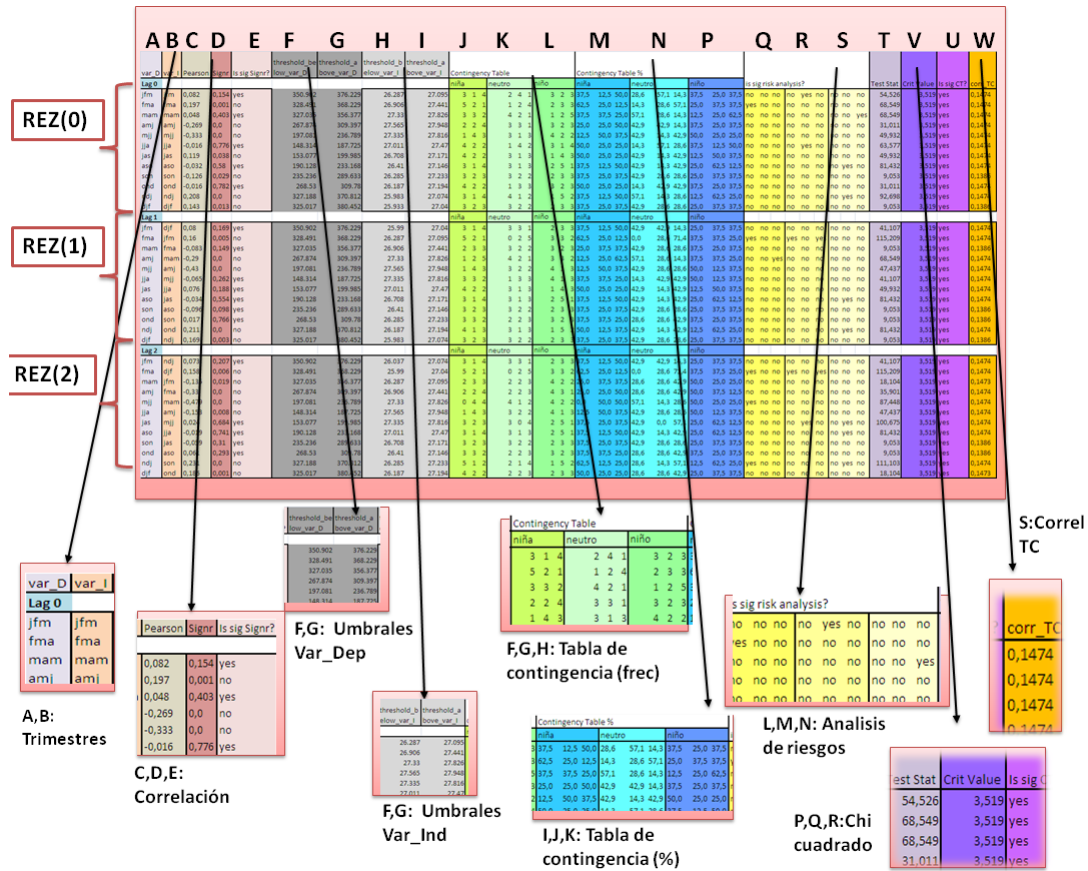


Figura 3.4: Esquema de la tabla general de resultados de Jazikû - clima

- K: Valores de la tabla de contingencia en frecuencias empíricas, para cuando se analiza la categoría por dentro de lo normal de la variable independiente. El primer valor corresponde a la categoría por debajo de lo normal de la variable dependiente, el segundo a la categoría dentro de lo normal de la variable dependiente y el tercero para la categoría por encima de lo normal.
- L: Valores de la tabla de contingencia en frecuencias empíricas, para cuando se analiza la categoría por encima de lo normal de la variable independiente. El primer valor corresponde a la categoría por debajo de lo normal de la variable dependiente, el segundo a la categoría dentro de lo normal de la variable dependiente y el tercero para la categoría por encima de lo normal.
- M,N y P: los mismos valores de las columnas agrupadas en J,K, y L, pero

en valores de porcentaje.

- Q,R y S: Análisis de riesgos para cada uno de los 9 valores de la tabla de contingencia.
- P,Q y R: Valor del estadístico Chi cuadrado, valor critico contra el cual es evaluado a un 95 % de confianza y selección de valores significativos en función del contraste de los dos valores anteriores (“yes” significativo, “no” no significativas)
- T: Correlación de la tabla de contingencia, calculada con el estadístico V de Cramer.



Figura 3.5: Esquema de la tabla general de resultados de Jazikü - pronóstico

Los resultados asociados al esquema Jazikü - pronóstico, son guardados en la carpeta “Jazikü-forecast”, esta carpeta a su vez guarda una carpeta denominada “maps”, en esta carpeta y por rezagos son guardados los valores de pronóstico de todas las estaciones incluidas en el análisis en función de su ubicación geográfica (ver Fig.3.5) para comprender los archivos allí presentados ver apéndice 8.6. Las otras carpetas, una por cada estación corresponden a los valores de pronóstico presentados en esquemas redondos (pie) de manera individual para cada rezago y de manera conjunta para la estación.

Parte II

Manual teórico

## Capítulo 4

# Acerca de la información inicial

Con el fin de automatizar el proceso de previsión en la metodología de análisis de compuesto, es necesario seleccionar las variables dependientes e independientes. Son variables dependientes, todas las variables meteorológicas que deben estar relacionados con los principales índices de la variabilidad climática, a continuación se describen.

### 4.1. Variable Dependiente

Se trata de utilizar las variables de mayor interés en el territorio nacional, la Tabla 4.1 muestra los valores para los cuales se estima es el rango de validez y las unidades típicas en las que deben ser incluidas las series de tiempo en Jazikü. En esta etapa el programa efectúa varias verificaciones a fin de tener la seguridad de que los valores son razonables de acuerdo con el comportamiento de las variables meteorológicas, de lo contrario se rechazan los datos [15].

Este tipo de verificación para las variables presión, humedad relativa y temperatura se corresponde con la sugerida por León (2007, [15]) en el documento: "Control de Calidad Automático de datos Meteorológicos", la cual restringe las observaciones para que estén dentro de un conjunto de límites de tolerancia (límites blandos), que fueron seleccionados para el área comprendida entre los 30°N y -20°N, de acuerdo con las especificaciones físicas de la variable que determinan en algunos casos límites fijos, por ejemplo la humedad relativa es una variable acotada entre 0 y 100 %, o que bien se pueden referir a límites climatológicos determinados estadísticamente, que pueden variar geográfica y/o temporalmente (tomado de [15]). En el caso de las variables número de días con lluvia y de la variable precipitación se analizaron las condiciones mas usuales de la climatología para todo el país, de acuerdo al Atlas Climatológico Nacional del IDEAM[19].

Cuadro 4.1: Rango de validez y unidades típicas en las cuales la variable dependiente debe ser incluida en Jazikü

Variable	Abreviación	Unidades	Rango de variación
Precipitación	Ppt	mm	$0 \text{ mm} \leq \text{Ppt} \leq 3500 \text{ mm}$
Numero de días con lluvia	NdPpt	#	Varia entre 0 a 31 para enero, a 28 o 29 para Febrero (dependiendo de si el año es bisiesto o no), a 31 para Marzo, a 30 para Abril, a 31 para Mayo, a 30 para Junio, a 31 para Julio, a 31 para Agosto, a 30 para Septiembre, a 31 para Octubre, a 30 para Noviembre, a 31 para Diciembre
Temperatura del aire (adherida, Máxima y Mínima Diaria))	T, Tmin, Tmax	°C	$-15^{\circ}\text{C} \leq \text{Tmin} \leq 50^{\circ}\text{C}$
Presión	P	mb	$400 \text{ mb} \leq P \leq 1100 \text{ mb}$
% HR	HR	-	$0 \% \leq \text{HR} \leq 100 \%$

## 4.2. Variable Independiente

El clima usualmente es entendido como el conjunto de condiciones medias del tiempo atmosférico, agrupadas por épocas del año, no obstante no es usual que los valores promedios mensuales de las variables meteorológicas que lo definen sean siempre estáticos, distintos fenómenos pueden influir en su variabilidad, alteraciones reflejadas a través de las anomalías climatológicas. Importantes estudios ya han sido realizados con el objetivo de identificar dichas afectaciones, no obstante un esquema muy completo de las teleconexiones mas importantes para la precipitación del país es el presentado por Montealegre (2009,[20]). Quien seleccionó del océano y de la atmósfera los componentes del sistema climático que mayormente inciden en la variabilidad interanual de la precipitación del país e igualmente acepto como hipótesis: “que los procesos oceánicos y atmosféricos de meso y gran escala, así como la variabilidad interanual de los océanos Pacífico Tropical y Atlántico Norte y Sur tienen una significativa relación con el comportamiento de la precipitación en Colombia”, así como además examino la sensibilidad de los índices que sintetizan estos procesos, de manera tan cuidadosa como para poder seleccionar los de mayor conveniencia para la detectar dicha asociación.

Montealegre (2009,[20]) analizó cerca de cincuenta de las variables que presentan los centros mundiales especializados, las que clasificó en 7 grandes grupos<sup>1</sup>, así como algunas otras producto del análisis de la correlación espacial; dado

<sup>1</sup>(Atmosféricas, Precipitación, ENOS, Temperatura superficial del mar (TSM) en el Pacífico, TSM en el Atlántico y De otra índole )

que encontró que es posible que algunos de estos índices caractericen el mismo proceso desde ángulos diferentes, en su estudio evitó tal redundancia mediante un cuidadoso proceso de selección, que le permitió concluir, que los fenómenos de mayor afectación y sus respectivos índices mas relevantes, con objetivos de teleconexiones son:

- El Niño – La oscilación del sur:
  1. Índice de la Oscilación del sur SOI
  2. Índice Oceánico del Niño-ONI
  3. Índice multivariado – MEI
  4. La temperatura superficial del mar en la región 3 y 3.4
- Vientos de la Alta troposfera tropical
  1. Valor de los vientos en la alta troposfera tropical
- NAO - North Atlantic Oscillation
  1. Índice de la oscilación del atlántico norte
- QBO – Oscilación Cuasibienal.
  1. Índice de la Oscilación Cuasibienal.
- OLR – Radiación de onda larga saliente
  1. Valor de la radiación de onda larga saliente
- ZCIT – Zona de Confluencia Intertropical
  1. Seguimiento sinóptico de la ZCIT
- Dipolo Amazónico
  1. Anomalía de humedad relativa.

Del estudio de Montealegre (2009,[20]) , son tomados como sobresalientes e incluidos como variables independientes dentro de Jazikü, todas excepto la ZCIT, por no poseer índice numérico para su evaluación. A continuación se presenta un breve descripción de cada índice y la Tabla 4.2, informa al usuario los rangos para los cuales el software acepta su ingreso.

