

**IMPLEMENTACIÓN DE UN MODELO DE AUTÓMATA CELULAR PARA EL
PRONÓSTICO DE LA PRECIPITACIÓN ESPACIAL. CASO DE ESTUDIO
CIUDAD DE BOGOTÁ (COLOMBIA).**

Proponente:
Ing. Ana Carolina Santos Rocha
Código: 295346

Director de proyecto:
Ing. MSc PhD Nelson Obregón Neira

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE BOGOTÁ
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA CIVIL Y AGRÍCOLA
MAESTRÍA EN RECURSOS HIDRÁULICOS
Bogotá, 2010

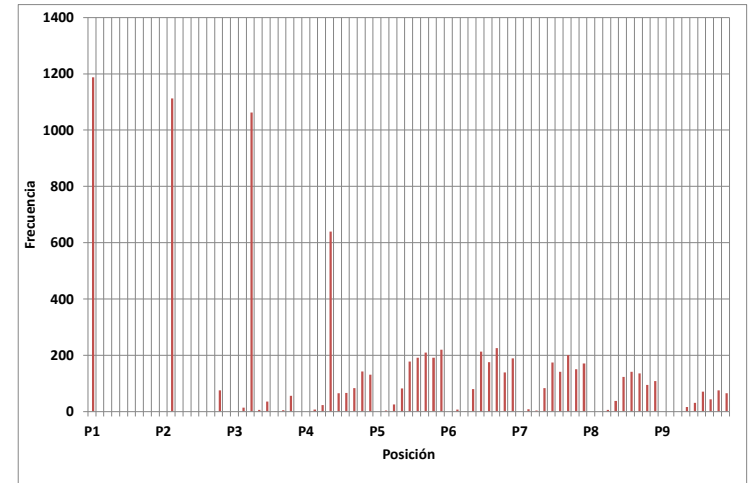
RESUMEN CLASIFICACIÓN

Clasificación tipo: 2 Modelo: clase2 Resolución temporal: diaria Año 1995

CLASIFICACIÓN		
Rango	Etiqueta	
0	0.30	11
0.30	1.00	12
1.00	10.00	13
10.00	20.00	14
20.00	30.00	15
30.00	40.00	16
40.00	50.00	17
50.00	60.00	18
60.00	70.00	19
70.00	80.00	20
80.00	90.00	21
90.00	100	22

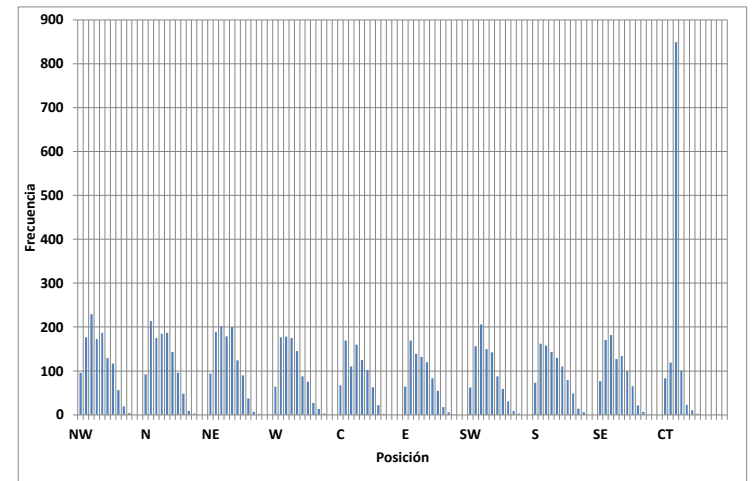
FRECUENCIA POR POSICIÓN Y ETIQUETA						
Posición	Frecuencia Vecino ¹	Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³
			Max	Mín		
P1	1	NW	13	20	11	1
P2	2	N	12	20	12	2
P3	3	NE	13	20	13	1
P4	4	W	13	20	14	2
P5	9	C	12	19	15	3
P6	7	E	12	20	16	2
P7	7	SW	13	20	17	1
P8	6	S	12	20	18	1
P9	8	SE	13	20	19	1
		Ct	13	17	20	8
					21	#N/A
					22	#N/A

ESTADÍSTICA DE REGLAS	
Número de patrones	9176
Porcentaje de reducción de información	29
Total de reglas	6500
Número de reglas usadas	1187
Porcentaje de reglas usadas	18
Número de reglas sin usar	5313
Porcentaje de reglas sin usar	82



- ¹ Corresponde al vecino que con mayor frecuencia le pregunta para tomar la decisión.
- ² Corresponde al valor de etiqueta más frecuente en el vecino
- ³ Corresponde al vecino donde con mayor frecuencia se presenta el valor de etiqueta

REGLAS MAS FRECUENTES																			
*Solo se muestran las primeras 40																			
Ranking	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8	V8	P9	OUT	FRECUENCIA
1	1	13	2	13	3	13	4	13	7	13	9	13	8	13	0	0	0	13	631418
2	1	12	2	12	3	12	4	12	9	12	7	12	5	12	6	12	0	13	202248
3	1	11	2	11	3	11	9	11	7	11	5	11	4	11	6	11	8	11	183681
4	1	14	2	14	3	14	4	14	9	14	6	14	7	14	0	0	0	13	72924
5	1	12	2	12	3	12	4	12	9	11	7	12	5	12	0	0	0	13	4244
6	1	13	2	13	3	12	4	13	7	13	8	13	5	13	9	13	0	13	3343
7	1	15	2	15	3	15	4	15	5	15	6	15	7	15	8	15	0	13	3285
8	1	12	2	12	3	11	4	12	8	12	5	12	9	12	0	0	0	13	3125
9	1	13	2	13	3	13	4	13	7	12	5	13	9	13	6	13	0	13	3104
10	1	11	2	11	3	11	9	12	4	11	8	12	0	0	0	0	0	13	2977
11	1	13	2	13	3	13	4	13	7	13	9	12	8	13	5	13	0	13	2955
12	1	12	2	12	3	12	4	12	9	12	7	11	6	12	5	12	0	13	2950
13	1	11	2	12	3	12	8	12	5	12	7	12	9	12	6	12	0	13	2912
14	1	12	2	11	3	11	6	11	8	11	5	11	0	0	0	0	0	13	2670
15	1	12	2	13	3	13	9	13	6	13	7	13	4	13	5	13	8	13	2636
16	1	12	2	13	3	13	9	13	6	13	7	12	0	0	0	0	0	13	2434
17	1	11	2	11	3	11	9	11	7	12	6	11	5	11	0	0	0	13	2360
18	1	12	2	12	3	12	4	12	9	13	7	13	5	12	0	0	0	13	2269
19	1	12	2	12	3	12	4	12	9	13	7	12	6	12	0	0	0	13	2244
20	1	11	2	11	3	12	7	12	8	12	6	12	0	0	0	0	0	13	2142
21	1	13	2	12	3	12	4	13	9	13	7	13	8	13	0	0	0	13	2115
22	1	13	2	13	3	13	4	13	7	13	9	14	0	0	0	0	0	13	2100
23	1	12	2	12	3	12	4	12	9	12	7	13	6	12	0	0	0	13	2066
24	1	12	2	12	3	12	4	11	8	11	9	11	0	0	0	0	0	13	2022
25	1	12	2	12	3	12	4	12	9	11	7	11	0	0	0	0	0	13	1928
26	1	12	2	12	3	12	4	13	8	13	6	12	7	13	0	0	0	13	1890
27	1	11	2	11	3	12	7	11	9	12	5	11	6	12	0	0	0	13	1871
28	1	12	2	12	3	13	4	12	9	12	5	12	8	12	0	0	0	13	1814
29	1	12	2	12	3	12	4	13	8	13	6	13	9	13	7	13	0	13	1714
30	1	13	2	13	3	13	4	13	7	12	5	13	9	12	0	0	0	13	1695
31	1	11	2	11	3	11	9	12	4	12	8	12	5	12	0	0	0	13	1660
32	1	11	2	11	3	12	7	11	9	11	4	11	8	11	0	0	0	13	1596
33	1	13	2	13	3	13	4	13	7	13	9	12	8	12	5	13	0	13	1593
34	1	11	2	11	3	11	9	12	4	11	8	11	7	11	0	0	0	13	1549
35	1	13	2	13	3	14	4	13	8	13	5	13	9	13	7	13	0	13	1363
36	1	13	2	12	3	12	4	12	6	12	7	12	8	12	5	12	0	13	1337
37	1	12	2	12	3	13	4	12	9	13	6	13	8	12	5	12	0	13	1261
38	1	11	2	12	3	12	8	12	5	12	7	11	4	11	9	12	6	12	1247
39	1	12	2	12	3	12	4	12	9	13	7	12	6	13	0	0	0	13	1236
40	1	12	2	12	3	13	4	12	9	13	6	13	8	13	7	12	0	13	1212



RESUMEN CLASIFICACIÓN

Clasificación tipo: 2 Modelo: clase2 Resolución temporal: diaria Año 1996

CLASIFICACIÓN

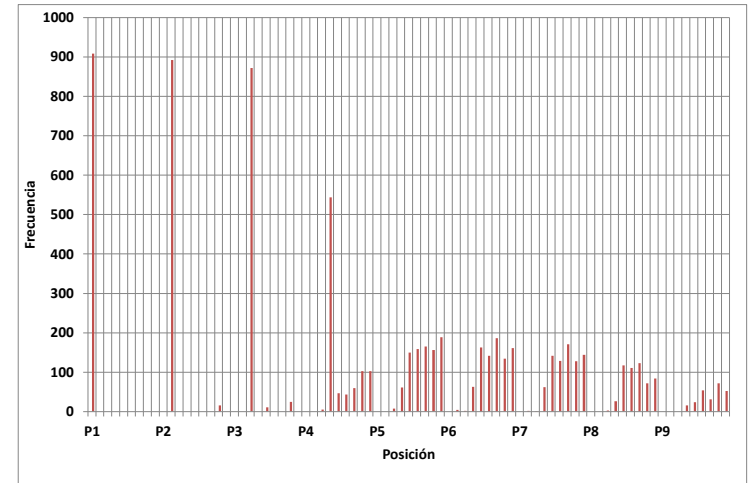
Rango	Etiqueta	
0	0.30	11
0.30	1.00	12
1.00	10.00	13
10.00	20.00	14
20.00	30.00	15
30.00	40.00	16
40.00	50.00	17
50.00	60.00	18
60.00	70.00	19
70.00	80.00	20
80.00	90.00	21
90.00	100	22

FRECUENCIA POR POSICIÓN Y ETIQUETA

Posición	Frecuencia Vecino ¹	Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³
			Max	Mín		
P1	1	NW	13	18	11	1
P2	2	N	12	18	12	2
P3	3	NE	13	18	13	1
P4	4	W	12	18	14	3
P5	9	C	12	18	15	1
P6	7	E	12	18	16	1
P7	7	SW	13	18	17	1
P8	7	S	12	18	18	1
P9	8	SE	13	18	19	#N/A
		Ct	13	15	20	#N/A
					21	#N/A
					22	#N/A

ESTADÍSTICA DE REGLAS

Número de patrones	9176
Porcentaje de reducción de información	29
Total de reglas	6500
Número de reglas usadas	908
Porcentaje de reglas usadas	14
Número de reglas sin usar	5592
Porcentaje de reglas sin usar	86

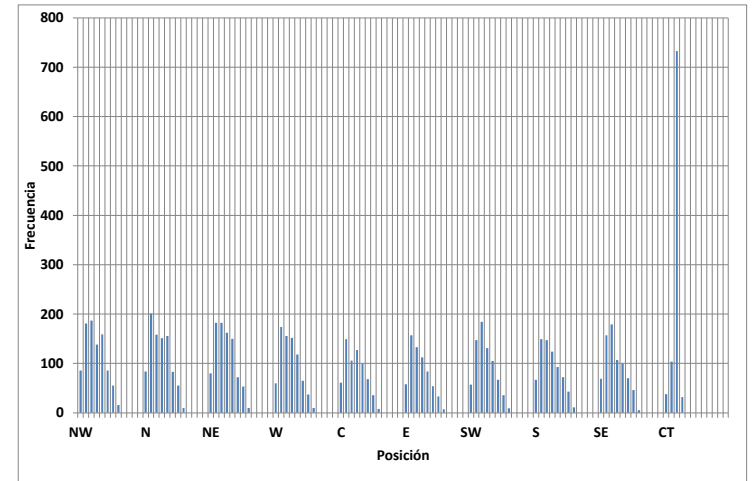


- ¹ Corresponde al vecino con mayor frecuencia le pregunta para tomar la decisión.
- ² Corresponde al valor de etiqueta más frecuente en el vecino
- ³ Corresponde al vecino donde con mayor frecuencia se presenta el valor de etiqueta

REGLAS MAS FRECUENTES

*Solo se muestran las primeras 40

Ranking	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8	V8	P9	OUT	FRECUENCIA
1	1	13	2	13	3	13	4	13	7	13	9	13	8	13	0	0	0	13	685420
2	1	12	2	12	3	12	4	12	9	12	7	12	5	12	6	12	0	13	196281
3	1	11	2	11	3	11	9	11	7	11	5	11	4	11	6	11	8	11	173950
4	1	14	2	14	3	14	4	14	9	14	6	14	7	14	0	0	0	13	33168
5	1	15	2	15	3	15	4	15	5	15	6	15	7	15	8	15	0	13	16672
6	1	16	2	16	3	16	8	16	4	16	9	16	7	16	0	0	0	13	5521
7	1	12	2	12	3	12	4	12	9	11	7	12	5	12	0	0	0	13	3728
8	1	13	2	13	3	13	4	13	7	12	5	13	9	13	6	13	0	13	3113
9	1	13	2	13	3	12	4	13	7	13	8	13	5	13	9	13	0	13	3101
10	1	13	2	13	3	13	4	13	7	13	9	12	8	13	5	13	0	13	2899
11	1	11	2	11	3	11	9	12	4	11	8	12	0	0	0	0	0	13	2808
12	1	12	2	12	3	11	4	12	8	12	5	12	9	12	0	0	0	13	2746
13	1	12	2	13	3	13	9	13	6	13	7	12	0	0	0	0	0	13	2673
14	1	11	2	12	3	12	8	12	5	12	7	12	9	12	6	12	0	13	2630
15	1	12	2	12	3	12	4	12	9	12	7	11	6	12	5	12	0	13	2534
16	1	12	2	11	3	11	6	11	8	11	5	11	0	0	0	0	0	13	2383
17	1	12	2	13	3	13	9	13	6	13	7	13	4	13	5	13	8	13	2372
18	1	11	2	11	3	11	9	11	7	12	6	11	5	11	0	0	0	13	2218
19	1	12	2	12	3	12	4	12	9	13	7	13	5	12	0	0	0	13	2158
20	1	12	2	12	3	13	4	12	9	12	5	12	8	12	0	0	0	13	2032
21	1	12	2	12	3	12	4	12	9	13	7	12	6	12	0	0	0	13	2020
22	1	11	2	11	3	12	7	12	8	12	6	12	0	0	0	0	0	13	1958
23	1	13	2	12	3	12	4	13	9	13	7	13	8	13	0	0	0	13	1863
24	1	12	2	12	3	12	4	12	9	12	7	13	6	12	0	0	0	13	1852
25	1	13	2	13	3	13	4	13	7	13	9	14	0	0	0	0	0	13	1761
26	1	12	2	12	3	12	4	12	9	11	7	11	0	0	0	0	0	13	1736
27	1	12	2	12	3	12	4	11	8	11	9	11	0	0	0	0	0	13	1683
28	1	13	2	13	3	13	4	13	7	12	5	13	9	12	0	0	0	13	1670
29	1	12	2	12	3	12	4	13	8	13	6	12	7	13	0	0	0	13	1657
30	1	12	2	12	3	12	4	13	8	13	6	13	9	13	7	13	0	13	1638
31	1	11	2	11	3	12	7	11	9	12	5	11	6	12	0	0	0	13	1636
32	1	13	2	13	3	13	4	13	7	13	9	12	8	12	5	13	0	13	1607
33	1	11	2	11	3	11	9	12	4	12	8	12	5	12	0	0	0	13	1510
34	1	11	2	11	3	11	9	12	4	11	8	11	7	11	0	0	0	13	1483
35	1	11	2	11	3	12	7	11	9	11	4	11	8	11	0	0	0	13	1463
36	1	13	2	12	3	12	4	12	6	12	7	12	8	12	5	12	0	13	1410
37	1	13	2	12	3	13	4	12	9	13	6	13	8	12	5	12	0	13	1345
38	1	13	2	13	3	12	4	13	7	13	8	13	5	13	9	12	6	13	1336
39	1	13	2	12	3	12	4	13	9	12	7	13	8	12	5	12	0	13	1289
40	1	13	2	13	3	13	4	12	6	12	5	12	8	12	9	12	0	13	1149



RESUMEN CLASIFICACIÓN

Clasificación tipo: 2 Modelo: clase2 Resolución temporal: diaria Año 1997

CLASIFICACIÓN

Rango	Etiqueta	
0	0.30	11
0.30	1.00	12
1.00	10.00	13
10.00	20.00	14
20.00	30.00	15
30.00	40.00	16
40.00	50.00	17
50.00	60.00	18
60.00	70.00	19
70.00	80.00	20
80.00	90.00	21
90.00	100	22

FRECUENCIA POR POSICIÓN Y ETIQUETA

Posición	Frecuencia Vecino ¹	Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³
			Max	Mín		
P1	1	NW	12	18	11	1
P2	2	N	12	18	12	2
P3	3	NE	12	17	13	1
P4	4	W	12	18	14	3
P5	9	C	12	18	15	1
P6	7	E	12	18	16	2
P7	7	SW	13	18	17	1
P8	7	S	12	18	18	1
P9	8	SE	12	18	19	#N/A
		Ct	13	14	20	#N/A
					21	#N/A
					22	#N/A

ESTADÍSTICA DE REGLAS

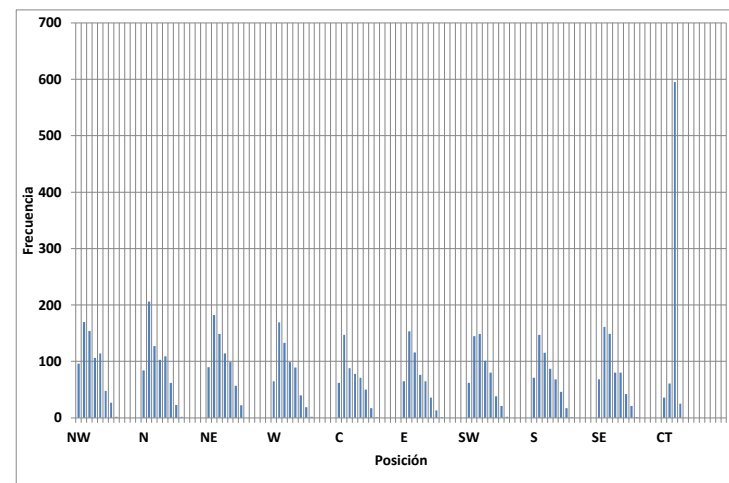
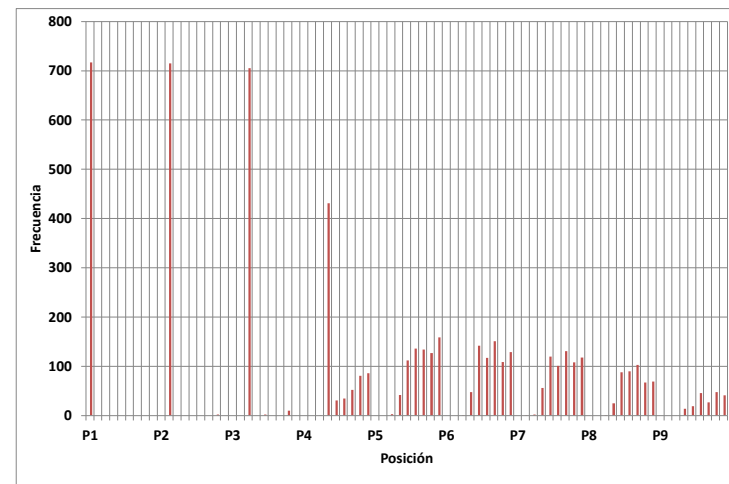
Número de patrones	9176
Porcentaje de reducción de información	29
Total de reglas	6500
Número de reglas usadas	717
Porcentaje de reglas usadas	11
Número de reglas sin usar	5783
Porcentaje de reglas sin usar	89

- ¹ Corresponde al vecino con mayor frecuencia le pregunta para tomar la decisión.
² Corresponde al valor de etiqueta más frecuente en el vecino
³ Corresponde al vecino donde con mayor frecuencia se presenta el valor de etiqueta

REGLAS MAS FRECUENTES

*Solo se muestran las primeras 40

Ranking	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8	V8	P9	V9	OUT	FRECUENCIA
1	1	13	2	13	3	13	4	13	7	13	9	13	8	13	0	0	0	0	13	501391
2	1	11	2	11	3	11	9	11	7	11	5	11	4	11	6	11	8	11	12	312921
3	1	12	2	12	3	12	4	12	9	12	7	12	5	12	6	12	0	0	13	274215
4	1	14	2	14	3	14	4	14	9	14	6	14	7	14	0	0	0	0	13	22976
5	1	11	2	11	3	11	9	12	4	11	8	12	0	0	0	0	0	0	13	4561
6	1	12	2	12	3	12	4	12	9	11	7	12	5	12	0	0	0	0	13	4237
7	1	11	2	12	3	12	8	12	5	12	7	12	9	12	6	12	0	0	13	3433
8	1	12	2	11	3	11	6	11	8	11	5	11	0	0	0	0	0	0	13	3216
9	1	11	2	11	3	11	9	11	7	12	6	11	5	11	0	0	0	0	13	3163
10	1	12	2	12	3	11	4	12	8	12	5	12	9	12	0	0	0	0	13	3159
11	1	12	2	12	3	12	4	12	9	12	7	11	6	12	5	12	0	0	13	3055
12	1	13	2	13	3	13	4	13	7	12	5	13	9	13	6	13	0	0	13	2655
13	1	13	2	13	3	12	4	13	7	13	8	13	5	13	9	13	0	0	13	2596
14	1	11	2	11	3	11	9	12	4	11	8	11	7	11	0	0	0	0	13	2533
15	1	11	2	11	3	12	7	12	8	12	6	12	0	0	0	0	0	0	13	2449
16	1	12	2	12	3	12	4	12	9	13	7	13	5	12	0	0	0	0	13	2361
17	1	12	2	12	3	12	4	11	8	11	9	11	0	0	0	0	0	0	13	2300
18	1	11	2	11	3	11	9	12	4	12	8	12	5	12	0	0	0	0	13	2285
19	1	11	2	11	3	12	7	11	9	12	5	11	6	12	0	0	0	0	13	2249
20	1	12	2	12	3	12	4	12	9	11	7	11	0	0	0	0	0	0	13	2247
21	1	12	2	12	3	12	4	12	9	13	7	12	6	12	0	0	0	0	13	2156
22	1	11	2	11	3	12	7	11	9	11	4	11	8	11	0	0	0	0	13	2148
23	1	12	2	13	3	13	9	13	6	13	7	13	4	13	5	13	8	13	13	2095
24	1	12	2	12	3	12	4	12	9	12	7	13	6	12	0	0	0	0	13	2086
25	1	13	2	13	3	13	4	13	7	13	9	12	8	13	5	13	0	0	13	2085
26	1	12	2	13	3	13	9	13	6	13	7	12	0	0	0	0	0	0	13	1880
27	1	12	2	12	3	12	4	13	8	13	6	13	9	13	7	13	0	0	13	1835
28	1	12	2	12	3	13	4	12	9	12	5	12	8	12	0	0	0	0	13	1834
29	1	12	2	12	3	12	4	13	8	13	6	13	7	13	0	0	0	0	13	1804
30	1	13	2	13	3	12	4	13	9	13	7	13	8	13	0	0	0	0	13	1658
31	1	13	2	13	3	13	4	13	7	12	5	13	9	12	0	0	0	0	13	1595
32	1	11	2	12	3	12	8	12	5	12	7	11	4	11	9	12	6	12	13	1517
33	1	15	2	15	3	15	4	15	5	15	6	15	7	15	8	15	0	0	13	1517
34	1	12	2	12	3	11	4	12	8	12	5	12	9	11	0	0	0	0	13	1329
35	1	13	2	13	3	13	4	13	7	13	9	14	0	0	0	0	0	0	13	1285
36	1	13	2	13	3	13	4	12	6	12	5	12	8	12	9	12	0	0	13	1140
37	1	13	2	13	3	13	9	12	4	12	8	12	5	12	0	0	0	0	13	1125
38	1	13	2	13	3	13	4	13	7	13	9	12	8	12	5	13	0	0	13	1120
39	1	13	2	13	3	13	4	12	6	13	8	12	9	12	0	0	0	0	13	1085
40	1	12	2	12	3	12	4	12	9	13	7	12	6	13	0	0	0	0	13	1083