Principales índices de variabilidad climática que pueden ser utilizados:

Cuadro 4.2: Rango de validez en el cual la variable independiente debe ser incluida en Jazikü

Índice	Unidades	Rango de variación	Fuente	Condición por encima de lo normal	Condición por debajo de lo normal
Oceanic Niño Index.ONI región 3.4-<ONI>	Anomalia	-5<=ONI<=5 <sup>1</sup>	NOAA, Climate Prediction Center, Cold and Warmepisodes by season. Período 1950-2007: <a href="http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml">http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml</a>	El Niño: cuando es mayor a 0.5[20]	La Niña: cuando es menor a -0.5[20]
Índice de la Oscilación del Sur SOI-<SOI>	Anomalia estandarizada	-7<=SOI<=7 <sup>1</sup>	NOAA, Climate Prediction Center, <a href="http://www.cpc.ncep.noaa.gov/data/indices/">http://www.cpc.ncep.noaa.gov/data/indices/</a>	La Niña: cuando es mayor a 0.9[20]	El Niño: cuando es menor a -1.2[20]
El Índice Multivariado del ENSO, MEI-<MEI>	Composición de valores	-4.552<=MEI<=6.078	<a href="http://www.esrl.noaa.gov/psd/enso/mei.ext/rank.ext.html">http://www.esrl.noaa.gov/psd/enso/mei.ext/rank.ext.html</a>	El niño: Cuando es mayor al p66[6]	La Niña: Cuando es menor al p33[6]
Radiación de longitud tropical de onda larga, OLR tropical OLR -<OLR>	W/m2	-6<=OLR<=6 <sup>1</sup>	<a href="http://www.cpc.ncep.noaa.gov/data/indices/">http://www.cpc.ncep.noaa.gov/data/indices/</a>	Cuando es mayor a 0.9[20]	Cuando es menor a -1.1[20]
El índice de la anomalía del viento a 200 hPa,-<W200 >	Anomalia estandarizada	-7.5<=W200<=7.5 <sup>1</sup>	<a href="http://www.cpc.ncep.noaa.gov/data/indices/">http://www.cpc.ncep.noaa.gov/data/indices/</a>	El niño: Cuando es mayor al p66[20]	La Niña: Cuando es menor al p33[20]
Temperatura superficial del Mar (3.4)-<TSM>	°C	-60<=TSM<=60 <sup>1</sup>	<a href="http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices">http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices</a>	El niño: Cuando es mayor al p66[20]	La Niña: Cuando es menor al p33[20]
%HR amazonas-<ARH>	%	-100<=AHR<=100	<a href="http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml">http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml</a>	El niño: Cuando es mayor al p66[20]	La Niña: Cuando es menor al p33[20]
Oscilación del Atlántico Norte-<NAO>	anomalía	-6.36<=NAO<=6.08	<a href="http://www.cpc.noaa.gov/products/precip/CWlink/pna/norm.nao.monthly.b5001.current.ascii">http://www.cpc.noaa.gov/products/precip/CWlink/pna/norm.nao.monthly.b5001.current.ascii</a>	Cuando es mayor a 0[20]	Cuando es menor a 0[20]
Oscilación Cuasibienal-<QBO>	Km/h	-59.1<=QBO<=33.24	<a href="http://www.cdc.noaa.gov/Correlation/qbo.data">http://www.cdc.noaa.gov/Correlation/qbo.data</a>	El niño: Cuando es mayor al p66[20]	La Niña: Cuando es menor al p33[20]

<sup>1</sup> Criterio: 2 veces el máximo y mínimo de la serie



- El índice Oceánico El Niño (ONI): Criterio del Centro de Predicción Climática, (NCEP, por sus siglas en ingles) de NOAA (National Oceanic and Atmospheric Administration, por sus siglas en ingles), es conocido corrientemente como El Niño oceánico. El ONI (Oceanic Niño Index) se obtiene mediante el promedio móvil de la anomalía de la temperatura de la superficie del mar, en relación con el período de referencia 1971-2000, con tres meses sucesivos: la anomalía del mes, la anomalía del mes que le antecede y la del mes que le sigue, en la región Niño 3.4 (5° N a 5° S y 120°W a 170°W)[9]. La formula mas general del ONI es:

$$ONI = (Ai - 1 + Ai + Ai + 1)/3 \quad (4.1)$$

Donde,  $Ai$ , es la anomalía de la Temperatura Superficial del Mar del mes  $Ai-1$ , y  $Ai+1$ , anomalía de la Temperatura Superficial del Mar de los meses precedentes y siguientes al mes dado. Un ONI es el promedio móvil de las anomalías de la superficie del mar de tres meses consecutivos, siendo el ONI de enero, por ejemplo, el constituido por la anomalía de enero, febrero, y la anomalía de diciembre del año anterior (DEF); el segundo ONI de tres meses, EFM, y así sucesivamente hasta el último ONI del año, con los datos de los meses noviembre, diciembre, y enero del año siguiente, (NDE).

Aplicando el concepto del índice ONI de la NOAA, la definición cuantitativa de El Niño, La Niña y El ENSO quedan simplificadas de la manera siguiente: Un Niño o evento Niño, (para la región Niño 3.4 del Océano Pacifico ecuatorial) como una secuencia de al menos 5 ONI, cada uno con magnitud mayor o igual que 0,5 °C. Una Niña o evento La Niña, también para la región Niño 3.4 como la secuencia de al menos 5 ONI, cada uno con magnitud igual o menor que -0,5°C. Un ENSO o evento ENSO es la secuencia en la región Niño 3.4, en el Océano Pacifico ecuatorial de al menos 5 ONI consecutivos, cada uno con valor mayor o igual que 0,5°C (un Niño) o igual o menor que -0,5°C(una Niña), es decir, que para cumplir con el criterio se requiere información de las anomalías de la temperatura del mar durante siete meses consecutivos [9].

Desde diciembre de 2008 el ONI es calculado usando la versión 3b de la reconstrucción de la serie extendida de la temperatura superficial del mar (por sus siglas en ingles extended reconstructed sea surface temperature, ERSST). Una mayor información de este nuevo conjunto de datos es dada en

([http://www.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/](http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/))

- Índice de la Oscilación del Sur: Los patrones de calentamiento y enfriamiento cíclicos de la atmósfera, que se desarrollan en las zonas del centro y del este del Pacífico, coexisten con un cambio significativo de la presión atmosférica a nivel del mar. Esta diferencia de presión corresponde a la denominada Oscilación del Sur (OS).

Cuando la presión atmosférica aumenta en el lado este del Pacífico (y desciende en el lado oeste), se dice que la OS está en su fase positiva (La Niña). Por el contrario, cuando la variación de la presión atmosférica cambia de sentido

hacia la fase negativa (El Niño), la presión se eleva en el oeste y disminuye en el este. Esta alternancia entre las fases positiva y negativa del SOI, contribuye a cambiar la intensidad de la circulación de los vientos descritos anteriormente, debido a que un aumento/ disminución de esta diferencia de presión causa que el aire superficial, que normalmente sopla hacia el oeste, se intensifique/ debilite dependiendo del signo del SOI imperante.

El Índice de Oscilación del Sur (SOI, por sus siglas en ingles), utiliza los registros de presión atmosférica de las estaciones meteorológicas de Darwin–Australia (lado oeste del Pacífico) y Papeete– Tahití (lado este del Pacífico). El SOI es negativo cuando los valores de presión atmosférica son relativamente más altos que la media normalizada sobre Indonesia y Australia (estación meteorológica de Darwin), y más bajos que la media normalizada sobre la Polinesia Francesa (estación meteorológica de Papeete). Si al mismo tiempo, se observan TSM más cálidas que lo normal sobre el Pacífico ecuatorial central y oriental, y ambas condiciones se mantienen por unos meses, podemos decir que nos encontramos en presencia de un evento El Niño en pleno desarrollo (tomado de [9]). Las anomalías de este índice son basadas respecto al periodo 1981-2010 y el SOI es calculado de la siguiente manera [9]:

$$StandardDeviationTahiti = SQRT(SUMMATION(1)/N) \quad (4.2)$$

Donde SUMMATION(1) es la suma de todos los  $((TA) ** 2)$ , N es el numero de meses y TA es la anomalía en Tahití, calculada como:

$$(actual(SLP) - mean(SLP)) \quad (4.3)$$

Donde SLP es el la presión a nivel de superficie. Luego el valor de estandarizado de la anomalía en Tahití SLP y la desviación estándar en Darwin (Standard Deviation Darwin) son dados como:

$$StandardizedTahiti = (ActualTahiti(SLP) - MeanTahiti(SLP)) \quad (4.4)$$

$$StandardDeviationDarwin = SQRT(SUMMATION(1)/N) \quad (4.5)$$

Donde de nuevo SUMMATION(1) es la suma de todos los  $((DA) ** 2)$ , N es el numero de meses y DA es la anomalía en Darwin, calculada como:

$$Darwinanomaly = (actual(SLP) - mean(SLP)) \quad (4.6)$$

El valor de estandarizado de la anomalía en Darwin SLP y la desviación estándar en Darwin) son dados como:

$$StandardizedDarwin = (ActualDarwin(SLP) - MeanDarwin(SLP)) \quad (4.7)$$

Por lo tanto para calcular la desviación estándar media se usa:

$$MonthlyStandardDeviation(MSD) = SQRT(SUMMATION(3)/N) \quad (4.8)$$

En esta ultima expresión SUMMATION(3) es la suma de ((Standardized Tahiti - Standardized Darwin) \*\* 2), para que finalmente la ecuación del índice SOI sea:

$$SOI = (StandardizedTahiti - StandardizedDarwin)/MSD \quad (4.9)$$

- Índice Multivariado del ENSO(MEI): Es un complejo índice diseñado por Klaus Wolter (1987[26]) de NOAA-CIRES Climate Diagnostics Center (CDC), para reconocer el fenómeno de El Niño- Oscilación del Sur, ENSO. Este índice emplea como componentes principales las seis variables: temperatura de la superficie del mar, temperatura del aire, presión atmosférica al nivel del mar, nubosidad, componente zonal del viento y componente meridional del viento.

Los datos de estas variables son publicadas en ICOADS (Internacional comprehensive Ocean Atmosphere Data Set, (anteriormente, COAD). Los sitios de ubicación de donde provienen los datos de las variables utilizadas en el MEI son señalados por Wolter y Timlin (1998[?]) en la Fig.4.1 . Según Wolter (1987[26]) y Wolter and Timlin (1998[?]) el índice MEI sería un promedio ponderado de los rasgos del ENSO de las variables empleadas en su cálculo, y es obtenido mensualmente para los dos meses precedentes. Por la metodología empleada, el MEI es el primer componente principal no rotado, de las seis variables utilizadas. Mediante la normalización con el período de referencia 1950-93, el índice MEI se hace comparable con otros índices ENSO. El MEI es una media móvil de dos meses: así, el primer valor del año es el promedio de diciembre del año anterior y enero; el segundo, de enero y febrero,...,hasta noviembre y diciembre del año siguiente. En este caso índices con signos positivos representan la fase caliente del ENSO (El Niño) y los negativos del MEI, la fase fría (La Niña) (Tomado de [9]).

En la Fig.4.1 :

P+ y P-, representan la diferencia entre la presión al nivel del mar en los sitios más occidental y oriental del Pacífico ecuatorial, respectivamente.

U, la anomalía de la componente zonal del viento.

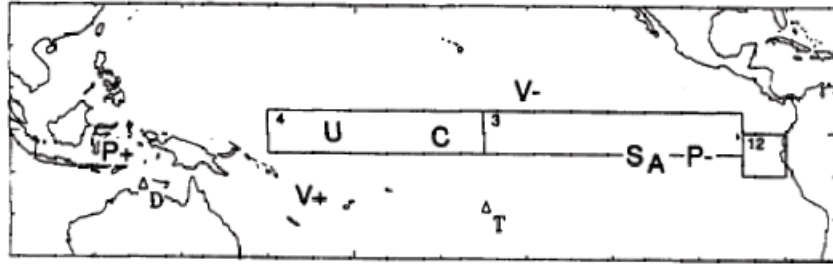
V+ y V-, la anomalía de la componente meridional del viento (viento del sur, positivo, viento del norte, negativo).

SA, ubicación de la temperatura media de la superficie del mar .

C, la anomalía de la nubosidad, en la región Niño 4.

T, Taití y D, Darwin.

1; 2; 3; 4, las regiones El Niño 1+2; 3, y 4, respectivamente.



Fuente: Wolter y Timlin (1998).

Figura 4.1: Sitios de observación del MEI

Para incluir la serie de tiempo del índice MEI, se aconseja al usuario utilizar las series de rangos del MEI disponible desde enero de 2001 en : <http://www.esrl.noaa.gov/psd/enso/mei.ext/rank.ext.html>. Para interpretar esta nueva forma del índice se pueden utilizar los terciles o los quintiles, en este caso son usados los terciles, marcando positivo (fase El Niño) a los valores por encima del percentil 66, siendo negativo (fase La Niña) para valores por debajo del percentil 33 [4].