Clasificación tipo: 2

Modelo: clase2

Resolución temporal: diaria

Año 1998

RESUMEN CLASIFICACIÓN

CLASIFICACIÓN

Rango	Etiqueta	
0	0.30	11
0.30	1.00	12
1.00	10.00	13
10.00	20.00	14
20.00	30.00	15
30.00	40.00	16
40.00	50.00	17
50.00	60.00	18
60.00	70.00	19
70.00	80.00	20
80.00	90.00	21
90.00	100	22

FRECUENCIA POR POSICIÓN Y ETIQUETA

Posición	Frecuencia Vecino ¹	Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³
			Max	Mín		
P1	1	NW	13	18	11	1
P2	2	N	12	18	12	2
P3	3	NE	12	18	13	1
P4	4	W	12	18	14	3
P5	9	C	12	18	15	1
P6	7	E	12	18	16	2
P7	7	SW	12	18	17	2
P8	6	S	12	18	18	1
P9	8	SE	13	18	19	#N/A
		Ct	13	16	20	#N/A
					21	#N/A
					22	#N/A

ESTADÍSTICA DE REGLAS

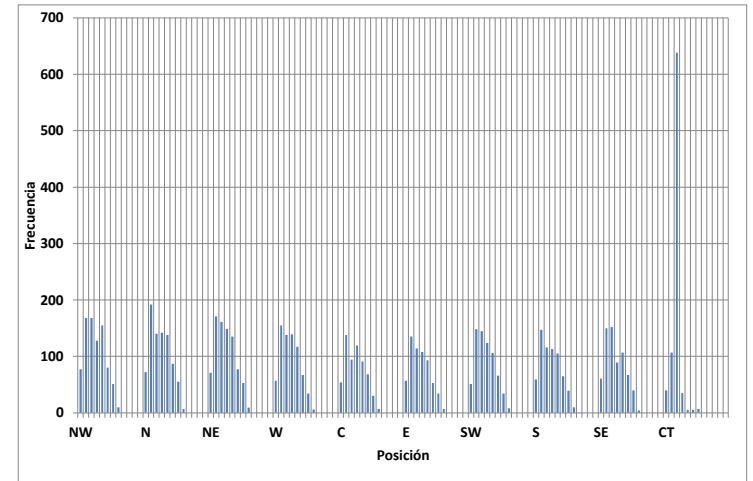
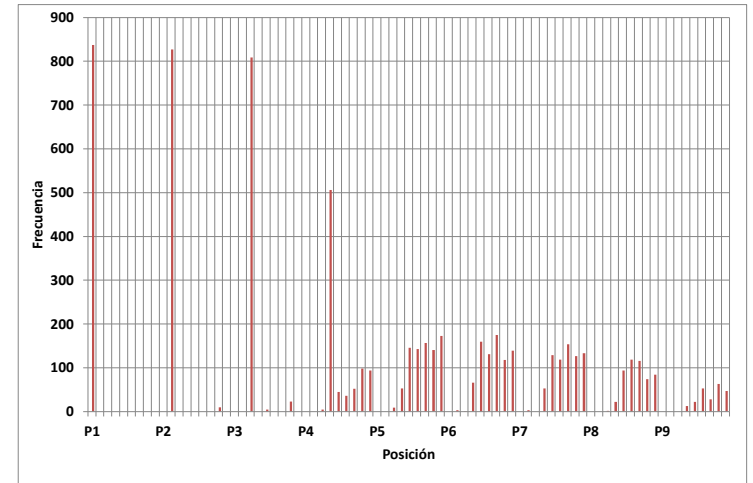
Número de patrones	9176
Porcentaje de reducción de información	29
Total de reglas	6500
Número de reglas usadas	837
Porcentaje de reglas usadas	13
Número de reglas sin usar	5663
Porcentaje de reglas sin usar	87

- ¹ Corresponde al vecino que con mayor frecuencia le pregunta para tomar la decisión.
- ² Corresponde al valor de etiqueta más frecuente en el vecino
- ³ Corresponde al vecino donde con mayor frecuencia se presenta el valor de etiqueta

REGLAS MAS FRECUENTES

*Solo se muestran las primeras 40

Ranking	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8	V8	P9	V9	OUT	FRECUENCIA
1	1	13	2	13	3	13	4	13	7	13	9	13	8	13	0	0	0	0	13	611895
2	1	12	2	12	3	12	4	12	9	12	7	12	5	12	6	12	0	0	13	221212
3	1	11	2	11	3	11	9	11	7	11	5	11	4	11	6	11	8	11	12	214234
4	1	14	2	14	3	14	4	14	9	14	6	14	7	14	0	0	0	0	13	46063
5	1	15	2	15	3	15	4	15	5	15	6	15	7	15	8	15	0	0	13	14297
6	1	16	2	16	3	16	8	16	4	16	9	16	7	16	0	0	0	0	13	5546
7	1	12	2	12	3	12	4	12	9	11	7	12	5	12	0	0	0	0	13	3585
8	1	11	2	11	3	11	9	12	4	11	8	12	0	0	0	0	0	0	13	3196
9	1	13	2	13	3	13	4	13	7	12	5	13	9	13	6	13	0	0	13	3142
10	1	13	2	13	3	12	4	13	7	13	8	13	5	13	9	13	0	0	13	2873
11	1	11	2	12	3	12	8	12	5	12	7	12	9	12	6	12	0	0	13	2641
12	1	12	2	12	3	11	4	12	8	12	5	12	9	12	0	0	0	0	13	2636
13	1	13	2	13	3	13	4	13	7	13	9	12	8	13	5	13	0	0	13	2627
14	1	12	2	12	3	12	4	12	9	12	7	11	6	12	5	12	0	0	13	2575
15	1	12	2	13	3	13	9	13	6	13	7	12	0	0	0	0	0	0	13	2381
16	1	11	2	11	3	11	9	11	7	12	6	11	5	11	0	0	0	0	13	2322
17	1	12	2	13	3	13	9	13	6	13	7	13	4	13	5	13	8	13	13	2297
18	1	12	2	11	3	11	6	11	8	11	5	11	0	0	0	0	0	0	13	2256
19	1	12	2	12	3	12	4	12	9	13	7	13	5	12	0	0	0	0	13	2164
20	1	12	2	12	3	13	4	12	9	12	5	12	8	12	0	0	0	0	13	2064
21	1	12	2	12	3	12	4	12	9	13	7	12	6	12	0	0	0	0	13	2051
22	1	12	2	12	3	12	4	12	9	12	7	13	6	12	0	0	0	0	13	1936
23	1	11	2	11	3	12	7	12	8	12	6	12	0	0	0	0	0	0	13	1921
24	1	12	2	12	3	12	4	11	8	11	9	11	0	0	0	0	0	0	13	1912
25	1	12	2	12	3	12	4	12	9	11	7	11	0	0	0	0	0	0	13	1812
26	1	13	2	12	3	12	4	13	9	13	7	13	8	13	0	0	0	0	13	1796
27	1	11	2	11	3	11	9	12	4	11	8	11	7	11	0	0	0	0	13	1696
28	1	12	2	12	3	12	4	13	8	13	6	12	7	13	0	0	0	0	13	1690
29	1	11	2	11	3	11	9	12	4	12	8	12	5	12	0	0	0	0	13	1648
30	1	12	2	12	3	12	4	13	8	13	6	13	9	13	7	13	0	0	13	1638
31	1	11	2	11	3	12	7	11	9	12	5	11	6	12	0	0	0	0	13	1610
32	1	13	2	13	3	13	4	13	7	12	5	13	9	12	0	0	0	0	13	1607
33	1	11	2	11	3	12	7	11	9	11	4	11	8	11	0	0	0	0	13	1563
34	1	13	2	13	3	13	4	13	7	13	9	14	0	0	0	0	0	0	13	1510
35	1	13	2	13	3	13	4	13	7	13	9	12	8	12	5	13	0	0	13	1424
36	1	13	2	12	3	12	4	12	6	12	7	12	8	12	5	12	0	0	13	1266
37	1	13	2	12	3	13	4	12	9	13	6	13	8	12	5	12	0	0	13	1141
38	1	13	2	13	3	13	4	12	6	12	5	12	8	12	9	12	0	0	13	1128
39	1	12	2	13	3	13	9	12	4	12	8	12	5	12	0	0	0	0	13	1110
40	1	13	2	13	3	13	4	12	6	13	8	12	9	12	0	0	0	0	13	1059



RESUMEN CLASIFICACIÓN

Clasificación tipo: 2 Modelo: clase2 Resolución temporal: diaria Año 1999

CLASIFICACIÓN

Rango	Etiqueta	
0	0.30	11
0.30	1.00	12
1.00	10.00	13
10.00	20.00	14
20.00	30.00	15
30.00	40.00	16
40.00	50.00	17
50.00	60.00	18
60.00	70.00	19
70.00	80.00	20
80.00	90.00	21
90.00	100	22

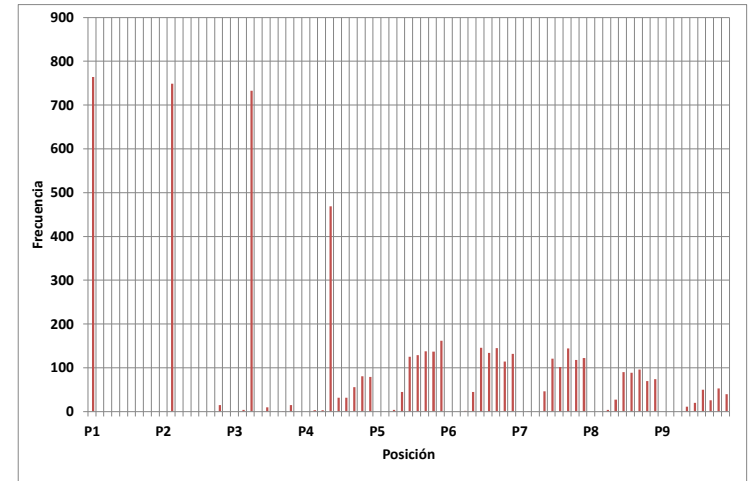
FRECUCIA POR POSICIÓN Y ETIQUETA

Posición	Frecuencia Vecino ¹	Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³
			Max	Mín		
P1	1	NW	13	19	11	1
P2	2	N	12	19	12	2
P3	3	NE	12	19	13	1
P4	4	W	12	18	14	3
P5	9	C	12	18	15	2
P6	5	E	12	19	16	1
P7	7	SW	13	19	17	1
P8	7	S	12	19	18	1
P9	8	SE	13	19	19	1
		Ct	13	17	20	#N/A
					21	#N/A
					22	#N/A

ESTADÍSTICA DE REGLAS

Número de patrones	9176
Porcentaje de reducción de información	29
Total de reglas	6500
Número de reglas usadas	764
Porcentaje de reglas usadas	12
Número de reglas sin usar	5736
Porcentaje de reglas sin usar	88

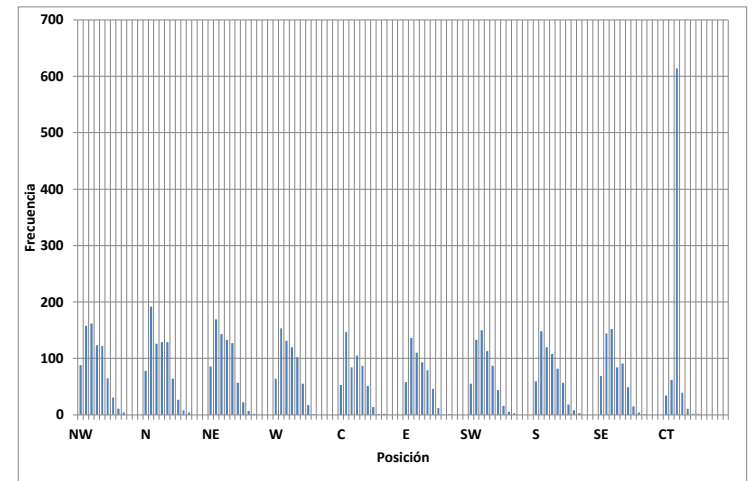
- ¹ Corresponde al vecino que con mayor frecuencia le pregunta para tomar la decisión.
² Corresponde al valor de etiqueta más frecuente en el vecino
³ Corresponde al vecino donde con mayor frecuencia se presenta el valor de etiqueta



REGLAS MAS FRECUENTES

*Solo se muestran las primeras 40

Ranking	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8	V8	P9	V9	OUT	FRECUCIA
1	1	13	2	13	3	13	4	13	7	13	9	13	8	13	0	0	0	0	13	636895
2	1	12	2	12	3	12	4	12	9	12	7	12	5	12	6	12	0	0	13	231331
3	1	11	2	11	3	11	9	11	7	11	5	11	4	11	6	11	8	11	12	202805
4	1	14	2	14	3	14	4	14	9	14	6	14	7	14	0	0	0	0	13	35116
5	1	12	2	12	3	12	4	12	9	11	7	12	5	12	0	0	0	0	13	4020
6	1	11	2	11	3	11	9	11	4	11	8	12	0	0	0	0	0	0	13	3510
7	1	13	2	13	3	12	4	13	7	13	8	13	5	13	9	13	0	0	13	3123
8	1	12	2	12	3	11	4	12	8	12	5	12	9	12	0	0	0	0	13	3120
9	1	15	2	15	3	15	4	15	5	15	6	15	7	15	8	15	0	0	13	3053
10	1	11	2	12	3	12	8	12	5	12	7	12	9	12	6	12	0	0	13	3023
11	1	13	2	13	3	13	4	13	7	12	5	13	9	13	6	13	0	0	13	2940
12	1	11	2	11	3	11	9	11	7	12	6	11	5	11	0	0	0	0	13	2826
13	1	13	2	13	3	13	4	13	7	13	9	12	8	13	5	13	0	0	13	2781
14	1	12	2	11	3	11	6	11	8	11	5	11	0	0	0	0	0	0	13	2773
15	1	12	2	12	3	12	4	12	9	12	7	11	6	12	5	12	0	0	13	2691
16	1	12	2	13	3	13	9	13	6	13	7	13	4	13	5	13	8	13	13	2567
17	1	12	2	12	3	12	4	12	9	13	7	13	5	12	0	0	0	0	13	2505
18	1	12	2	13	3	13	9	13	6	13	7	12	0	0	0	0	0	0	13	2464
19	1	12	2	12	3	12	4	12	9	13	7	12	6	12	0	0	0	0	13	2462
20	1	12	2	12	3	12	4	12	9	12	7	13	6	12	0	0	0	0	13	2281
21	1	11	2	11	3	12	7	12	8	12	6	12	0	0	0	0	0	0	13	2193
22	1	12	2	12	3	13	4	12	9	12	5	12	8	12	0	0	0	0	13	2011
23	1	12	2	12	3	12	4	13	8	13	6	12	7	13	0	0	0	0	13	1981
24	1	12	2	12	3	12	4	13	8	13	6	13	9	13	7	13	0	0	13	1926
25	1	13	2	12	3	12	4	13	9	13	7	13	8	13	0	0	0	0	13	1925
26	1	11	2	11	3	11	9	12	4	11	8	11	7	11	0	0	0	0	13	1890
27	1	11	2	11	3	11	9	12	4	12	8	12	5	12	0	0	0	0	13	1890
28	1	11	2	11	3	12	7	11	9	12	5	11	6	12	0	0	0	0	13	1785
29	1	13	2	13	3	13	4	13	7	12	5	13	9	12	0	0	0	0	13	1735
30	1	12	2	12	3	12	4	11	8	11	9	11	0	0	0	0	0	0	13	1732
31	1	12	2	12	3	12	4	12	9	11	7	11	0	0	0	0	0	0	13	1659
32	1	11	2	11	3	12	7	11	9	11	4	11	8	11	0	0	0	0	13	1516
33	1	13	2	13	3	13	4	13	7	13	9	12	8	12	5	13	0	0	13	1449
34	1	13	2	12	3	12	4	12	6	12	7	12	8	12	5	12	0	0	13	1381
35	1	12	2	12	3	12	4	12	9	13	7	12	6	13	0	0	0	0	13	1283
36	1	13	2	13	3	13	4	13	7	13	9	14	0	0	0	0	0	0	13	1280
37	1	13	2	12	3	13	4	12	9	13	6	13	8	12	5	12	0	0	13	1260
38	1	13	2	13	3	12	4	13	7	13	8	13	5	13	9	12	6	12	13	1258
39	1	13	2	13	3	13	4	12	6	12	5	12	8	12	9	12	0	0	13	1208
40	1	13	2	12	3	12	4	13	9	12	7	13	8	12	5	12	0	0	13	1192



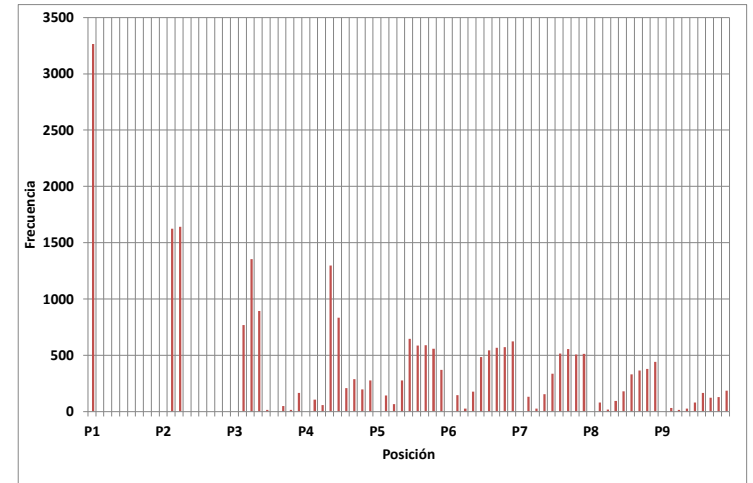
RESUMEN CLASIFICACIÓN

Clasificación tipo: 1 Modelo: clase1 Resolución temporal: diaria Año 1999

CLASIFICACIÓN		
Rango	Etiqueta	
0	0.20	11
0.20	0.50	12
0.50	0.70	13
0.70	1.00	14
1.00	5.00	15
5.00	10.00	16
10.00	1000.00	17
		18
		19
		20
		21
		22

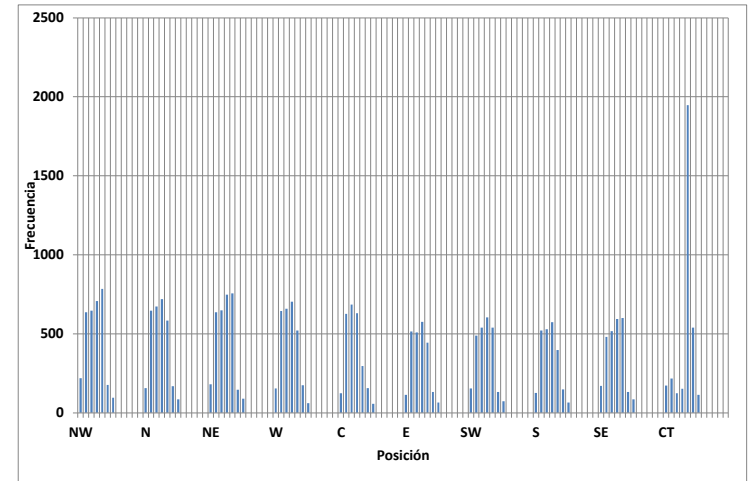
FRECUENCIA POR POSICIÓN Y ETIQUETA						
Posición	Frecuencia Vecino ¹	Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³
			Max	Mín		
P1	1	NW	15	17	11	1
P2	3	N	14	17	12	2
P3	3	NE	15	17	13	5
P4	4	W	14	17	14	3
P5	5	C	13	17	15	1
P6	9	E	14	17	16	1
P7	7	SW	14	17	17	1
P8	9	S	14	17	18	
P9	9	SE	15	17	19	
		Ct	15	17	20	
					21	
					22	

ESTADÍSTICA DE REGLAS	
Número de patrones	9176
Porcentaje de reducción de información	30
Total de reglas	6414
Número de reglas usadas	3265
Porcentaje de reglas usadas	51
Número de reglas sin usar	3149
Porcentaje de reglas sin usar	49



- ¹ Corresponde al vecino con mayor frecuencia le pregunta para tomar la decisión.
- ² Corresponde al valor de etiqueta más frecuente en el vecino
- ³ Corresponde al vecino donde con mayor frecuencia se presenta el valor de etiqueta