- Índice de radiación de longitud de onda larga (OLR tropical, o simplemente OLR): Se refiere a la radiación emitida de onda larga por el tope de las nubes en el océano Pacífico tropical, expresado como anomalías en  $W/m^2$ , constituye una medida de la temperatura de ese tope nuboso y la actividad convectiva. Donde OLR es negativo, el tope de la nube es alto, indicando la presencia de convección y potencialmente, zonas más lluviosas, mientras que si el OLR es positivo, restringe las convecciones, la nubosidad y las posibilidades de precipitación (Tomado de[9]).
- La temperatura superficial del mar (TSM): Valioso indicador de la estructura térmica de la capa superior de los océanos. En consecuencia, se utiliza la TSM como variable explicativa de esta dinámica, tomando en consideración también que esta variable está estrechamente relacionada con los procesos convectivos que se generan en la atmósfera que se yace sobre ella. Para la vigilancia de la variabilidad de la TSM en la franja tropical del Pacífico, se han establecido cuatro regiones muy bien definidas a saber: Occidental (región Niño 4), Central (región Niño 3), Centro-occidental: (región Niño 3-4) y Oriental: (región Niño 1+2). En cada una de estas regiones se mide la temperatura de la superficie del mar en grados Celsius y se determinan sus anomalías, evaluadas como la diferencia entre un valor mensual cualquiera y su promedio climatológico o valor normal [20].

Como El Niño y La Niña son fenómenos de carácter oceánico, la comunidad científica ha desarrollado una gran cantidad de índices para caracterizar estos

fenómenos. Probablemente los más utilizados en la actualidad son las anomalías de TSM en las regiones Niño 3 y Niño 3-4. Analizadas las estadísticas descriptivas de las variables explicativas y de sus relaciones de asociación, los ciclos y frecuencias comunes, la complejidad del proceso físico involucrado, el grado de asociación con las precipitaciones, su variabilidad y algunos otros aspectos clasificatorios, Montealegre 2009 ([20]) encontró que para el desarrollo del análisis de Contingencia y el análisis Compuesto la TSM del Pacífico tropical centro occidental, (El Niño - región N-34) es uno de los índices más adecuados.

- Índice de la anomalía del viento a 200 hPa (WIND200): Permite apreciar el patrón del viento zonal en la alta troposfera (200 hPa) en comparación con el flujo al nivel isobárico de 850 hPa, a razón de anomalías opuestas. El mayor interés por este flujo de viento del oeste en altura sobre los trópicos, expresado por su anomalía se explica por la relación encontrada con la circulación en las zonas extratropicales. (Tomado de [9]).
- Índice de la Oscilación del Atlántico Norte: este fenómeno es un subí y baja a gran escala en la presión barométrica de las masas de aire atmosférico, situadas entre dos celdas, la de la alta presión subtropical del Atlántico (Azores), ubicada a los 38° de latitud y la baja presión polar del Ártico (Islandia) a los 60° de latitud. Esta oscilación tiene dos fases que se expresan mayoritariamente en invierno. La fase positiva sucede cuando las presiones barométricas del anticiclón de las Azores son más altas con respecto al valor promedio, mientras que las presiones en Islandia son aún más bajas de lo normal. Ésta marcada diferencia de presión intensifica los vientos del Oeste entre los 50 y 60° de latitud Norte produciendo tormentas que cruzan el Atlántico en dirección Nordeste [23]; Durante la fase negativa, las diferencias de presión entre las Azores e Islandia son menores. El "anticiclón" de las Azores está débil y la celda ciclónica de Islandia se encuentra poco activa; ambas se sitúan en latitudes más al Sur y, en consecuencia, los vientos del Oeste son más tenues y aportan menos humedad y calor sobre el Norte de Europa. La NAO fue identificada por primera vez en a principios del siglo XX por Sir Gilbert Walker (1924) e históricamente ha sido definida mediante un índice que mide la diferencia de presión en superficie entre Ponta Delgada (Azores) y la estación de Stykkisholmur (Islandia), la versión mas reciente es la propuesta por Jones et al. (1997,[11]) que sustituye Ponta Delgada por Gibraltar y Stykkisholmur por Reykjavik [11].
- Índice de la Oscilación Cuasibienal: la oscilación Cuasibienal es una oscilación de largo plazo en el viento zonal de la baja y media estratosfera ecuatorial, con un período irregular que varía entre un poco menos de dos años (20 meses) y hasta casi tres años (35 meses), con una duración media de 28 meses [20]. Consiste en la alternancia de vientos del Este, dominantes durante una parte del período, con vientos del Oeste en el resto del período. La Oscilación Cuasibienal se calcula sobre el flujo zo-

nal del viento en un nivel alto de la atmósfera a 30mb sobre la región tropical[16] .

- Dipolo Amazónico:Índice de meso-escala creado por Montealegre (2009 [20])con base en el contraste de humedad observado en dos puntos de la selva Amazónica al nivel de la media Troposfera (dipolo del Amazonas), se trata del promedio de humedad en el Amazonas a 500hPa, región  $2^{\circ}$  N –  $73^{\circ}$  W y  $6^{\circ}$  S –  $68^{\circ}$  W, provenientes del reanálisis NCEP/NCAR versión 1.

## Capítulo 5

### Jazikü - Clima

#### 5.1. Información calculada por Jazikü

#### 5.2. Análisis compuesto y Tablas de contingencia

En estadística las tablas de contingencia se emplean para registrar y analizar la relación entre dos o más variables, la asociación entre dos variables continuas está determinada por la distribución de probabilidad conjunta. En general esta distribución es desconocida, y en su defecto, se sustituye por la tabla de contingencia. Para construir esta tabla las variables continuas se dividen en categorías, la variable independiente X en M categorías y variable dependiente Y, en N categorías.

Cada pareja de valores  $(x_i, y_j)$  pertenece a una y sólo una de las  $M \times N$  categorías conjuntas. Luego se calculan las frecuencias empíricas  $f_{ij}$  que son el número de parejas que pertenecen a la categoría  $ij$ . Si la asociación entre las dos variables es muy débil, la población de la  $M \times N$  categorías conjuntas es similar (ver Fig.5.1) En un gráfico de tres dimensiones la superficie aparece plana porque los valores de X dentro de la categoría i pueden estar asociados indistintamente con valores que pertenecen a cualquiera de las N categorías de Y.

Si la asociación es fuerte (correlación directa), la superficie alcanza valores grandes a lo largo de una de las diagonales y valores bajos en las esquinas que no pertenecen a la diagonal y los valores altos se dan a lo largo de la diagonal mayor y en el caso de asociación negativa (correlación inversa), a lo largo de la diagonal menor (ver Fig.5.2).

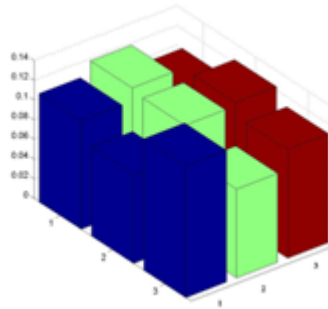


Figura 5.1: Ejemplo de variables no correlacionadas. Tomado de: [5]

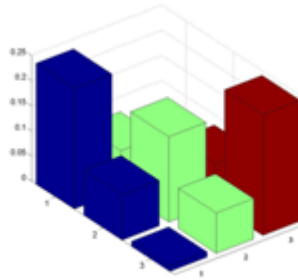


Figura 5.2: Ejemplo de variables correlacionadas. Tomado de: [5]

### 5.3. Construcción e interpretación de la Tabla de contingencia

Para construir una Tabla de contingencia que permita interpretar gráficos como los presentados en las Figs. 5.1 y 5.2, debe dividirse el conjunto de datos de la variable dependiente, en conjuntos seleccionados, que en este caso son los terciles por lo que se obtienen 3 categorías conjuntas que permiten cierto grado de resolución y un número manejable de posibilidades. Del mismo modo la variable independiente es categorizada en 3 conjuntos que también permiten hacer una selección de  $3 \times 3$  que frecuentemente es buscada para categorizaciones tales como: normal, arriba de lo normal y abajo de lo normal.

Se trata entonces de encontrar las probabilidades empíricas a partir de la división de las frecuencias empíricas y el número total de parejas se que se obtienen de ello. Es decir  $p_{ij} = f_{ij} / n$ , es la probabilidad empírica que una pareja de valores pertenezca a la categoría conjunta  $ij$ .

Entonces si se quiere hacer una validación cruzada entre un conjunto de



Cuadro 5.1: Ejemplo de Tabla de contingencia

categorías Vardep/VarIndp	Ind. Bajo	Ind. Normal	Ind. Alto
No. De dato debajo del primer tercil	OB-PA	OB-PA	OB-PA
No.de datos entre los dos terciles	ON-PA	ON-PA	ON-PA
No. De dato encima del primer tercil	OA-PA	OA-PA	OA-PA

observaciones como la variable independiente, o predictora y un conjunto de valores cuya proyección a futuro se desea pronosticar como variable dependiente, predictante. Las observaciones son divididas en terciles, dentro de las categorías de la variable explicativa como los valores bajos (Obs. Bajo), los valores normales (Obs. Normal) y valores altos (Obs. Alto), de manera que cada observación o valor tiene que pertenecer a una sola categoría y estas son excluyentes a su vez, lo que garantiza que las tres categorías agotan todas las posibilidades dentro de la “muestra” o conjunto de observaciones o valores tomados.

Como las categorías son terciles, cada categoría es un tercio del área total. Al considerar las parejas de (valor observado, valor pronosticado) son identificables nueve categorías excluyentes y exhaustivas, por ejemplo, un valor bajo de la observación puede estar asociado con un valor bajo (OB-PB), uno normal (OB-PN) o uno alto (OB-PA) del valor pronosticado, de ahí que en el caso trivial a cada categoría le corresponde un tercio del área total. Entonces si la variable explicativa es un índice cualquiera divisible en categorías alto, normal y bajo; y la explicada una variable meteorológica, es posible construir una Tabla de la frecuencia relativa de la variable explicada por debajo del primer tercil, por encima de este, o entre estos dos (que con frecuencia es llamado en lo normal “ON”), si es dado un índice alto, bajo o normal (ver Tabla 5.1) .

Generalmente, los valores individuales de  $p_{ij}$  o  $f_{ij}$  de la Tabla 5.1 (los valores OB-PA, OB-PN, OB-PB..., OA-PB) pueden ser expresados en términos de porcentajes, según la Ec.5.2

$$P_{ij} = \frac{f_{ij}}{\sum_{j=1}^N f_{ij}} \quad (5.1)$$

Debe notarse que para eventos independientes, la probabilidad que ocurra simultáneamente es multiplicativa. Por lo tanto, la probabilidad de que una pareja (observación, pronóstico) corresponda a este caso es dada por la Ec.5.2

$$P_r \{Q_2 \cap P_3\} = P_r \{Q_2\} P_r \{P_3\} = 1/3 * 1/3 = 1/9 \quad (5.2)$$

### 5.3.1. Correlación lineal

De especial interés para el uso de la metodología del análisis compuesto, se encuentra el calculo del coeficiente de correlación, este permite cuantificar el grado de asociación de las dos variables sin considerar a una como dependiente y a la otra independiente, puesto que se asume que las dos tienen el mismo

estatus[22]. Este coeficiente de correlación  $r$ , también denominado de Pearson, se define como:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (5.3)$$

Donde  $X$  e  $Y$  pueden intercambiar su lugar y el resultado no varia, el coeficiente de correlación puede arrojar valores dentro del rango  $-1 \leq r \leq 1$ , donde el signo indica el sentido de la asociación (directa o inversa, siempre lineal); las asociaciones más fuertes son  $-1$  y  $1$ , para  $r = 0$  se habla de falta de asociación. Una conclusión acerca de este resultado, se puede basar en el cálculo de la significancia estadística de la correlación, si se asume una distribución normal esta prueba de hipótesis se hace con el estadístico  $t$  para un  $r$  dado, según[12]:

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{1-r^2}} \quad (5.4)$$

Donde el numero de grados de libertad se da por  $n-2$ . En esta prueba la hipótesis nula es que  $r = 0$ , por lo tanto la alternativa es que las variables de hecho están correlacionadas[12]. Típicamente esta formula es usada para evaluar la significancia tanto del coeficiente de correlación de Pearson, como el de Sperman[3] (Glasser & Winter, 1961[8]; Kendall, Kendall, & Smith, 1939[14]). El usuario de Jazikü, tendrá a su disposición el valor del estadístico  $r$  y de la diferencia  $p$ -valué, para que los contraste a cualquier margen de incertidumbre, se aconseja 0.05 (95 % de confianza)

### 5.3.2. Prueba de bondad de ajuste $\chi^2$

El primer enfoque inferencial aplicado a datos categóricos se desarrollo a principios de siglo y trataba de determinar si un conjunto de probabilidades de una tabla de contingencia eran iguales a ciertas cantidades prefijadas[18]. Este es el clásico problema de bondad de ajuste. El proceso analítico se propone evaluar la discrepancia existente entre las frecuencias observadas ( $n_{ij}$ ). Freeman (1987[7]) distingue tres pruebas estadísticas de bondad de ajuste basadas en la distribución  $\chi^2$ , se presentan dos:

- La mas conocida es la clásica prueba  $\chi^2$ , introducida por K.Pearson (1900[24]), que se define como [18]:

$$\chi^2 = N \sum_{i=1}^a \sum_{j=1}^b \frac{(p_{ij} - \pi_{ij})^2}{\pi_{ij}} \quad (5.5)$$

$$\chi^2 = \sum_{i=1}^a \sum_{j=1}^b \frac{(n - m_{ij})^2}{m_{ij}} \sim \chi^2_{(a-1)(b-1)} \quad (5.6)$$

La primera formulación en términos de probabilidades empíricas y esperadas y la segunda en términos de valores observados y esperados[18].