REGLAS MAS FRECUENTES																FRECUENCIA				
Ranking	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8		V8	P9	V9	OUT
1	1	15	3	15	2	15	4	15	5	15	8	15	7	15	6	15	0	0	15	465120
2	1	16	2	16	4	16	8	16	3	16	5	16	7	16	9	16	6	16	15	144237
3	1	11	3	11	9	11	7	11	2	11	5	11	6	11	6	11	4	11	14	144178
4	1	12	3	12	4	12	5	12	8	12	2	12	7	12	0	0	0	0	15	121974
5	1	14	2	14	3	14	4	14	9	14	6	14	5	14	0	0	0	0	15	63375
6	1	13	2	13	3	13	4	13	5	13	9	13	6	13	0	0	0	0	15	47048
7	1	17	2	17	3	17	5	17	4	17	9	17	8	17	7	17	0	0	16	42093
8	1	11	3	12	2	12	9	12	6	12	5	12	0	0	0	0	0	0	15	3352
9	1	13	2	13	3	13	4	13	5	13	9	14	0	0	0	0	0	0	15	3252
10	1	12	3	12	4	12	5	12	8	13	9	13	0	0	0	0	0	0	15	2918
11	1	12	3	12	4	12	5	12	8	11	2	12	0	0	0	0	0	0	15	2836
12	1	14	2	14	3	14	4	15	9	15	8	15	0	0	0	0	0	0	15	2730
13	1	15	3	16	9	15	7	15	4	15	8	15	0	0	0	0	0	0	15	2611
14	1	15	3	15	2	15	4	15	5	15	8	14	9	14	0	0	0	0	15	2578
15	1	14	2	15	3	15	4	15	5	15	9	15	6	15	7	15	0	0	15	2443
16	1	12	3	13	4	12	5	12	2	12	8	12	7	12	0	0	0	0	15	2428
17	1	14	2	14	3	14	4	14	9	15	5	14	7	14	0	0	0	0	15	2422
18	1	15	3	16	9	16	6	16	8	16	7	16	5	16	0	0	0	0	15	2379
19	1	12	3	11	2	12	7	12	4	12	9	12	8	12	0	0	0	0	15	2357
20	1	15	3	14	2	15	4	15	5	15	7	15	6	15	0	0	0	0	15	2254
21	1	15	3	15	2	15	4	15	5	15	8	16	9	16	0	0	0	0	15	2233
22	1	11	3	11	9	12	2	11	6	12	8	12	7	12	0	0	0	0	15	2142
23	1	16	2	16	4	16	8	15	6	15	9	15	0	0	0	0	0	0	15	2115
24	1	11	3	11	9	12	2	11	6	11	7	12	8	12	0	0	0	0	15	2115
25	1	15	3	15	2	15	4	15	5	15	8	15	7	14	9	15	6	15	15	2032
26	1	14	2	15	3	15	4	14	5	15	6	15	8	15	0	0	0	0	15	1919
27	1	12	3	11	2	12	7	12	4	12	9	11	0	0	0	0	0	0	15	1890
28	1	14	2	14	3	14	4	14	9	13	5	14	8	13	0	0	0	0	15	1879
29	1	16	2	16	4	16	8	16	3	15	5	16	7	16	0	0	0	0	15	1800
30	1	14	2	14	3	14	4	14	9	15	5	14	7	15	0	0	0	0	15	1746
31	1	12	3	11	2	11	7	12	8	12	0	12	0	0	0	0	0	0	15	1743
32	1	13	2	13	3	13	4	13	5	13	9	12	7	13	0	0	0	0	15	1734
33	1	11	3	12	2	11	9	12	8	12	5	12	6	12	0	0	0	0	15	1659
34	1	11	3	11	9	11	7	12	4	11	6	11	0	0	0	0	0	0	15	1645
35	1	12	3	12	4	11	6	12	9	12	0	0	0	0	0	0	0	0	15	1625
36	1	11	3	11	9	12	2	11	6	11	7	11	0	0	0	0	0	0	15	1530
37	1	16	2	16	4	15	7	15	9	15	3	16	8	15	0	0	0	0	15	1466
38	1	16	2	15	3	15	7	15	5	6	15	9	15	8	15	0	0	0	15	1448
39	1	15	3	16	9	16	6	16	8	16	7	15	0	0	0	0	0	0	15	1441
40	1	13	2	13	3	12	9	13	5	13	8	13	4	13	0	0	0	0	15	1438



**IMPLEMENTACIÓN DE UN MODELO DE AUTÓMATA CELULAR PARA EL
PRONÓSTICO DE LA PRECIPITACIÓN ESPACIAL. CASO DE ESTUDIO
CIUDAD DE BOGOTÁ (COLOMBIA).**

Ana Carolina Santos Rocha
Código: 295346

Trabajo de grado para optar el título de Magíster en Recursos Hidráulicos

Director de proyecto:
Ing. MSc PhD Nelson Obregón Neira

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE BOGOTÁ
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA CIVIL Y AGRÍCOLA
MAESTRÍA EN RECURSOS HIDRÁULICOS
Bogotá, 2010

AGRADECIMIENTOS

Gracias a Dios y a todos los Santos que siempre me apoyaron y favorecieron para cumplir esta meta en mi vida. A mi director de tesis Nelson Obregón porque con su apoyo y estímulo logro que la motivación de la investigación cada día fuera mayor. A Miguel Angel le doy un agradecimiento especial, pues que fue un gran colaborador en el desarrollo de la presente investigación.

A los ingenieros Fabio Bernal y Víctor Peñaranda, quienes aportaron con información y conocimiento al desarrollo de la investigación.

Al IDEAM, quien participo como cofinanciador de la investigación y colaboro de manera amable en la adquisición de la información, de manera muy especial agradecimiento al ingeniero Alberto Pardo, quien fue el puente de comunicación entre la Universidad Nacional y el IDEAM

A mis compañeros del Instituto Geofísico quienes siempre me brindaron una voz de aliento y aportaron al desarrollo de la investigación con ideas y críticas constructivas.

CONTENIDO

	Pág.
INTRODUCCIÓN	14
ALCANCES Y DEFINICIÓN DEL PROBLEMA DE INVESTIGACIÓN.....	18
OBJETIVOS.....	19
OBJETIVO GENERAL	19
OBJETIVOS ESPECÍFICOS	19
1 MARCO TEÓRICO.....	20
1.1 SISTEMAS DINÁMICOS Y AUTÓMATAS CELULARES.....	20
1.1.1 Revisión de la Historia y Evolución de los Autómatas Celulares	20
1.1.2 Definición de los Elementos de los Autómatas Celulares.....	22
1.2 MINERÍA DE DATOS Y EXTRACCIÓN DE CONOCIMIENTO.....	26
1.3 PRECIPITACIÓN.....	32
1.3.1 Complejidad y Patrones Espaciales en la Precipitación	32
1.3.2 Modelación de la Precipitación con AC.....	35
2 CARACTERIZACIÓN DE LA ZONA DE ESTUDIO E INFORMACIÓN DISPONIBLE	37
2.1 INFORMACIÓN PLUVIOGRÁFICA DISPONIBLE	38
2.2 INFORMACIÓN DE VIENTO.....	45
2.3 PRE-PROCESAMIENTO DE LOS DATOS.....	50
2.3.1 Análisis Exploratorio de Datos (AED).....	50
2.3.2 Análisis Exploratorio de Datos Espaciales (AEDE)	56
2.4 SELECCIÓN DEL PERÍODO DE SIMULACIÓN Y EVENTOS DE PRECIPITACIÓN	59
3 ESTIMACION DE LOS CAMPOS DE PRECIPITACIÓN.....	61
3.1 CONSIDERACIONES PARA LA CONSTRUCCIÓN DE LOS CAMPOS DE PRECIPITACIÓN	64
3.1.1 Selección del Tamaño de Celda	64
3.1.2 Alineación de la Malla	68

3.1.3 Cobertura Espacial de la Interpolación.....	69
3.2 SELECCIÓN DE MÉTODO DE INTERPOLACIÓN -VALIDACIÓN CRUZADA.....	71
4 PROPUESTA METODOLÓGICA PARA LA IMPLEMENTACIÓN DE UN AC EN EL ESTUDIO DE LA DINÁMICA Y PRONÓSTICO DE CAMPOS DE LLUVIA 75	
4.1 PROCESAMIENTO DE LA INFORMACIÓN	75
4.2 MODELO DE AUTÓMATA CELULAR.....	77
4.2.1 Definición de los Elementos del Autómata	77
4.2.2 Reproducción del Modelo de Autómata Celular	83
4.2.3 Evaluación del Desempeño del Modelo	86
4.2.4 Análisis de Sensibilidad	89
5 IMPLEMENTACIÓN DE LA METODOLOGIA PARA EL MODELAMIENTO DE LA PRECIPITACIÓN CON UN AUTÓMATA CELULAR	91
5.1 PROCESAMIENTO DE LA INFORMACIÓN	91
5.2 MODELO DE AUTÓMATA CELULAR.....	92
5.2.1 Definición de los Elementos del Autómata Celular	92
5.2.2 Reproducción del Modelo de Autómata Celular	99
5.2.3 Desempeño del Modelo.....	104
5.2.4 Análisis de Sensibilidad	106
6 ESTADISTICA DE REGLAS.....	121
6.1 REDUCCIÓN DE LA INFORMACIÓN	121
6.2 USO DE REGLAS EN EL PRONÓSTICO DE LA PRECIPITACIÓN	123
6.3 ANÁLISIS DE FRECUENCIAS POR POSICIÓN Y ETIQUETA	125
7 RESULTADOS Y DISCUSIÓN.....	129
8 CONCLUSIONES Y RECOMENDACIONES	134
9 TRABAJO FUTURO	136
BIBLIOGRAFÍA.....	137
ALGORITMOS.....	143
APENDICES.....	152

LISTA DE TABLAS

Tabla 2-1. Resumen de estaciones con información de precipitación a nivel diario y subhoraria, para el periodo comprendido entre enero 1 de 1995 al 31 de diciembre de 1996, en la ciudad de Bogotá	39
Tabla 2-2. Listado de estaciones disponibles con registro de precipitación subhoraria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá...	40
Tabla 2-3. Listado de estaciones disponibles con registro de precipitación diaria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá.	42
Tabla 2-4. Listado de estaciones disponibles con registro de dirección y magnitud de viento, para el período comprendido entre el 01-09-1997 al 01-01-2002. Ciudad de Bogotá.	45
Tabla 2-5. Eventos seleccionados para el estudio de pronóstico.	59
Tabla 3-1. Clasificación de diferentes métodos de interpolación espacial	62
Tabla 3-2. Parámetros del semivariograma para precipitación diaria. Caso 1: Estaciones localizadas en el casco urbano, Caso 2: Estaciones localizadas en el casco urbano y alrededores	63
Tabla 3-3. Comparación de metodologías de interpolación en función del tamaño de celda.	72
Tabla 4-1. Ejemplo de asignación de etiquetas a la serie de datos de precipitación diaria.	79
Tabla 4-2. Posibles medidas de desempeño del modelo de pronóstico de precipitación.	86
Tabla 5-1. Asignación de etiquetas de acuerdo a los rangos de precipitación diaria [mm].	94
Tabla 5-2. Asignación de etiquetas de acuerdo a los rangos de precipitación [mm] por agregación temporal en evento.....	94
Tabla 5-3. Ejemplo de patrones con igual valor en la entrada pero diferente salida.	97
Tabla 5-4. Número de modelos propuestos para el pronóstico de precipitación diaria, en función de la estimación de los patrones.....	98
Tabla 5-5. Número de patrones y reglas por clasificación para una plantilla Moore de radio 1. Reglas para el pronóstico de la precipitación diaria.....	99
Tabla 5-6. Tiempo de cómputo del pronóstico de precipitación diaria para los meses de marzo, abril y mayo.....	99
Tabla 5-7. Asignación de etiquetas de acuerdo a los rangos de precipitación por agregación temporal para el evento de abril 12 de 1995.	113

Tabla 5-8. Frecuencias de las reglas más utilizadas en el modelo de AC para la tormenta de abril 12 de 1995 con resolución temporal 2h.....	114
Tabla 5-9. Organización de reglas de acuerdo a la frecuencia por salida de las reglas más utilizadas en el modelo de AC para la tormenta de abril 12 de 1995 con resolución temporal 2h.....	116
Tabla 6-1. Numero de reglas utilizadas y no utilizadas en cada modelo de pronóstico de precipitación.....	124
Tabla 6-2. Vecino más frecuente al que se le hace para tomar una decisión usando una plantilla MOORE.	127
Tabla 6-3. Resumen de estadística de reglas para el modelo Moda4 en el año 1995. .	128
Tabla 6-4. Resumen de estadística de reglas para el modelo Moda5 en el año 1995. .	128
Tabla 6-5. Resumen de estadística de reglas para el modelo Moda1 en el año 1995. .	128