Pearson planteó la utilización del estadístico  $\chi^2$  para analizar la independencia. La hipótesis nula a contrastar ( $H_0$ ) será la de independencia entre los factores, siendo la hipótesis alternativa ( $H_a$ ) la de dependencia entre ellos [13]. El valor calculado de  $\chi^2$  se compara con el valor generado por Jazikü de una distribución  $\chi^2$  para un nivel de confianza del 95 % y N-1 grados de libertad (siendo N el número total de trimestres de las series de la variable dependiente e independiente). Si el valor calculado es mayor que el valor generado de la distribución  $\chi^2$  (de Jazikü), se rechaza la hipótesis nula, con lo cual podemos concluir que si existe una relación entre las variables; por el contrario si el valor calculado es menor no se rechaza la hipótesis nula aceptando que no existe ninguna relación entre las variables (todo a un nivel de confianza del 95 %)<sup>1</sup>[13].

$$si \chi^2 > \chi^2_{\text{Crit}} \Rightarrow \text{Rechaza } H_0(\text{dependencia}) \quad (5.7)$$

$$si \chi^2 \leq \chi^2_{\text{Crit}} \Rightarrow \text{Acepta } H_0(\text{independencia}) \quad (5.8)$$

El problema de la prueba con el estadístico  $\chi^2$  es que está influenciada por el tamaño muestral, esto es que a mayor número de casos analizados (a mayor N), el valor de  $\chi^2$  tiende a aumentar, por lo que cuanto mayor sea la muestra más fácil será rechazada la hipótesis nula de independencia; otro aspecto a tener en cuenta a la hora de realizar este contraste, es que para que el contraste sea estadísticamente válido en cada celda de la tabla deberían existir un mínimo de 5 observaciones [13].

### 5.3.3. Coeficiente V de Cramer

Medida de asociación de las variables de entrada de una tabla de contingencia de escala nominal h x k, con h y k > 2 (usualmente cuando h, k = 2, se usa el coeficiente Phi)[17], este coeficiente sigue la expresión:

$$V = \sqrt{\frac{\chi^2}{n(m-1)}} \quad (5.9)$$

siendo: n el tamaño muestral

m el mínimo entre h y k (el número de categorías de la variable independiente y k número de categorías de la variable dependiente)

Se verifica que  $0 \leq V \leq 1$ , en general valores cercanos o iguales a 0 indican muy baja asociación y valores cercanos a 1, alta asociación. Es importante resaltar que este coeficiente sólo proporciona información sobre la relación entre las variables y no sobre el sentido de la misma[10], de la misma manera que se ha encontrado que el grado de asociación puede ser subestimado[17].

---

<sup>1</sup>(Lo mismo que decir que el valor numérico p valúe arrojado por Jazikü debe ser menor a 0.05 para rechazar la hipótesis nula, con lo cual podemos concluir que si existe una relación entre las variables y lo contrario si el valor calculado es mayor que 0.05, con lo que no se rechaza la hipótesis nula aceptando que no existe ninguna relación entre las variables. Ver código fuente)

### 5.3.4. Análisis de riesgos

#### 5.3.4.1. Distribución hipergeométrica<sup>2</sup>

Notación:  $h(x; N, n, M)$

Definición: Muchas veces en la práctica es difícil realizar pruebas con reposición ó reemplazamiento. Por ejemplo, si en el control de calidad se pierde el elemento que se prueba, pues no se puede hacer reposición directamente. Se plantea entonces la prueba sin reposición, donde los elementos de la muestra se toman todos a la vez y no individualmente ó donde el elemento seleccionado no se reintegra al experimento ó a la muestra nuevamente.

La diferencia mas simple con la binomial es la forma de aplicar el muestreo. En efecto, en:

*Binomial*: Muestreo con reemplazamiento e independencia de pruebas ó ensayos.

*Hipergemoetrica*: Muestreo sin reemplazamiento y sin indepedencia entre pruebas ó ensayos.

Sus aplicaciones estan en areas con uso considerable de muestreo de aceptación, pruebas electronicas y de aseguramiento de la calidad, fabricación de piezas, etc.

En la distribución Hipergeométrica  $X$  cantidad de resultados éxitos en una muestra aleatoria (sin reposición) de tamaño  $n$ , tomada de una población de tamaño  $N$  y de la cual  $M$  satisface una característica ó propiedad (éxito) antes del muestreo y  $N - M$  no la satisface (fracaso).

Criterios ó propiedades que la caracterizan.

1. La población  $N$  del conjunto de unidades ó elementos es de orden finito, de los cuales una parte:  $M$  "son éxitos", y otra parte:  $N - M$  son "fracasos".
2. Cada elemento puede ser caracterizado como éxito ó fracaso.
3. Se obtiene una muestra aleatoria de  $n$  elementos todos a la vez (sin reemplazamiento) y no de forma independiente. No son pruebas repetidas.
4. El tamaño de la muestra aleatoria  $n$  es grande relativamente en comparación con el tamaño de la población. Generalmente:  $\frac{n}{N} > 5\%$
5. Se busca la probabilidad de  $x$  número de éxitos a partir de los  $M$  resultados ó elementos y  $(n - x)$  fracasos a partir de los  $N - M$  elementos así clasificados, al obtener una muestra aleatoria de tamaño  $n$

Planteamiento:

Supongamos un lote de  $N$  productos de los cuales: Obtenemos muestra de  $n$  productos, todos a la vez. Interesa entonces la probabilidad de sacar  $x$  productos defectuosos (Exito), o sea:  $p(x)$ . Podemos hacer el siguiente raciocinio: De una población de  $N$  elementos se pueden extraer muestras de tamaño  $n$  de  $\binom{N}{n}$  formas diferentes (distintas muestras de tamaño  $n$ ). Al extraer muestras de tamaño  $n$  productos, el número de formas de obtener  $x$  productos defectuosos de  $M$  de ellos será:  $\binom{M}{x}$  y entonces  $\binom{N - M}{n - x}$  sera el nú-

<sup>2</sup>Tomado de <[http://www.virtual.unal.edu.co/cursos/ciencias/2001065/html/un2/distribucion\\_hipergeometrica.html](http://www.virtual.unal.edu.co/cursos/ciencias/2001065/html/un2/distribucion_hipergeometrica.html)>

mero de formas de obtener  $(n - x)$  productos no defectuosos entre  $N - M$  de ellos.

Como es el mismo evento compuesto, entonces el número de formas de seleccionar  $x$  productos defectuosos esta ligado con el número de formas de obtener  $(n - x)$  productos no defectuosos. Luego el total de formas posibles

sera:  $\binom{M}{x} \binom{N-M}{n-x}$

Combinando los casos

$$h(x, n, M, N) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \text{para } x = 0, 1, 2, \dots, nx \leq M; (n-x) \leq N-M \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Los parámetros de la distribución Hipergeométrica son entonces:

$N$ : Tamaño de población.

$M$ : Número de elementos de  $N$  con una característica ó propiedad específica (éxitos).

$n$ : Tamaño de muestra aleatoria extraída.

Nota: Algunos tratadistas simbolizan esta distribución con:  $h(x; N, n, M)$

Características de la Distribución Hipergeométrica.

En la practica, si  $\frac{n}{N} < 5\%$ , no se aplica el (F.C.F.) pues su valor tendera a cero (1).

La función de distribución acumulativa quedará definida entonces por:

$$P(X \leq x) = f(x, n, M, N) = \sum_{x=0}^n \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad (5.10)$$

Ejemplo

En una empresa industrial diariamente se producen 90 unidades de unidad metalmecánica, de las cuales generalmente 5 salen defectuosas. Se examina en un dia cualquiera una muestra de 5 unidades. Hallar la probabilidad de  $x$  unidades defectuosas.

$$\left\{ \begin{array}{l} N = 90 \\ M = 5 \\ N - M = 85 \\ n = 5 \\ x = 0, 1, \dots, 5 \end{array} \right\} P(x) = \sum_{x=0}^n \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\binom{5}{x} \binom{85}{5-x}}{\binom{90}{5}}$$

que resolviendo permite definir la tabla de distribución de probabilidad:

x	0	1	2	3	4
P(x)	0.746	0.230	0.0225	0.000812	0.00000967

### 5.3.4.2. Metodología del análisis de riesgos

Con el objetivo de que los pronósticos arrojados por el programa, sean fiables, a continuación se presenta el marco teórico del análisis de riesgos. En este tipo de análisis una distribución hipergeométrica es utilizada para describir la distribución de probabilidad entre todas las posibles categorías de la variable independiente. Este tipo de distribución hipergeométrica es muy útil para procesos muestrales sin reposición, en los que se investiga la presencia o ausencia de cierta característica. Por tanto, esta distribución es la equivalente a la binomial, pero cuando el muestreo se hace sin reposición. La expresión general de la distribución es:

$$P(X = x) = f(x, n, M, N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \quad (5.11)$$

Donde:

$n$ : es el número de veces que se presentó una categoría por encima, por debajo ó dentro de lo normal de la variable independiente.

$N$ : corresponde al valor total de eventos de las tres categorías (por encima, por debajo y dentro de lo normal) de la variable independiente.

$M$ : Es número total de situaciones para las que se dio que la variable dependiente marcó condiciones por encima, por debajo y dentro de lo normal, en conjunción para una de las tres categorías de la variable independiente.

$x$ : El número de casos por encima, por debajo ó dentro de lo normal bajo alguna de estas mismas categorías para la variable independiente.

Esta distribución de probabilidad es aplicada a cada uno de los valores de la tabla de contingencia, en términos de frecuencia. A continuación un ejemplo.

Sea la tabla de contingencia 5.2, para cuando se quiere analizar la variación de la precipitación en función del fenómeno ENOS.

Cuadro 5.2: Tabla de contingencia de ejemplo

Var dep/Var Ind	El Niño	Neutro	La Niña
Por debajo de lo normal	5	8	7
Dentro de lo normal	2	5	7
Por encima de lo normal	7	10	3
Total	14	23	17

Supóngase que se desea saber si el valor de  $x=3$  correspondiente a 3 casos para los cuales la precipitación en la climatología se observó por encima de lo normal, bajo condiciones La Niña, es un valor significativo y puede ser usado

Cuadro 5.3: Valores de la distribución hipergeométrica para el ejemplo tratado, en la categoría por encima de lo normal para la variable dependiente.

Parámetro	Por encima Var D/El Niño	Por encima Var D/Neutro	Por encima Var D/La Niña
n	14	23	17
M	20	20	20
x	7	10	3
N	54	54	54

Cuadro 5.4: Valores de la distribución hipergeométrica para el ejemplo tratado, en la categoría dentro de lo normal para la variable dependiente.

Parámetro	Dentro de lo normal Var D/El Niño	Dentro de lo normal Var D/Neutro	Dentro de lo normal Var D/La Niña
n	14	23	17
M	14	14	14
x	2	5	7
N	54	54	54

Cuadro 5.5: Valores de la distribución hipergeométrica para el ejemplo tratado, en la categoría por debajo de lo normal para la variable dependiente.

Parámetro	Por encima Var D/El Niño	Por encima Var D/Neutro	Por encima Var D/La Niña
n	14	23	17
M	20	20	20
x	5	8	7
N	54	54	54

para futuros pronósticos. En este caso claramente  $n=17$ ,  $M=20$  y  $N=54$ . La Tablas 5.3, 5.4 y 5.5, presentan los parámetros de la distribución hipergeométrica para cada uno de los valores de la Tabla 5.2 .

La interpretación de los resultados es la siguiente: Una vez hallados los valores de los parámetros de la distribución hipergeométrica para cada uno de los valores asociados a la tabla de contingencia en cuestión, se halla el valor de la probabilidad discreta de que un determinado evento ocurra según

la distribución hipergeométrica  $P(x)$  o valor de probabilidad distribuida. La significancia estadística se determina mediante la comparación del valor  $P(x)$  respecto de si cae en alguna de las colas, derecha o izquierda de la distribución de probabilidad. Estos en relación al intervalo de confianza del 90 % de la distribución acumulada, así:

$$\begin{aligned} \text{Si } P_{CDF}(x) \leq 0,10 &\implies \text{El valor cae sobre la cola izquierda} \\ \text{Si } 1 - P_{CDF}(x) \leq 0,10 &\implies \text{El valor cae sobre la cola derecha} \end{aligned} \quad (5.12)$$

De comprobarse que una de las dos anteriores condiciones se cumple, es posible utilizar el valor  $x$ , para condiciones de pronóstico, de lo contrario la climatología y otras técnicas deben ser utilizadas. En el ejemplo indicado, para el que  $x=3$ , el valor de  $P(x) = 0,0337$ ,  $P_{CDF}(x) = 0,042116$  y  $1 - P_{CDF}(x) = 0,991537$ , indicando que el valor cae en la cola izquierda y es significativo por ser menor a 0.10.



## Capítulo 6

### Jazikü - Pronóstico

#### 6.1. Pronóstico operativo mensual de predicción climática de los posibles escenarios de impacto, en la variable dependiente debido al fenómeno ENOS

El Objetivo es realizar el pronóstico de, dada una fase del fenómeno ENOS la variable dependiente se vea afectada. Esto es posible si se tienen en cuenta los pronósticos de ocurrencia para cada fenómeno de la oscilación que es realizado por el IRI (International Research Institute for Climate and Society) y que corresponden a un análisis de la distribución de acontecimientos históricos ponderado por el centro de predicción climática en la región El Niño 3.4, en función de las previsiones de temperatura superficial del mar, que están disponibles en su página web : <http://portal.iri.columbia.edu/> en el link Enso Forecast y que son actualizados cada mes. Las Ecs. 6.1a 6.3, permiten hallar la probabilidad de que dadas las probabilidades de que en el próximo periodo el fenómeno El Niño – La oscilación del sur, se dé una fase El Niño, La Niña o Neutro, la variable dependiente en la locación de interés sea afectada por disminución, aumento o tendencia a la normalidad en su régimen.

$$P_{V-DepAbove} = \left\{ \begin{array}{l} [P_{V-DepAbove}(sí V - Ind_{above}) * (P_{IRI}V - Ind_{Above})] \\ + [P_{V-DepAbove}(sí V - Ind_{Normal}) * (P_{IRI}V - Ind_{Normal})] \\ + [P_{V-DepAbove}(sí V - Ind_{Below}) * (P_{IRI}V - Ind_{Below})] \end{array} \right\} \quad (6.1)$$

$$P_{V-DepNormal} = \left\{ \begin{array}{l} [P_{V-DepNormal}(sí V - Ind_{above}) * (P_{IRI}V - Ind_{Above})] \\ + [P_{V-DepNormal}(sí V - Ind_{Normal}) * (P_{IRI}V - Ind_{Normal})] \\ + [P_{V-DepNormal}(sí V - Ind_{Below}) * (P_{IRI}V - Ind_{Below})] \end{array} \right\} \quad (6.2)$$

$$P_{V-DepNormal} = \left\{ \begin{array}{l} [P_{V-DepBelow}(sí V - Ind_{above}) * (P_{IRI}V - Ind_{Above})] \\ + [P_{V-DepBelow}(sí V - Ind_{Normal}) * (P_{IRI}V - Ind_{Normal})] \\ + [P_{V-DepBelow}(sí V - Ind_{Below}) * (P_{IRI}V - Ind_{Below})] \end{array} \right\} \quad (6.3)$$

*Nota. En las Ecs. 6.1a 6.3, las primeras entradas de cada paréntesis corresponden a las salidas de la tabla de contingencia y las segundas a los pronósticos del fenómeno ENOS.*

Para presentar los resultados de las Ecs. 6.1a 6.3, se pueden utilizar distintos tipos de gráficos, particularmente la NOAA [1], recomienda el uso de diagramas circulares para representar la división porcentual de la probabilidad de tendencia deficitaria, de ganancia y a la normalidad de la variable dependiente analizada para una estación bajo las probabilidades de que el fenómeno se encuentre en su fase cálida, neutra o fría. Sin embargo gracias al uso operativo, este tipo de grafico puede ser cambiado a conveniencia del pronosticador.

Antes de emitir cualquier pronóstico, es necesario determinar si las relaciones halladas en la tabla de contingencia son significativas estadísticamente. Esto se logra mediante el ajuste a una distribución de probabilidad hipergeométrica de la distribución de probabilidad entre todos los posibles resultados de una categoría dentro El Niño o La Niña o Neutro. De manera tal, que la distribución hipergeométrica sea utilizada como una aproximación para describir la distribución de probabilidad entre todos los posibles resultados de una categoría dentro de las fases del fenómeno (ver sección 5.3.4). Entonces la creación de una curva de distribución basada en los valores de probabilidad discreta de la ocurrencia de un determinado evento, determina la significancia estadística del valor  $P(x)$  si se encuentra en la cola izquierda o derecha de la distribución [25], esto es si el valor  $x$  de la distribución asociado al valor  $n$ , en un test del 10 % de significancia se encuentra en algunas de las colas de la distribución, se acepta que el valor  $x$  es estadísticamente significativo y se puede aceptar con confianza el valor de la probabilidad de  $x$  [25].

## Capítulo 7

# Reconocimientos y licencias de uso

Jazikü mantiene la definición de software libre que permite a los usuarios ejecutar, copiar, distribuir, estudiar, cambiar y mejorar el software. Más precisamente<sup>1</sup>:

- \* Se tiene libertad de ejecutar el programa, para propósito académicos (i.e) (libertad 0) .

- \* La libertad de estudiar cómo trabaja el programa, y cambiarlo para propósitos propios.

- \* La libertad de redistribuir copias para que pueda ayudar al prójimo (libertad 2).

- \* La libertad de distribuir copias de sus versiones modificadas a terceros (la 3<sup>a</sup> libertad). Si lo hace, puede dar a toda la comunidad una oportunidad de beneficiarse de sus cambios. El acceso al código fuente es una condición necesaria para ello.

Es producto de los contratos 070/2011 y 066/2011 a nombre de Inés Sánchez Rodríguez y de Xavier Corredo Llano respectivamente, y dirigidos por la Dra. Gloria León Aristizabal. La propiedad y uso es del IDEAM, pero la autoría es de sus directos creadores. Agradecemos citar:

Sánchez Rodríguez, Inés Concepción; Corredor Llano, Xavier & León Aristizabal, Gloria. Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)-Grupo de tiempo y clima.(2011).En: Notas técnicas contratos <IDEAM - 070 de 2011> y <IDEAM - 066 de 2011> .

---

<sup>1</sup>Mayor información en: <<http://www.gnu.org/licenses/gpl.html>> ; <[www.gnu.org/philosophy/free-sw.es.html](http://www.gnu.org/philosophy/free-sw.es.html)>

## Parte III

# Apéndices

## Capítulo 8

# Apéndice 1: Posibles errores arrojados por Jazikü

### 8.1. Error por mala definición del período al correr Jazikü

Es posible que el usuario incluya en el comando estándar de corrida de Jazikü (ver sección 2), un período para el cual alguna de las variables (dependiente e independiente) no este definida. El mensaje mostrado es:

```
Setting period in: 1990-2100
```

```
##### Station: 4801501-01 (4801501)
```

```
Processing climate
```

```
ERROR: period define in argument 1990-2100 is outside the common  
period 1985-2009
```

```
For more help run program with argument: -h
```

### 8.2. Errores en el modulo de entrada de datos.

Un error general para cuando los valores fueron construidos como texto e implícitamente el editor los ha guardado bajo comillas es:

```
ERROR: 3 Reading from file 4801501_ppt.txt in line: 1. 4 invalid  
literal for int() with base 10: '1985'
```

#### 8.2.1. Variable precipitación

Un caso típico es incluir en la variable precipitación valores negativos o fuera del rango; a continuación se presentan el avisos dado por Jazikü para

cuando en la variable dependiente de la estación 4801501 en la línea 18 y 20 presentan dichos errores.

*ERROR: Reading from file 4801501\_ppt.txt in line: 20. exception validation in dependent variable: value out of range: ?PPT? type does not allow the value: -312.3 precipitation not valid*

*ERROR: Reading from file 4801501\_ppt.txt in line: 18. exception validation in dependent variable: value out of range: ?PPT? type does not allow the value: 10000.0 precipitation not valid*

### 8.2.2. Variable numero de días con lluvia

Para cuando se incluyen valores negativos y cuando la variable supera los valores de máximos de días para los meses, incluyendo los años bisiestos los errores son:

*ERROR: Reading from file 4801501\_NdPpt.txt in line: 14. exception validation in dependent variable: value out of range: ?NDPPT? type does not allow the value: -1.0 precipitation not valid*

*ERROR: Reading from file 4801501\_NdPpt.txt in line: 11. exception validation in dependent variable: value out of range: ?NDPPT? type does not allow the value: 31.0 day is out of range for month*

### 8.2.3. Variable temperatura media, máxima y mínima

Errores comunes en la inclusión de la variable temperatura, son principalmente el uso de valores fuera del rango, lo que es usual por errores en la transcripción de las cifras significativas. Cuando se presenta esta situación el error reportado por Jazikü es:

*ERROR: Reading from file 4801501\_Temp.txt in line: 219. exception validation in dependent variable: value out of range: «TMAX» type does not allow the value: 220.0 maximum temperature not valid*

*For more help run program with argument: -h*

### 8.2.4. Variable presión

En relación a la variable presión, son frecuentes errores en la transcripción de los datos, que pueden llevar a valores por fuera del rango de validez ya establecido, este error es reflejado por Jazikü como:

*ERROR: Reading from file 4801501\_ppt.txt in line: 233. exception validation in dependent variable: value out of range: «PATM» type does not allow the value: -15.0 atmosphere pressure not valid*

*For more help run program with argument: -h*

*ERROR: Reading from file 4801501\_ppt.txt in line: 180. exception validation in dependent variable: value out of range: «PATM» type does not allow the value: 3000.0 atmosphere pressure not valid*

*For more help run program with argument: -h*

### 8.2.5. Variable humedad relativa

Debido a que esta variable es dada en porcentaje, es posible que el usuario en vez de incluir valores en el rango de 0 a 100 lo haga en el rango de 0 a 1, por lo cual se pide extrema cautela al respecto. El error mas común también para esta variable, se relaciona con la transcripción de los datos, que pueden llevar a valores por fuera del rango de validez ya establecido, este error es reflejado por Jazikü como :

```
ERROR: Reading from file 4801501_Hrel.txt in line: 2. exception
validation in dependent variable: value out of range: «HR» type does
not allow the value: 232.9% of relative humidity not valid
```

```
For more help run program with argument: -h
```

```
ERROR: Reading from file 4801501_Hrel.txt in line: 1. exception
validation in dependent variable: value out of range: «HR» type does
not allow the value: -3.0% of relative humidity not valid
```

```
For more help run program with argument: -h
```

### 8.3. Errores en la variable independiente

El siguiente es un claro aviso de Jazikü para cuando el usuario esta intentado incluir en la variable independiente de temperatura superficial del mar fuera del rango de validación y/o negativo

```
ERROR: Reading from file TSM34.txt in line: 442. exception
validation in independent variable: value out of range: ?TSM?
type does not allow the value: -151.0 Sea surface temperature
not valid
```

```
ERROR: Reading from file TSM34.txt in line: 441. exception validation
in independent variable: value out of range: ?TSM? type does not allow
the value: 100.0 Sea surface temperature not valid
```

### 8.4. Error por inconsistencia de la tabla de contingencia.

Supóngase que el usuario ha escogido unos umbrales de tal tipo que (ver sección 2.1), para una determinada categoría de la variable independiente, no existe ni un caso para el cual la variable dependiente mostró variación alguna. Esto es, que en efecto los tres valores correspondientes a las categorías por encima, por debajo y dentro de lo normal de la variable dependiente son cero (dada la categoría de la variable independiente), o lo que es lo mismo un “n=0” en alguna de las categorías de la Ec. 5.11, usada para el calculo de la distribución Hipergeométrica.