LISTA DE FIGURAS

Figura 1-1. Algunas <i>lattice</i> , celdas y vecindades para autómatas celulares bidimensionales	23
Figura 1-2. Algunos tipos de vecindad utilizadas en los autómatas celulares bidimensionales.....	24
Figura 1-3. Posibles condiciones de frontera en la configuración de un autómata celular.	25
Figura 1-4. Posibles comportamientos de los autómatas celulares unidimensionales. ...	26
Figura 1-5. Proceso del descubrimiento del conocimiento de bases de datos (KDD). ...	28
Figura 1-6. Conjunto de patrones que describen el nivel del caudal de escorrentía en función de variables topográficas e hidrológicas.....	29
Figura 1-7. Árbol de decisión que define el nivel del caudal de escorrentía en función de variables topográficas e hidrológicas.....	30
Figura 1-8. Mapas de isoyetas para la tormenta del 13 de enero de 1993 en la ciudad de Bogotá.	34
Figura 1-9. Trayectoria de los centros de tormenta, para el evento registrado el 13 de enero de 1993 en la ciudad de Bogotá.	35
Figura 2-1. Localización de la zona de estudio.....	37
Figura 2-2. Carta pluviográfica de la estación LOS TAQUES.....	39
Figura 2-3. Estaciones disponibles con registro de precipitación subhoraria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá.....	41
Figura 2-4. Estaciones disponibles con registro de precipitación diaria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá.....	44
Figura 2-5. Estaciones disponibles con registro de viento con resolución subhoraria, para el período comprendido entre el 01-09-1997 al 01-01-2002. Ciudad de Bogotá.	46
Figura 2-6. Localización de las estaciones con información pluviográfica y dirección de viento. Ciudad de Bogotá.....	49
Figura 2-7. Serie de tiempo de registros de precipitación diaria de la estación pluviográfica ALCO para el período comprendido entre 01-01-1995 y 31-12-1999.	51
Figura 2-8. Diagrama de cajas para el análisis de registros de precipitación diaria para la estación ALCO. Período de registro 01-01-1995 y 31-12-1999.....	52
Figura 2-9. Función de autocorrelación para el análisis de registros de precipitación diaria para la estación ALCO. Período de registro 01-01-1995 y 31-12-1999.	53
Figura 2-10. Espectro de potencia de registros de precipitación diaria de la estación pluviográfica ALCO para el periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.	54

Figura 2-11. Histograma de frecuencias y ajuste a distribuciones de probabilidad para los registros de precipitación diaria de la estación pluviográfica ALCO para el periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.....	55
Figura 2-12. Distribución espacial de la precipitación diaria para periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.	56
Figura 2-13. Distribución espacial de la precipitación total diaria multianual [mm] para periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.....	57
Figura 2-14. Número de datos registrados, no registrados y con valor cero por estación, para el periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.....	58
Figura 2-15. Diagrama de cajas para el análisis de registros de precipitación diaria para la estación ALCO. Período de registro 01-01-1995 y 31-12-1999.	60
Figura 3-1. Representación grafica de los modelos de semivariograma para algunos registros de precipitación diaria.	64
Figura 3-2. Frecuencia de la mínima distancia de desplazamiento del centro de tormenta.	65
Figura 3-3. Desplazamiento del centro de tormenta en la ciudad de Bogotá.....	67
Figura 3-4. Rosa de los vientos anual construida a partir de registros de las estaciones de Calidad de Aire de Bogotá. Año 1997 a 2000.	68
Figura 3-5. Rosa de los vientos construida partir de registros de las estaciones de Calidad de Aire de Bogotá con resolución cada 10 minutos del periodo comprendido entre los años 1997 a 2006.....	69
Figura 3-6. Cobertura de la interpolación espacial, de acuerdo con la cercanía de la malla del TRMM. Ciudad de Bogotá.....	70
Figura 3-7. Zonificación pluviográfica de las estaciones localizadas en la ciudad de Bogotá y alrededores.	71
Figura 3-8. Comparación del ajuste de las variaciones del método de Kriging.....	73
Figura 3-9. Comparación del ajuste obtenido por cada una de las metodologías empleadas.	74
Figura 4-1. Esquema general de la metodología de implementación de u modelo de AC en el estudio de pronóstico de precipitación.	75
Figura 4-2. Metodología del procesamiento de la información para la implementación del modelo de autómeta celular.	76
Figura 4-3. Metodología propuesta de la implementación del modelo de autómeta celular para el pronóstico de la precipitación espacial.....	78
Figura 4-4. Ejemplo de clasificación del mapa interpolado para la serie de datos de precipitación diaria.....	79
Figura 4-5. Tamaños y formas de vecindad.....	80

Figura 4-6. Ejemplo de la definición de una frontera adiabática.	80
Figura 4-7. Construcción de la tabla de patrones para una vecindad Moore de radio 1.82	
Figura 4-8. Ejemplo de un árbol de decisión con etiquetas que representan la categoría de precipitación.	83
Figura 4-9. Ejemplo del pronóstico del estado de una celda.....	84
Figura 4-10. Asignación de condiciones iniciales para el caso de pronóstico actualizando las condiciones iniciales.	85
Figura 4-11. Asignación de condiciones iniciales para el caso de pronóstico sin actualizar las condiciones iniciales.	85
Figura 4-12. Ejemplo de la evaluación del desempeño del modelo de pronóstico de precipitación usando el Error Absoluto.	87
Figura 4-13. Diagrama de dispersión para indicar el desempeño del pronóstico.	87
Figura 4-14. Comparación de la precipitación total en el mapa clasificado y el mapa de pronóstico.	88
Figura 4-15. Remplazo de los valores de etiquetas a valores de precipitación en milímetros de acuerdo a la clasificación asignada.	89
Figura 4-16. Comparación de la precipitación total en milímetros del campo de precipitación y los valores remplazados en el mapa de pronóstico.	89
Figura 5-1. Descripción de las plantillas utilizadas en el modelo de autómata celular.	93
Figura 5-2. Establecimiento de la condición de frontera para la vecindad de Moore y Von Neuman.	95
Figura 5-3. Pasos a seguir para la construcción de un árbol de decisión.....	96
Figura 5-4. Variación de la cantidad de patrones por año según la vecindad.	97
Figura 5-5. Ejemplo de pronóstico de precipitación diaria para el 2 de marzo de 1995 usando diferentes modelos.	102
Figura 5-6. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 desde las 10 am y terminando el 13-abril-1995 a las 4 am. Resolución temporal 2 horas.	103
Figura 5-7. Variación de la precipitación total para la tormenta del 12 de abril. Resolución temporal 30 minutos.....	105
Figura 5-8. Diagrama de dispersión para comparar el valor del celda clasificado y el valor pronosticado para las primeras 14 horas de la tormenta de abril 12 de 1995.	106
Figura 5-9. Frecuencia de posibles salidas en la decisión del árbol de decisión para el modelo de pronóstico de precipitación diaria	107
Figura 5-10. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 cambiando la condición inicial para cada instante de tiempo.	109

Figura 5-11. Comparación de los diagrama de precipitación total variando las condiciones iniciales para cada instante de tiempo	110
Figura 5-12. Comparación de los diagrama de dispersión variando las condiciones iniciales para cada instante de tiempo	110
Figura 5-13. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 con resolución temporal de 30 minutos.	112
Figura 5-14. Diagrama de dispersión para comparar el valor del celda clasificado y el valor pronosticado	113
Figura 5-15. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 con resolución temporal de 2h. Simulación solo con las 13 reglas mas frecuentes del modelo inicial.....	115
Figura 5-16. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 con resolución temporal de 2h. Simulación con 33 reglas.....	117
Figura 5-17. Comparación de los diagramas de dispersión para el evento ocurrido el 12-abril-1995 con resolución temporal de 2h, variando la cantidad de reglas de evolución.	118
Figura 5-18. Matriz de coeficientes de filtraje. Kernel para el uso de un filtro de paso bajo.	119
Figura 5-19. Aplicación de un filtro de paso bajo a las imágenes del modelo de AC para el evento ocurrido el 12-abril-1995 con resolución temporal de 2h	120
Figura 6-1. Secuencia de la serie de pronóstico para el periodo comprendido desde mayo 19 a mayo 23 de 1995 para los modelos "Moda1" y "Moda4".....	123
Figura 6-2. Secuencia de la serie de pronóstico para el periodo comprendido desde abril 18 a abril 20 de 1996 para los modelos "Clase2" y "Moda4".....	125
Figura 6-3. Identificación de los niveles, nodos y conectores del árbol de decisión ...	126
Figura 7-1. Variación de la precipitación total diaria para el periodo marzo, abril y mayo en los modelos Moda1, Clase1, Moda5 y Clase5.	130
Figura 7-2. Variación de la precipitación total para el modelo "Clase6" en el año 1995.	132

LISTA DE ALGORITMOS

Algoritmo 1. Interpolación especial con el método del inverso cuadrático de la distancia	143
Algoritmo 2. Construcción de la tabla de patrones	145
Algoritmo 3. Reproducción del autómata celular	148
Algoritmo 4. Estadística de reglas	150

LISTA DE APÉNDICES

Apéndice A. Caracterización de la precipitación diaria por estación.....	153
Apéndice B. Métodos de interpolación espacial.....	308
Apéndice C. Caracterización de la los modelos propuestos de clasificación.....	315

INTRODUCCIÓN

El conocimiento del comportamiento de variables climáticas como la radiación, temperatura, humedad relativa, precipitación etc., en el campo rural, y para ser más específicos en el campo agrícola es de gran importancia para la toma de decisiones en el sistema agropecuario. Gracias al entendimiento de los fenómenos climáticos, sus ocurrencias y temporalidades se pueden plantear las estrategias de cultivo y evitar algunos problemas desencadenados por el clima. De igual manera ocurre en el ámbito urbano, donde el conocimiento del clima contribuiría a mejorar el diseño de estructuras hidráulicas de desagüe en las edificaciones y vías, entre otros aspectos. Sin dejar de lado un problema reconocido a nivel mundial, como lo son las inundaciones desencadenadas después de fuertes eventos de precipitación, donde el papel del pronóstico es relevante en planeación de sistemas de alarma y evacuación.

La precipitación es quizás el proceso más importante dentro del ciclo hidrológico global. Sobre la superficie la precipitación aporta volúmenes de agua a lagos y ríos, a nivel subsuperficial contribuye a las zonas de almacenamiento, las cuales son la fuente primaria de abastecimiento de animales y plantas.

El entendimiento de la dinámica de la precipitación y en especial la presentada en el trópico se dificulta por varias razones (Mejía y Poveda, 2005): a) La complejidad inherente a la dinámica espacio-temporal de la precipitación tropical; b) La ausencia de información detallada con alta resolución espacio-temporal, y c) Las creencias generalizadas de que los procesos del clima tropical son poco variables, y que responden a los mismos patrones y leyes de la precipitación en latitudes extra-tropicales. Por lo que se puede decir que el rasgo más notable de la precipitación tropical es su extraordinaria variabilidad espacio-temporal, más aun en escalas pequeñas de espacio y tiempo.

La observación de la precipitación se remonta desde la antigüedad, por lo que los métodos e instrumentos de medición han evolucionado, tanto que en la actualidad se cuenta con la información suministrada por sensores remotos con una alta resolución espacial y temporal.

En cuanto al pronóstico, el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), cuenta con dos modelos numéricos: MM5 para el pronóstico del tiempo (“*weather*”) y el CCM3 para la predicción del clima (“*climate*”). Por medio de su sitio Web, el IDEAM brinda a sus usuarios de manera gratuita, el pronóstico de la precipitación, junto a otras variables meteorológicas. Para la ciudad de Bogotá (<http://bart.ideam.gov.co/wrfideam/bogota/index.php>), los resultados son mostrados en forma de mapas de isoyetas, por jornadas durante el día para cada 3 horas (mañana, tarde,

noche y madrugada), con datos iniciales del *Global Forecast System* (GFS) de baja resolución, con una resolución espacial de 5 kilómetros.

Es importante resaltar este esfuerzo de pronóstico, no obstante la resolución espacial y temporal es todavía muy baja en el caso de ser requerida en el diseño de obras de drenaje o sistemas de alerta, a pesar que para otras necesidades de planeación de recursos o estudio del comportamiento de la precipitación pueda ser suficiente.