Jazikü no realiza conversión a valores de porcentaje de la tabla de contingencia indicando con un valor “null” la imposibilidad de la operación. No obstante continua haciendo todo el proceso del modulo clima para las otras categorías de la variable independiente que si fuere el caso, si presentan valores

de frecuencias asociadas (casos en los que bajo una categoría específica de la variable Independiente, la respuesta de la variable dependiente es de un número diferente de cero eventos por debajo de lo normal, dentro de lo normal ó por encima de lo normal).

### 8.5. Error en generación de pronóstico .

Cuando se encuentra que uno de los valores provenientes de la tabla de contingencia utilizados en las Ecs.6.1,6.2,6.3, no pasa el análisis de riesgos (yes/no, ver sección3, Fig.3.4), Jazikü no cuenta con información suficiente para el calculo del pronostico asociado y no realiza el calculo. Se espera que los valores de la tabla de contingencia en cada una de las categorías sean siempre los tres confiables (pasen el análisis de riesgos) para poder obtener un pronostico óptimo.

### 8.6. Acerca del contenido de las carpetas “maps” generadas por el software

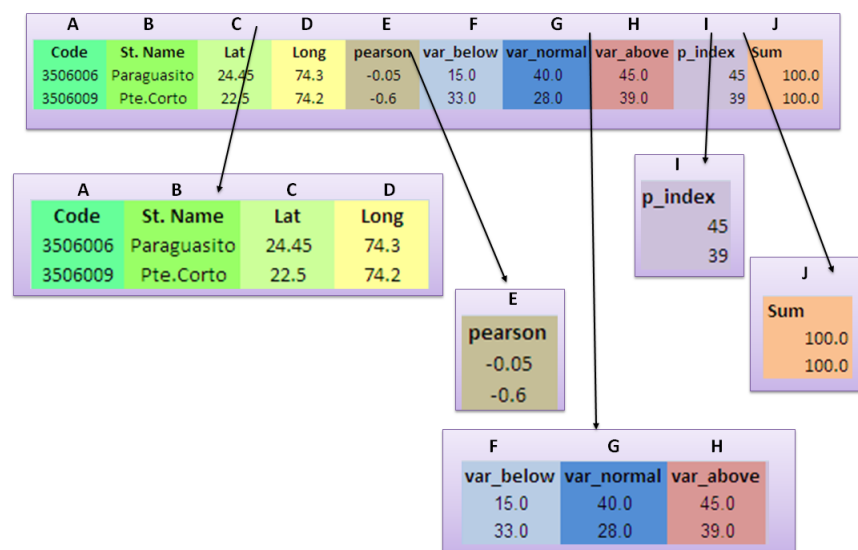


Figura 8.1: Esquema de las tablas de resultados de las carpetas maps

Los archivos guardados en la carpeta “maps” de los resultados de los módulos de clima y pronóstico de Jazikü, siguen el formato mostrado en la Fig. 8.1. En cualquier caso las columnas A,B,C y D, corresponden a el código de la estación, nombre de la estación y ubicación geográfica, conforme la información incluida por el usuario al inicio del proceso (ver sección 2.1). Cuando la



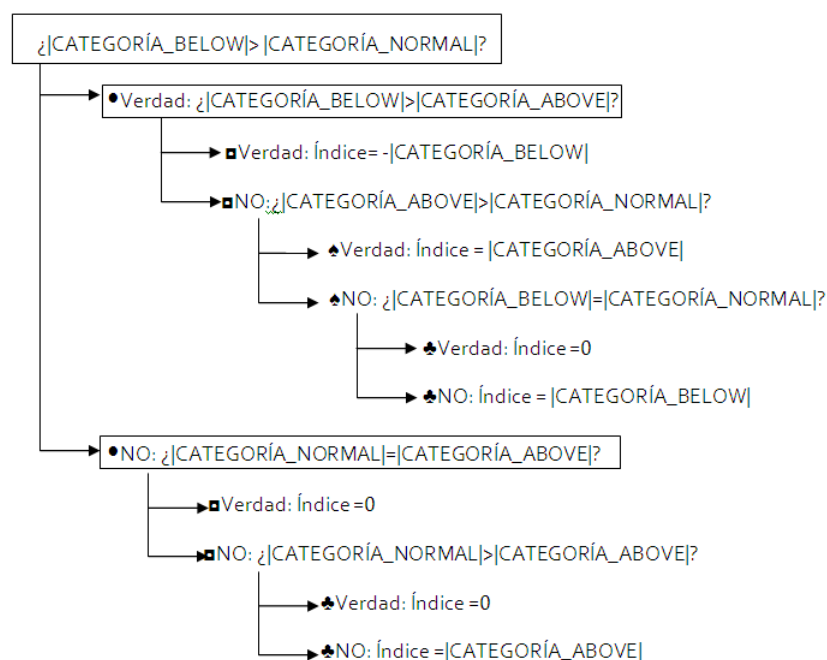


Figura 8.2: Ciclo IF para la columna índice.

tabla proviene del módulo Jazikü - clima, las columnas E,F y G, corresponden a los valores en porcentaje de la tabla de contingencia, para la fase del fenómeno, el trimestre y el rezago indicado en la dirección absoluta del archivo. En cambio para cuando los resultados provienen del módulo de pronóstico, las columnas E,F y G, corresponden al pronóstico de la variable dependiente según las probabilidades de la variable dependiente, resultado de la combinación lineal mostrada en la sección 6.1

La columna índice (H) se calcula según el ciclo IF de la Fig.8.2 y la columna suma (I) que es la suma de cada uno de los valores por debajo, por encima y dentro de lo normal bajo la categoría El Niño, La Niña o Neutro, siempre debe ser 100.

## 8.7. Ejemplo de aplicación del software

A continuación se presenta un ejemplo de aplicación del software para la serie de precipitación de la estación 4801501 (código IDEAM) correspondiente al aeropuerto Vázquez Cobo de la ciudad de Leticia en el departamento del

Cuadro 8.1: Información de la ubicación de la estación de la cual se extrajo la información de la variable precipitación utilizada para la revisión del software

DEPTO	CODIGO	NOMBRE	CAT	ELEV	LA_G	LO_G	MUNICIPIO
AMAZONAS	4801501	APTO VASQUEZ COBO	SP	84	-4°12'S	69°57'W	LETICIA

Amazonas colombiano (ver Tabla 8.1), en el período 1985-2009. Ya se ha comprobado que la serie es homogénea. La Fig.8.3 presenta la serie de la que de ahora en adelante será llamada variable dependiente; por su parte fue seleccionada la serie de la Temperatura Superficial del Mar en el pacifico tropical (región EL Niño 3.4) como variable independiente, su serie en el período 1950-2009 es mostrada en la Fig.8.4.

Para correr el ejemplo, digite en la Shell

```
jaziku -station stations.csv --climate --forecasting
-p_below "La Niña" -p_normal "Neutro" -p_above "El Niño"
```

Se espera que el software seleccione el período común como 1986-2009 . El archivo “stations.csv”, debe tener la estructura <sup>1</sup> (ver sección2.1).

```
4801501;4801501_ppt;"4.20";"-84.06";../4801501_ppt.txt;"PPT";
"/../TSM34.txt";"SST";p33;p66;36;63;1;10;82;8;40;58;2;1
```

Donde se ha indicado que el código de la estación es: “4801501”, el nombre de la estación es :”4801501\_ppt”, su latitud y longitud son: “4.20” y”-84.06, la ubicación del archivo que contiene la serie de la variable dependiente es: “../4801501\_ppt.txt”, la variable asociada a la variable dependiente es: "PPT" (precipitación), la ubicación del archivo que guarda la serie de la variable independiente es: “../TSM34.txt”, la clase de la variable independiente es: "SST" (temperatura superficial del mar), los umbrales para dar categorías por encima y por debajo de lo normal de la variable independiente son: “p33” y “p66”, los valores de pronóstico de la variable independiente para el rezago 0 en el trimestre EFM son: por debajo de lo normal “36”, dentro de lo normal “63” por encima de lo normal “1”, respectivamente para el rezago 1 son: “10”, “82” y “8”, mientras que para el rezago 2 son: “40”, “58” y “2”. Finalmente el trimestre para el cual se realiza el pronóstico es: “1”.

Dado que se ha decidido correr los módulos pronóstico y clima, es de esperarse obtener las carpetas

1. Jazikü-climate
  - 1.1 4801501\_4801501\_ppt,maps
    - 1.1.1 Contingency\_table,graphics, time\_series y result\_table\_4801501\_4801501\_ppt\_PPT\_SST\_(1985-2009) (tabla de resultados)
- 2.Jazikü-forecasting

<sup>1</sup>No olvide incluir las direcciones absolutas en las que han sido guardados las series de la variables dependiente e independientes

```
2.1. 4801501_4801501_ppt,maps
2.2 Mosaic_prob_var_D_trim_1_(jfm)_(1985-2009),
prob_var_D_lag_0_trim_1_(jfm)_(1985-2009),
prob_var_D_lag_1_trim_1_(jfm)_(1985-2009),
prob_var_D_lag_2_trim_1_(jfm)_(1985-2009)
```

En las subcarpetas 1.1.1, es posible encontrar las tablas de contingencia por rezagos, las gráficas de las tablas de contingencia, las series de tiempo de los trimestres utilizados para aplicar el análisis compuesto y la tabla de resultados asociada a la estación (ver sección 3.3). Compruebe que la tabla de resultados se corresponde con la mostrada en la Fig.8.5 Consecuentemente el pronóstico asociado al trimestre 1 como el marcado en el archivo stations.csv, debe corresponder con el mostrado en la Fig.8.6.