Este es un gran aporte para la planificación de actividades dentro de la ciudad, infortunadamente con los problemas de inundaciones que se han presentado en los últimos años, un pronóstico con ese tipo de resolución espacio-temporal (por jornada) no es el más adecuado. Considerando que los tiempos de respuesta en las cuencas urbanas, en algunas ocasiones son menores a 1 hora, la validez de un pronóstico de horas en que se presente un evento de tormenta de considerable magnitud es muy pobre.

Entonces se hace necesario un modelo que tenga la fortaleza de realizar pronósticos a una resolución temporal horaria o subhoraria. También cabe anotar que la exactitud de los pronósticos depende de la medición efectiva de las variables, la calidad del modelo y la interpretación. Por ello contar con información pluviográfica de alta resolución es indispensable para la construcción de un modelo y de esta manera se podrían identificar patrones de la dinámica pluvial que en resoluciones diarias son difíciles de obtener.

Dentro de la comunidad hidrológica, se han desarrollado modelos de pronóstico de precipitación aplicando diferentes técnicas de predicción como lo son: a) redes neuronales (Gamal y Smith, 2002; Hsu et al., 2000; Bodri y Čermák, 2000); b) sistemas bayesianos (Krzysztofowicz y Maranzano, 2004); c) modelos estadísticos (Barnston et al., 1999), d) modelos ARMA (Burlando et al., 1993,) entre otros. También son reconocidos los modelos globales y regionales, el *Global Spectral Model* y *Regional Spectral Model* (RSM), respectivamente, y el *Numerical Weather Prediction* (NWP). Todos estos modelos y/o técnicas de pronóstico toman información registrada en superficie o desde sensores remotos, y es así como se puede estimar la precipitación a partir de mediciones de pluviógrafos o de radar meteorológico (Krajewski, 1987 y Anhert et al., 1986).

Todas estas aproximaciones son válidas. Algunos modelos se basan en la capacidad para representar adecuadamente los procesos físicos que ocurren en la atmósfera. Si las variables de entrada como la topografía, los usos de suelos, las condiciones iniciales y de frontera, son representadas de manera errónea, entonces estos modelos pueden fallar. Existen casos donde los resultados no reproducen de manera fiel el fenómeno, pues se presenta sobreestimación en el pronóstico. Por lo que el problema sigue abierto. En ese sentido vale la pena el esfuerzo de nuevos modelos que contribuyan a los sistemas de toma de decisión para la evaluación de alternativas.

Debido a la naturaleza espacial de la precipitación, junto con otras variables meteorológicas (p.e. temperatura, humedad, viento, radiación), es común que se realice un acople de los modelos de simulación con los Sistemas de Información Geográfica (SIG), lo

que permite un manejo y análisis espacial de los datos asociado al modelo (Qi y Wu, 1996). Es así como los SIG's proporcionan ayuda para: a) La construcción de variables de entrada para la modelación; b) La identificación de patrones espaciales en los datos; c) La cuantificación de los cambios temporales observados o predichos; d) La evaluación de factores que operan cruzando una variedad de escalas, y e) La visualización de los resultados.

También se puede observar la variación de los fenómenos dentro del espacio mediante el uso de los modelos basados en transición espacial como las Cadenas de Markov y el Autómata Celular (AC) como en la investigación de Zhou y Liebhold, 1995 donde la implementación se basó en el estudio de dinámicas de población. Estos modelos asumen explícitamente que las áreas vecinas influyen en la probabilidad de transición del área o celda central. Un AC es un sistema que se caracteriza por ser discreto en el espacio y en el tiempo. Está compuesto por un conjunto de células (o celdas) dispuestas en el espacio, todas ellas son idénticas y pueden tener infinitos estados, estos dependen de reglas de comportamiento. En síntesis un modelo de AC puede generar patrones complejos a través de reglas simples (Wolfram, 2002). Según Packard y Wolfram (1985), los ACs son una simple idealización matemática de los sistemas naturales, inspirados en la observación del comportamiento colectivo de muchos de los sistemas naturales. Pueden ser descritos como una arreglo de células (o celdas) idénticas y discretas, donde el estado de las celdas evoluciona en un tiempo discreto, de acuerdo a reglas deterministas que están en función de los valores de las células vecinas. También pueden considerarse como idealizaciones discretas de ecuaciones diferenciales. En cualquier instancia los ACs son adecuados para la descripción matemática de modelos de sistemas en el que muchos componentes simples actúan conjuntamente para producir patrones con comportamiento complejo.

Con el fin de acercarse al problema de pronóstico de la precipitación a diferentes escalas temporales y estudiar el comportamiento de las posibles reglas de la naturaleza, se propone en esta investigación concebir e implementar un modelo de simulación espacio-temporal de la dinámica de la precipitación dentro de la ciudad de Bogotá haciendo uso de los fundamentos básicos de los de Automatas Celulares (ACs). Un modelo de pronóstico hidrológico que contemple la distribución espacial y temporal del fenómeno sugiere que se puede demostrar de la variabilidad de la precipitación mostrando patrones de comportamiento y tendencias entre otras, contribuyendo a la toma de decisiones en los sistemas de alarma y evacuación de localidades expuestas a inundaciones.

La presente investigación es una primera aproximación a la reproducción de la lluvia en el espacio y tiempo aplicando los AC, se dispone de registros pluviográficos con finas resoluciones temporales (subhorario), correspondientes a 37 estaciones pluviográficas localizadas en la ciudad de Bogotá y sus alrededores, para un período de dos años. De igual forma se evaluó el desempeño de la modelación de la precipitación con resoluciones horarias y diarias de 77 estaciones para un período de cinco años. En su gran mayoría las estaciones pertenecen a la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB), en menor cantidad a la Corporación Autónoma Regional (CAR) o al IDEAM.

El desarrollo de la modelación espacio temporal de la precipitación con AC incluye los componentes típicos del protocolo de modelación hidrológica, solo que se presta énfasis en la determinación de las reglas de evolución de la precipitación. Este aspecto es el equivalente a la calibración de los modelos tradicionales. Esta modelación tiene un gran soporte en la geoestadística, debido a que sobre la estimación de los campos de precipitación se realiza el estudio de patrones y estimación de las reglas.

El modelo presenta varios escenarios con diferentes resoluciones temporales que pasan de la diaria a la 30 minotal, lo que contribuye a realizar modelaciones hidrológicas a diferentes escales de tiempo y comprender los problemas de escala y agregación de la precipitación.

El desarrollo del documento inicia con una revisión bibliográfica de los sistemas dinámicos y la explicación de los conceptos básicos de la teoría de los autómetas celulares, junto con una breve reseña del estado del arte de la modelación de la precipitación y la evidencia de su comportamiento caótico. El segundo capítulo realiza una descripción de la zona de estudio y el recuento de la información disponible y el procesamiento y caracterización de la misma. El capítulo tres muestra la base geoestadística de la estimación de los campos de precipitación, es un capítulo donde el lector encontrara las consideraciones que se tuvieron en cuenta para la interpolación espacial, como la alineación de la malla, tamaño de celda y selección del método de interpolación. Ya entrando al interés de la investigación en el cuarto capítulo señala la metodología empleada para la modelación de la precipitación con AC, siguiendo todo un protocolo de modelación, haciendo énfasis en la estimación de las reglas del autómeta y la validación de los resultados. El quinto capítulo muestra algunos resultados de la implementación de la metodología de pronóstico. En el sexto capítulo se presenta la estadística de reglas, este capítulo tiene como objetivo mostrar la frecuencia y la identificación de las reglas más importantes en la implementación de la metodología. En el siguiente capítulo se presenta la discusión de los resultados y para finalizar el lector encontrara las conclusiones y recomendaciones; de manera adicional se sugieren trabajos futuros con los que se espera se profundice en el estudio y entendimiento de la precipitación y se señalan algunas sugerencia para la modelación con AC con otras variables de interés en el campo de la hidrología y de ámbito medio ambiental a nivel de atmósfera.

ALCANCES Y DEFINICIÓN DEL PROBLEMA DE INVESTIGACIÓN

De acuerdo con lo anterior, el impacto más importante de la presente investigación radica en que los autómatas celulares han sido aplicados en dinámicas de coberturas de suelo, escorrentía, transporte urbano, incluso en el pronóstico de variables del clima (Mora, 2005), pero no es muy reconocido en el campo de la precipitación, por ello se valora el trabajo futuro que se lograría realizar desde el grupo de investigación GIREH en el desarrollo de nuevos modelos de aproximación que puedan contribuir al entendimiento de los fenómenos hidroclimáticos.

El alcance de la presente investigación se enmarca en el estudio del pronóstico de precipitación haciendo uso de un modelo de Autómata Celular, reproduciendo el fenómeno para diferentes niveles de agregación temporal, obedeciendo a un periodo de tiempo comprendido en entre los años 1995 a 1999. La metodología expuesta plantea un modelo guiado por datos que se caracteriza por ser robusta y genérica. Se parte de la información pluviográfica la ciudad de Bogotá, el interés de la modelación hidrológica y por ende del pronóstico es generar escenarios críticos que contribuyan a las entidades a tomar decisiones en cuanto al diseño de estructuras o definir estrategias en los planes y/o proyectos de sistemas de alertas. Por el hecho de partir de información de estaciones puntuales se advierte que como resultado del proceso de interpolación para generar los campos de precipitación se encuentra una incertidumbre espacial, este tipo de problemas se podrían obviar si la adquisición de la información se realizara con tecnologías como el radar.

La autora advierte que ningún modelo podrá reproducir con total certeza la naturaleza, ya que existe evidencia de su comportamiento caótico y que técnicamente hablando los procesos que se presentan en ella, (como los hidrometeorológicos) no son lineales, así que se aclara que el modelo presentado en esta investigación representa un avance en la modelación y no es una fiel reproducción del fenómeno de la precipitación; más aún cuando el insumo del modelo solo contó con información de registros de lluvia y no de otras variables meteorológicas y microclimáticas.

OBJETIVOS

Objetivo general

Mediante el uso de Autómatas Celulares desarrollar un modelo de simulación espacial y temporal que permita realizar el pronóstico de la precipitación en una zona de la ciudad de Bogotá.

Objetivos específicos

- Establecer un modelo matemático y computacional basado en la teoría de Autómatas Celulares, que permita simular el pronóstico de la precipitación, de manera que replique su dinámica espacial y temporal.
- Calibrar y validar el modelo de Autómatas Celulares que reproduzca el fenómeno de precipitación para un sector de la ciudad de Bogotá.
- Evaluar el desempeño de un modelo de Autómatas Celulares en el pronóstico de la precipitación.
- Identificar bajo qué condiciones de resolución temporal, tipo de plantilla y período de simulación el modelo de AC responde de mejor manera al pronóstico de lluvia.

1 MARCO TEÓRICO

En esta sección de este documento se describen de manera general los aspectos teóricos de los autómatas celulares, la precipitación y la minería de datos con el objetivo de contextualizar los métodos y conceptos de la modelación de la precipitación a aplicar en esta propuesta. Se parte de la conceptualización de los autómatas celulares (AC) como modelos que representan un sistema dinámico, se describen de manera breve los componentes de los AC y para terminar se muestra el estado del arte de la modelación y/o pronóstico de la precipitación evidenciando la naturaleza compleja del fenómeno.

1.1 SISTEMAS DINÁMICOS Y AUTÓMATAS CELULARES

Dentro del concepto de lo que representa o caracteriza un sistema dinámico está el hecho que reproduce un comportamiento que evoluciona en el tiempo tomando valores de estado diferentes, el comportamiento es complejo y se ve influenciado por los límites del sistema y los elementos o variables que incluye el modelo. Existen varios tipos de sistemas dinámicos y por ende modelos matemáticos de lo representan uno de ellos es el modelo de los autómatas celulares.