Año	Ppt(mm)	Año	Ppt(mm)	Año	Ppt(mm)	Año	Ppt(mm)	Año	Ppt(mm)
1985-01	86.4	1986-01	456.9	1987-01	489.9	1988-01	553.8	1989-01	405.8
1985-02	232.9	1986-02	332.2	1987-02	293.9	1988-02	340.1	1989-02	339.3
1985-03	316.4	1986-03	349.1	1987-03	343.9	1988-03	161.1	1989-03	397.1
1985-04	306.2	1986-04	301	1987-04	441.1	1988-04	328.9	1989-04	354.9
1985-05	333.1	1986-05	268.1	1987-05	256.6	1988-05	439	1989-05	233.1
1985-06	159.3	1986-06	241.7	1987-06	151	1988-06	169.7	1989-06	330.2
1985-07	261.8	1986-07	187.2	1987-07	181.9	1988-07	67.7	1989-07	137.5
1985-08	305.7	1986-08	312.3	1987-08	78.6	1988-08	70.2	1989-08	266.4
1985-09	319.7	1986-09	230.7	1987-09	107.6	1988-09	162.4	1989-09	210.4
1985-10	147.5	1986-10	289.6	1987-10	479.4	1988-10	263.5	1989-10	381.9
1985-11	156.4	1986-11	475.5	1987-11	359.7	1988-11	385.5	1989-11	322.1
1985-12	301.9	1986-12	327.5	1987-12	316.8	1988-12	305.3	1989-12	187.5
1990-01	328.4	1991-01	282.9	1992-01	450.1	1993-01	351.6	1994-01	406.2
1990-02	327.7	1991-02	449.2	1992-02	282.1	1993-02	304.4	1994-02	517.3
1990-03	412.6	1991-03	365.6	1992-03	448.9	1993-03	324.2	1994-03	360.8
1990-04	376	1991-04	525.2	1992-04	219.9	1993-04	457.4	1994-04	291.1
1990-05	353	1991-05	289.7	1992-05	220	1993-05	324.1	1994-05	316.5
1990-06	187	1991-06	245.7	1992-06	265.2	1993-06	264.6	1994-06	324
1990-07	179.2	1991-07	89.3	1992-07	93.2	1993-07	133	1994-07	126.9
1990-08	257.1	1991-08	109.5	1992-08	171.6	1993-08	244.5	1994-08	181.4
1990-09	179.3	1991-09	285	1992-09	233.9	1993-09	350.9	1994-09	360.7
1990-10	225.2	1991-10	164.4	1992-10	89.4	1993-10	310.7	1994-10	150.8
1990-11	376.3	1991-11	290.2	1992-11	370.5	1993-11	421.7	1994-11	354.8
1990-12	451	1991-12	275.8	1992-12	442.4	1993-12	370.3	1994-12	272.4
1995-01	350.1	1996-01	311.2	1997-01	226.3	1998-01	370.4	1999-01	384.3
1995-02	345.4	1996-02	361.4	1997-02	427	1998-02	195.1	1999-02	252.9
1995-03	519.1	1996-03	384.6	1997-03	398.6	1998-03	327.9	1999-03	216.6
1995-04	301.6	1996-04	227.7	1997-04	303.2	1998-04	446	1999-04	284.8
1995-05	159	1996-05	302.9	1997-05	372.9	1998-05	351.1	1999-05	518.6
1995-06	169.2	1996-06	221	1997-06	107.1	1998-06	127.4	1999-06	184.6
1995-07	105.7	1996-07	98.6	1997-07	95.2	1998-07	168.7	1999-07	234.1
1995-08	124.9	1996-08	237.7	1997-08	93.2	1998-08	121.5	1999-08	182.3
1995-09	199.3	1996-09	216.9	1997-09	315.6	1998-09	146.5	1999-09	169.8
1995-10	242.7	1996-10	251	1997-10	99.9	1998-10	312.3	1999-10	213.3
1995-11	366.5	1996-11	152.1	1997-11	193.4	1998-11	231.5	1999-11	292
1995-12	287.6	1996-12	214	1997-12	429.9	1998-12	101	1999-12	260.7
2000-01	155.2	2001-01	558.5	2002-01	414.7	2003-01	298.8	2004-01	274.9
2000-02	376.9	2001-02	409.2	2002-02	398.9	2003-02	531.3	2004-02	216.5
2000-03	221.6	2001-03	326.5	2002-03	269.2	2003-03	299.5	2004-03	488.2
2000-04	373.9	2001-04	362.8	2002-04	442.3	2003-04	312.2	2004-04	391.6
2000-05	542.4	2001-05	409.4	2002-05	229.9	2003-05	245.7	2004-05	183.3
2000-06	198.6	2001-06	214.1	2002-06	190.4	2003-06	220	2004-06	291.4
2000-07	161.5	2001-07	270.3	2002-07	331.3	2003-07	67.8	2004-07	121.5
2000-08	149.3	2001-08	38.9	2002-08	154	2003-08	212.4	2004-08	155.7
2000-09	387.6	2001-09	141.4	2002-09	114.5	2003-09	112.9	2004-09	297.9
2000-10	227.1	2001-10	139.4	2002-10	316.3	2003-10	177.6	2004-10	168.4
2000-11	375.4	2001-11	124.1	2002-11	408.1	2003-11	362.1	2004-11	405
2000-12	297	2001-12	265.8	2002-12	347.1	2003-12	99999	2004-12	310.7
2005-01	312.2	2006-01	605.9	2007-01	400.8	2008-01	344.2	2009-01	594
2005-02	208.8	2006-02	350.1	2007-02	291.7	2008-02	372.5	2009-02	324.2
2005-03	319.7	2006-03	536.8	2007-03	405.1	2008-03	254.9	2009-03	522.9
2005-04	466	2006-04	308.8	2007-04	383.2	2008-04	282.3	2009-04	375.2
2005-05	210.3	2006-05	255.9	2007-05	244.4	2008-05	506.5	2009-05	235.1
2005-06	57.7	2006-06	108.5	2007-06	173.5	2008-06	213.1	2009-06	197
2005-07	52.5	2006-07	187.1	2007-07	87.6	2008-07	176.2	2009-07	188.8
2005-08	99.8	2006-08	150.6	2007-08	50.4	2008-08	149	2009-08	99999
2005-09	76.9	2006-09	269.5	2007-09	295	2008-09	274.9	2009-09	99999
2005-10	574.9	2006-10	166.3	2007-10	429.2	2008-10	283.7	2009-10	99999
2005-11	260.4	2006-11	176.6	2007-11	546.1	2008-11	237	2009-11	99999
2005-12	296.1	2006-12	476.6	2007-12	347.9	2008-12	283.5	2009-12	99999

Figura 8.3: Series de ejemplo de variable dependiente, precipitacion estación 4801501 en el período 1986-2009

CAPÍTULO 8. APÉNDICE 1: POSIBLES ERRORES ARROJADOS POR JAZIKÜ

53

Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)	Año	TSM(°C)
1950-01	25.01	1951-01	25.47	1952-01	26.88	1953-01	27.02	1954-01	27.08	1980-01	27.13	1981-01	26.11	1982-01	26.72	1983-01	29.36	1984-01	25.64		
1950-02	24.92	1951-02	26.62	1952-02	27.03	1953-02	27.09	1954-02	26.95	1980-02	27.18	1981-02	26.16	1982-02	26.7	1983-02	29.13	1984-02	26.39		
1950-03	26.41	1951-03	26.44	1952-03	26.89	1953-03	27.41	1954-03	27.36	1980-03	27.28	1981-03	26.92	1982-03	27.2	1983-03	29.03	1984-03	26.86		
1950-04	26.75	1951-04	27.74	1952-04	28.12	1953-04	28.4	1954-04	27.38	1980-04	27.86	1981-04	27.28	1982-04	28.02	1983-04	28.91	1984-04	27.39		
1950-05	26.3	1951-05	27.54	1952-05	27.58	1953-05	27.9	1954-05	27.46	1980-05	27.96	1981-05	27.46	1982-05	28.54	1983-05	28.89	1984-05	27.39		
1950-06	26.79	1951-06	27.53	1952-06	27.06	1953-06	27.87	1954-06	27.06	1980-06	28.12	1981-06	27.41	1982-06	28.75	1983-06	28.24	1984-06	26.86		
1950-07	26.59	1951-07	27.48	1952-07	26.82	1953-07	27.65	1954-07	26.32	1980-07	27.32	1981-07	26.65	1982-07	28.1	1983-07	27.07	1984-07	26.74		
1950-08	26.11	1951-08	27.44	1952-08	26.83	1953-08	26.82	1954-08	25.65	1980-08	26.59	1981-08	26.01	1982-08	27.93	1983-08	26.53	1984-08	26.34		
1950-09	25.56	1951-09	27.02	1952-09	26.49	1953-09	27.46	1954-09	25.55	1980-09	26.6	1981-09	26.57	1982-09	28.11	1983-09	26.44	1984-09	26.43		
1950-10	26.03	1951-10	27.4	1952-10	26.7	1953-10	26.91	1954-10	25.8	1980-10	26.51	1981-10	26.8	1982-10	28.64	1983-10	25.87	1984-10	25.93		
1950-11	25.42	1951-11	27.31	1952-11	26.37	1953-11	26.94	1954-11	25.65	1980-11	26.6	1981-11	26.41	1982-11	28.81	1983-11	25.58	1984-11	25.41		
1950-12	25.54	1951-12	27.19	1952-12	25.99	1953-12	26.43	1954-12	25.53	1980-12	26.91	1981-12	26.58	1982-12	29.21	1983-12	25.59	1984-12	25		
1955-01	25.12	1956-01	25.32	1957-01	26.16	1958-01	28.42	1959-01	27.06	1985-01	25.43	1986-01	25.79	1987-01	27.32	1988-01	27.32	1989-01	24.53		
1955-02	26.27	1956-02	26.18	1957-02	26.88	1958-02	28.24	1959-02	27.3	1985-02	25.67	1986-02	25.94	1987-02	28.02	1988-02	27.22	1989-02	25.33		
1955-03	26.31	1956-03	26.37	1957-03	27.37	1958-03	28.36	1959-03	27.24	1985-03	26.23	1986-03	26.65	1987-03	28.47	1988-03	27.31	1989-03	25.9		
1955-04	27.01	1956-04	26.91	1957-04	28.16	1958-04	28.21	1959-04	27.96	1985-04	26.8	1986-04	27.44	1987-04	28.8	1988-04	27.32	1989-04	26.69		
1955-05	26.66	1956-05	27.63	1957-05	28.13	1958-05	28.11	1959-05	27.92	1985-05	27.11	1986-05	27.5	1987-05	28.75	1988-05	26.48	1989-05	27.09		
1955-06	26.72	1956-06	27.16	1957-06	27.91	1958-06	28.1	1959-06	27.37	1985-06	26.86	1986-06	27.69	1987-06	29.03	1988-06	26.11	1989-06	26.98		
1955-07	26.39	1956-07	26.32	1957-07	27.82	1958-07	27.33	1959-07	26.62	1985-07	26.69	1986-07	27.37	1987-07	28.8	1988-07	25.57	1989-07	26.74		
1955-08	25.92	1956-08	25.98	1957-08	27.97	1958-08	27.22	1959-08	26.33	1985-08	26.5	1986-08	27.15	1987-08	28.58	1988-08	25.24	1989-08	26.33		
1955-09	25.01	1956-09	25.79	1957-09	27	1958-09	26.13	1959-09	26.04	1985-09	26.25	1986-09	27.33	1987-09	28.39	1988-09	25.43	1989-09	26.25		
1955-10	24.83	1956-10	26.3	1957-10	27.39	1958-10	26.7	1959-10	26.52	1985-10	26.19	1986-10	27.57	1987-10	28.07	1988-10	24.62	1989-10	26.26		
1955-11	24.34	1956-11	25.63	1957-11	27.93	1958-11	26.71	1959-11	26.4	1985-11	26.19	1986-11	27.73	1987-11	27.99	1988-11	24.27	1989-11	26.24		
1955-12	25.05	1956-12	26.2	1957-12	28.08	1958-12	27.02	1959-12	26.39	1985-12	26.11	1986-12	27.7	1987-12	27.6	1988-12	24.33	1989-12	26.38		
1960-01	26.49	1961-01	26.23	1962-01	26.17	1963-01	25.88	1964-01	27.42	1990-01	26.55	1991-01	27.01	1992-01	28.41	1993-01	26.69	1994-01	26.6		
1960-02	26.35	1961-02	26.61	1962-02	26.38	1963-02	26.19	1964-02	27.14	1990-02	26.95	1991-02	26.93	1992-02	28.63	1993-02	26.97	1994-02	26.59		
1960-03	27.17	1961-03	27.03	1962-03	26.87	1963-03	27.14	1964-03	26.96	1990-03	27.46	1991-03	27.25	1992-03	28.83	1993-03	27.66	1994-03	27.27		
1960-04	27.84	1961-04	27.76	1962-04	27.28	1963-04	27.83	1964-04	27.07	1990-04	28.02	1991-04	27.98	1992-04	29.14	1993-04	28.59	1994-04	27.9		
1960-05	27.65	1961-05	27.84	1962-05	27.24	1963-05	27.55	1964-05	26.82	1990-05	28.06	1991-05	28.35	1992-05	28.99	1993-05	28.82	1994-05	28.04		
1960-06	27.35	1961-06	28.05	1962-06	27.41	1963-06	27.61	1964-06	26.58	1990-06	27.58	1991-06	28.36	1992-06	28.02	1993-06	28.28	1994-06	27.99		
1960-07	27.07	1961-07	26.9	1962-07	27.05	1963-07	28.01	1964-07	26.53	1990-07	27.25	1991-07	27.92	1992-07	27.53	1993-07	27.55	1994-07	27.35		
1960-08	26.9	1961-08	26.45	1962-08	26.81	1963-08	27.62	1964-08	26	1990-08	27.05	1991-08	27.44	1992-08	26.64	1993-08	26.84	1994-08	27.35		
1960-09	26.63	1961-09	25.93	1962-09	26.1	1963-09	27.42	1964-09	25.52	1990-09	26.75	1991-09	27.07	1992-09	26.48	1993-09	26.92	1994-09	27		
1960-10	26.28	1961-10	26.03	1962-10	26.17	1963-10	27.59	1964-10	25.83	1990-10	26.98	1991-10	27.63	1992-10	26.34	1993-10	26.93	1994-10	27.49		
1960-11	26.12	1961-11	26.16	1962-11	25.97	1963-11	27.45	1964-11	25.32	1990-11	26.72	1991-11	27.86	1992-11	26.51	1993-11	26.91	1994-11	27.87		
1960-12	26.31	1961-12	26.28	1962-12	25.83	1963-12	27.58	1964-12	25.38	1990-12	26.91	1991-12	28.37	1992-12	26.73	1993-12	26.76	1994-12	27.87		
1965-01	25.82	1966-01	27.82	1967-01	26.15	1968-01	25.88	1969-01	27.69	1995-01	27.55	1996-01	25.74	1997-01	25.96	1998-01	29.1	1999-01	24.9		
1965-02	26.41	1966-02	27.67	1967-02	26.42	1968-02	25.77	1969-02	27.81	1995-02	27.45	1996-02	25.85	1997-02	26.36	1998-02	28.86	1999-02	25.41		
1965-03	27.15	1966-03	28.03	1967-03	26.58	1968-03	26.55	1969-03	27.99	1995-03	27.63	1996-03	26.62	1997-03	27.03	1998-03	26.67	1999-03	26.25		
1965-04	27.75	1966-04	28.39	1967-04	26.87	1968-04	27.28	1969-04	28.17	1995-04	27.93	1996-04	27.36	1997-04	28.03	1998-04	28.56	1999-04	26.84		
1965-05	28.09	1966-05	27.66	1967-05	27.5	1968-05	27.2	1969-05	28.57	1995-05	27.73	1996-05	27.37	1997-05	28.6	1998-05	28.47	1999-05	26.97		
1965-06	28.27	1966-06	28.03	1967-06	27.74	1968-06	27.6	1969-06	28.09	1995-06	27.59	1996-06	27.32	1997-06	28.94	1998-06	26.72	1999-06	26.6		
1965-07	28.23	1966-07	27.45	1967-07	26.93	1968-07	27.47	1969-07	27.28	1995-07	27.01	1996-07	27.09	1997-07	28.92	1998-07	25.94	1999-07	26.35		
1965-08	28.05	1966-08	26.79	1967-08	26.48	1968-08	27.14	1969-08	27.33	1995-08	26.33	1996-08	26.56	1997-08	28.84	1998-08	25.49	1999-08	25.59		
1965-09	27.93	1966-09	26.51	1967-09	26.04	1968-09	26.69	1969-09	27.27	1995-09	25.96	1996-09	26.35	1997-09	28.93	1998-09	25.61	1999-09	25.71		
1965-10	28.16	1966-10	26.51	1967-10	26.09	1968-10	26.92	1969-10	27.44	1995-10	25.67	1996-10	26.24	1997-10	29.23	1998-10	25.34	1999-10	25.64		
1965-11	28.13	1966-11	26.29	1967-11	26.19	1968-11	27.36	1969-11	27.24	1995-11	25.66	1996-11	26.19	1997-11	29.32	1998-11	25.18	1999-11	25.12		
1965-12	28.31	1966-12	26.17	1967-12	26.11	1968-12	27.27	1969-12	27.62	1995-12	25.57	1996-12	26.02	1997-12	29.26	1998-12	24.79	1999-12	24.9		
1970-01	27.39	1971-01	25.01	1972-01	25.84	1973-01	28.25	1974-01	24.58	2000-01	24.65	2001-01	25.74	2002-01	26.5	2003-01	27.76	2004-01	26.74		
1970-02	27.1	1971-02	25.27	1972-02	26.49	1973-02	27.98	1974-02	25.21	2000-02	25.19	2001-02	26.11	2002-02	26.95	2003-02	27.49	2004-02	26.86		
1970-03	27.42	1971-03	26.11	1972-03	27.08	1973-03	27.82	1974-03	26.05	2000-03	26.08	2001-03	26.84	2002-03	27.32	2003-03	27.81	2004-03	27.1		
1970-04	28.14	1971-04	26.64	1972-04	28.08	1973-04	27.53	1974-04	26.9	2000-04	27.01	2001-04	27.52	2002-04	27.94	2003-04	27.81	2004-04	27.84		
1970-05	27																				

Figura 8.5: Tabla de resultados obtenida después de usar el software Jazikü cuando se utilizó a manera de ejemplo como variable dependiente, precipitación estación 4801501 en el período 1986-2009 y como variable independiente la serie de TSM en la región El Niño 3.4 en el período 1950-2009.

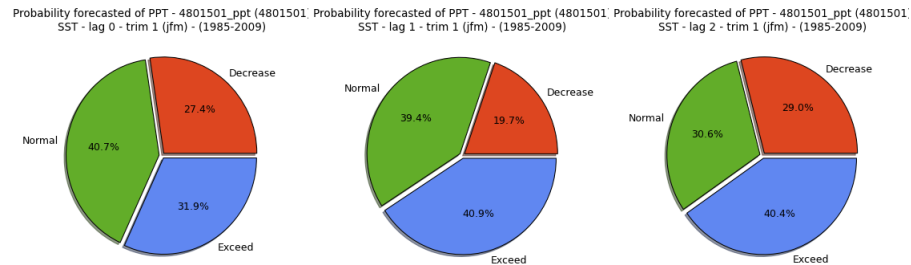


Figura 8.6: Pronóstico asociado para el comportamiento de la variable precipitación de la estación 4801501, para probabilidades de la variable independiente (Por debajo de lo normal “36”, dentro de lo normal “63” por encima de lo normal “1”, respectivamente para el rezago 1 “10”, “82” y “8”, mientras que para el rezago 2 “40”, “58” y “2”), Temperatura Superficial del Mar en la región El Niño 3.4, trimestre EFM.

# Bibliography

- [1] National Oceanic and Atmospheric Administration (NOAA) , University Corporation for Atmospheric Research (UCAR)). Creating a Local Climate Product Using Composite Analysis - Print Version of Webcast -(En Linea). 1997-2010:COMET Website at <http://meted.ucar.edu/>, 1997.
- [2] A. Leetmaa Barnston, A. G. NCEP Forecasts of the El Niño of 1997 1998 and Its U.S. Impacts. *Bull. Amer. Met. Soc*, 80:1829 – 1852, 1999.
- [3] B.D. Zumbo & R.H. Williams D.W. Zimmerman. Bias in Estimation and Hypothesis Testing of Correlation. *Psicológica*, 24:133–158, 2003.
- [4] Earth System Research Laboratory & National Oceanic and Atmospheric Administration. Dirección Nacional de Servicios Académicos Virtuales Seminario de Investigación. <http://www.esrl.noaa.gov/psd/enso/mei.ext/rank.ext.html> (En línea), 2011.
- [5] Eric J. Alfaro & F. Javier Soley. Uso de una Tabla de contingencia para Aplicaciones Climáticas. *Physical Oceanography Division, Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration*,, 2003.
- [6] Eric J. Alfaro & Jorge A. Amador. Algunas características de la estratósfera sobre Mesoamérica. *Top. Meteor. Oceanog*, 19(8):1 – 9, 2001.
- [7] D.H Freeman. (Citado por Ato & López (1996)) *Applied Categorical Data Analysis*. Marcel Dekker, NY-USA, 1987.
- [8] & Winter Glasser, G.J. (Citado por Zimmerman et al (2003)). Critical values of the coefficient of rank correlation for testing the hypothesis of independence. *Biometrika*, (48):444–448, 1961.
- [9] J.M. Guevara Díaz. El abc de los índices usados en la Identificación y definición cuantitativa de El niño - oscilación del sur (ENSO). *Terra*, XXIV(35):85 –140, 2008.
- [10] M. A. Lopez Sánchez I. Espejo Miranda, F.Fernandez Palacín. *Estadística descriptiva y probabilidad*. Universidad de Cadíz, ISBN: 978-84-9828-058-6, 3 edition, 2002.



- [11] Geir Ottersen & Martin Visbeck James W. Hurrell, Yochanan Kushnir. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, volume 134, chapter An Overview of the North Atlantic Oscillation, pages 1812 – 1835. Copyright 2003 by the American Geophysical Union, 2003.
- [12] J.F. Gravetter & L. B. Wallnau. *Statistics for the Behavioral Sciences*. ISBN: 0-495-602-205. Wadsworth Cengage Learning, 8 edition, 2007.
- [13] José Vicéns Otero & Eva Medina Moral. ANÁLISIS DE DATOS CUALITATIVOS. Technical report, Departamento de Economía Aplicada, Universidad Autónoma de Madrid, 2005.
- [14] Kendall Kendall, M.G. (Citado por Zimmerman et al (2003)). The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times. *Biometrika*, (30):251–273, 1939.
- [15] G.E León Aristizábal. CONTROL DE CALIDAD AUTOMÁTICO DE DATOS METEOROLÓGICOS: IDEAM - 2007. *Instituto de hidrología, meteorología y estudios ambientales, IDEAM - Subdirección de Meteorología*, page 51, 2007.
- [16] Luis Alberto García Guirola. Régimen de lluvia y caudales en El Salvador, su relación con la variabilidad climática (forzante oceánicos -atmosféricos) para la construcción de mapas de amenazas por déficit de lluvias. Estudios de posgrado en ciencias de la atmósfera, para optar al grado de magister scientiae en hidrología, Universidad de Costa Rica: Sistema de estudios de posgrado, Ciudad Universitaria Rodrigo Facio, Costa Rica., 2009.
- [17] M.A. Sánchez Ramos. Uso metodológico de las tablas de contingencia en la ciencia política. *Espacios públicos*, 8(16):60–84, 2005.
- [18] Manuel Ato García & Juan José López García. *Análisis Estadístico para datos categóricos*. Editorial Síntesis S.A., 1996.
- [19] meteorología y estudios ambientales Ministerio de medio ambiente, vivienda y desarrollo territorial & Instituto de hidrología. *Atlas Climatológico de Colombia*. IDEAM, 2005.
- [20] J.E Montealegre Bocanegra. Estudio de la variabilidad climática de la precipitación en Colombia asociada a procesos oceánicos y atmosféricos de meso y gran escala: Informe final de las actividades desarrolladas en el marco del contrato de prestación de servicios No IDEAM 022 - 2009. *Instituto de hidrología, meteorología y estudios ambientales, IDEAM - Subdirección de Meteorología*, page 54, 2009.
- [21] M. B. Richman Montroy, D.L. Observed Nonlinearities of Monthly Teleconnections between Tropical Pacific Sea Surface Temperature Anomalies

- and Central and Eastern North American Precipitation. *Journal of Climate*, 11:1812 – 1835, 1998.
- [22] M.R. Risk. *Cartas sobre Estadística de la Revista Argentina de Bioingeniería*. ISBN 987-43-6130-1. Universidad Tecnológica Nacional, 2003.
- [23] M. Signoret-Poillon y R. Garduño-López N. Sánchez-Santillán. La Oscilación del Atlántico Norte: un fenómeno que incide en la variabilidad climática de México. *INGENIERÍA Investigación y Tecnología*, 2:85–95, 2006.
- [24] Person K. On the criterion that a given system of deviation from the probable in the case of a correlated system of varibale is such that it can be reasonable, supposed that have arisen from random sampling. *Philosophical Magazine*, 50(5):157–175, 1900.
- [25] Alberto Luceño. Vasquez. *Metodos estadisticos para medir, describir y controlar la variabilidad*. Universidad de Cantabria, España , isbn:54810237502 edition, 2004.
- [26] K. Wolter. (Citado por Guevara Díaz (2008)). The Southern Oscillation in surface circulation and climate over the tropical Atlantic, Eastern Pacific, and Indian Oceans as captured by cluster analysis. *Journal of Applied Meteorology*, (26):540 – 558, 1987.