1.1.1 Revisión de la Historia y Evolución de los Autómatas Celulares

El Autómata Celular (AC) es un modelo matemático que fue introducido en los años '40 por John Von Neumann y Stanislaw Ulam (Cooper, 1987), con el objeto de crear un modelo real del comportamiento de sistemas extensos y complejos, mediante la integración de células finitas, que interactúan de acuerdo a reglas simples. Existen dos características que a su vez son ventajas de la aplicación de un autómata celular:

- A partir de reglas simples se logra reproducir un fenómeno complejo
- El carácter espacial y de vecindad que implica el estado de un individuo

En síntesis, la idea central que gobierna la modelación con AC, es la reproducción de patrones de observaciones de un sistema complejo mediante el uso de reglas simples, capaces de modelar un comportamiento global consecuencia de las interacciones entre estas mismas reglas (Ilachinski, 2001 y Weimar, 1997). Un sistema complejo está conformado por un gran número de elementos idénticos, que interaccionan localmente y simulan un comportamiento global que no se explica fácilmente a partir de las propiedades de un solo individuo, sino de las interacciones del colectivo.

Uno de los ACs más reconocidos es la llamado “Juego de la vida”, propuesta por el matemático John Horton Conway (1970), y está basado en la evolución de un autómata celular bidimensional.

Consiste en un conjunto de células (o celdas) dispuestas en un plano, donde el estado de una célula es definido por reglas, de manera que las células pueden adoptar dos estados: vivas o muertas, y los cuales se van modificando con el paso del tiempo atendiendo a una serie de reglas sencillas:

- **Nacimiento:** Una célula muerta en un momento dado, cambia de estado en el instante siguiente si tiene exactamente 3 células vivas a su alrededor, en caso contrario permanece muerta en el siguiente instante.
- **Supervivencia:** Una célula viva en un instante dado, permanece viva en el instante siguiente si hay 2 o 3 células vivas a su alrededor; en caso contrario pasa a estar muerta en el instante siguiente.
- **Muerte:** Cada célula viva con ninguna, una, o más de tres células vivas a su alrededor pasa a estar muerta.

Los ACs son sistemas dinámicos donde el tiempo y el espacio son discretos y están compuestos por un conjunto de celdas elementales homogéneas ordenadas en una malla regular espacial de n dimensiones. Cada celda representa un lugar espacial susceptible de encontrarse en cualquiera de los estados finitos que componen el sistema. Este estado puede ser cambiado o alterado de un período al siguiente a través de ciertas reglas de evolución de estado, mediante un conjunto de reglas locales (conocidas también como funciones de transición) que involucran el estado de la celda y de sus celdas cercanas (vecindad), suponiendo que el tiempo se incrementa en forma discreta (Weimar, 1997). Las reglas de interacción que generalmente se aplican a los vecinos inmediatos pueden o no guardar una semejanza con las leyes físicas que gobiernen el fenómeno que se desea reproducir. La actualización de estos estados se hace en paralelo, siguiendo en esencia una regla local que considera sólo la vecindad de cada celda.

Las simulaciones computacionales de los ACs pueden actualizar millones de celdas en un lapso de tiempo muy corto, y son aplicadas en modelos simples de ecuaciones diferenciales como las ecuaciones del calor, de propagación de ondas o de Navier-Stokes. En particular los modelos de autómatas celulares se utilizan frecuentemente en dinámica de fluidos y gases. En la Física, se han desarrollado estudios en los sistemas dinámicos, la aparición de fenómenos colectivos, la turbulencia, el caos y la fractalidad.

Los ACs proveen una gran colección de modelos representativos de estos fenómenos, que pueden ser aislados y estudiados de forma sencilla. El uso sistemático de los autómatas celulares en este campo fue iniciado por Stephen Wolfram a mediados de los años 80. También existen aplicaciones en el estudio de los fenómenos ferromagnéticos, simulación de los procesos de percolación, propagación de incendios, dinámica de poblaciones, estudio del desarrollo urbano, procesos de cristalización, tránsito de vehículos, propagación de epidemias, ocurrencia de sismos (Jiménez et. al., 2000), etc.

La teoría del AC ha sido utilizada con éxito para la simulación del movimiento de fluidos sobre una superficie como el modelo de simulación de escurrimiento de lava volcánica (Crisci et.al., 2004) o al del escurrimiento del agua (Dalponte et. al., 2007;

Rinaldi et. al., 2007 y Vénere et. al., 2004). Algunos estudios donde la aplicación de AC en el área de hidrología son referentes al análisis y pronóstico de zonas de inundación, como el citado por Coppola et. al. (2007)

1.1.2 Definición de los Elementos de los Autómatas Celulares

Los ACs pueden definirse con los siguientes elementos básicos (Karafyllidis y Thanailakis, 1997 y Muñoz, 1996):

- Un plano o espacio n- dimensional, dividido en un número de subespacios homogéneos, conocidos como células o celdas.
- La geometría de celda, cada celda puede estar en uno de un conjunto finito o numerable de estados.
- Una Configuración, la que consiste en asignarle un estado a cada celda del autómata.
- Una Vecindad definida para cada celda, que consiste en un conjunto contiguo de celdas, indicando sus posiciones relativas respecto a la celda misma.
- Una Regla de evolución, la cual define cómo debe cada celda cambiar de estado, dependiendo del estado inmediatamente anterior de su vecindad.
- Un “Reloj virtual” de cómputo conectado a cada celda del autómata, el cual generará pulsos simultáneos a todas las celdas indicando que debe aplicarse la regla de evolución y de esta forma cada celda cambiará de estado.

El tiempo es discreto y sirve como índice en la configuración sucesiva de las simulaciones. Es también homogéneo y por lo tanto esas evoluciones toman lugar de la misma forma, independiente del lugar donde ocurran.

1.1.2.1 Lattice

La *lattice* hace referencia al plano donde evoluciona el AC, la geometría de ésta está dada por la geometría de las celdas. La definición del AC, requiere que la *lattice* sea regular. Para un AC unidimensional, la configuración en el espacio solo puede ser lineal. Un AC bidimensional, pueden tener diferentes tipos de geometría de celda. Los más comunes son: triangular, hexagonal o rectangular (Figura 1-1). La geometría más usada en 3D es la de un cubo.

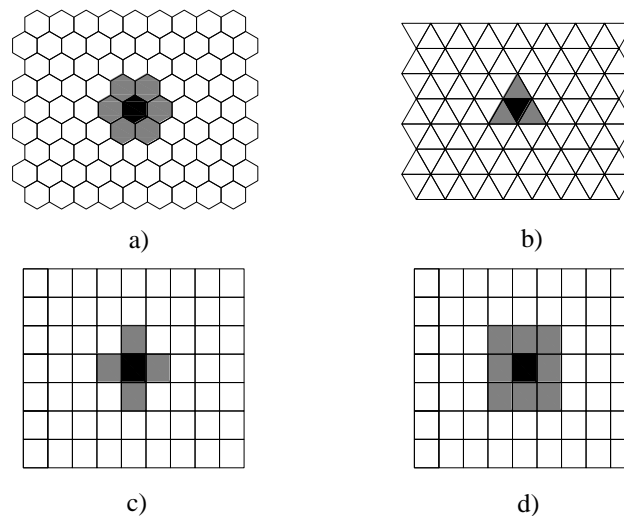


Figura 1-1. Algunas *lattice*, celdas y vecindades para autómatas celulares bidimensionales
a) Celda hexagonal, b) Celda triangular, c) Celda rectangular con vecindad Von Neumann, d) Celda rectangular con vecindad Moore. La celda oscura indica la celda central, la cual actualiza su estado de acuerdo al estado de sus celdas vecinas (celdas sombreadas)

1.1.2.2 Celda y Estados de Celda (Configuración)

La celda es el elemento básico de los ACs, simboliza un lugar del espacio a modelar y guarda un estado. Representa un espacio físico, por lo cual las celdas se distribuyen en un espacio n-dimensional discreto, pudiendo ser $n = 1, 2$ ó 3 , es decir, una representación lineal, bidimensional o tridimensional respectivamente del fenómeno a modelar. Todas las celdas deben ser idénticas y se encuentran formando una malla regular (*lattice*).

El estado es un número o una propiedad de la celda, que puede representar el número de una población de individuos en esa celda o en el caso de esta investigación el valor de precipitación que posee. El estado de las celdas cambia de un instante a otro de acuerdo a un conjunto de reglas de transición comunes a todas las celdas.

1.1.2.3 Vecindad

La vecindad puede ser definida de diferentes formas, de manera que cada celda del autómatas celular se relaciona con las inmediatamente adyacentes. Las vecindades más comunes en una malla de 2-dimensiones son la de Von Neumann que considera sólo las cuatro celdas adyacentes que comparten un borde con la celda central, y la de Moore que considera las celdas diagonales, es decir las ocho celdas circundantes, en la Figura 1-2 se muestran las vecindades más comunes encontradas en la literatura.

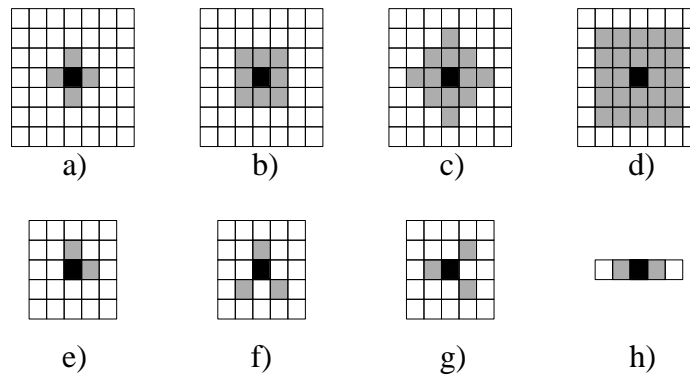


Figura 1-2. Algunos tipos de vecindad utilizadas en los autómatas celulares bidimensionales.
a) Von Neumann con radio 1, b) Moore con radio 1, c) Von Neumann con radio 2, d) Moore con radio 2, e) Smith, f) Cole 1, g) Cole 2, h) Vecindad para un dimensión. La celda oscura indica la celda central, la cual actualiza su estado de acuerdo al estado de sus celdas vecinas (celdas sombreadas).

Fuente: Chen 2004 y Jiménez et.al., 2000.

1.1.2.4 Reglas de Transición

Las reglas de transición son funciones locales ya que ellas definen el estado de la celda en el tiempo $t+1$ basada solamente en su propio estado y el de la vecindad en el instante t . También son funciones uniformes ya que las reglas son las mismas en cualquier lugar, exceptuando las celdas de los bordes de malla, las que pueden tener reglas especiales. De esta forma, las reglas de transición dependen de la geometría de la malla, la vecindad y el conjunto de estados posibles. Las reglas de transición pueden ser definidas de la siguiente manera (Weimar 1997):

- Reglas Determinísticas: En las reglas de este tipo el estado de la celda no es determinado por función de probabilidad. El estado de la celda queda determinada por el estado de sus vecinas. Un ejemplo de este tipo de regla es el “juego de la vida” (Gardner, 1970), en que el estado de una celda queda definida si un determinado número de celdas vecinas se encuentran en un estado específico.
- Reglas Probabilísticas: Este tipo de regla se produce cuando una función entrega uno ó más estados con cierta probabilidad, considerando que la suma de probabilidades de todos los estados posibles es uno. De esta forma, los ACs se convierten en procesos estocásticos en tiempo discreto.

La determinación de las reglas de transición del AC es un proceso que de alguna manera depende del conocimiento del modelador acerca del fenómeno, es decir es un proceso intuitivo. No obstante las reglas de transición pueden también ser definidas a partir de expresiones matemáticas como: el vecino más cercano, funciones de decaimiento, parámetros matriciales, ecuaciones de evaluación multicriterio (EMC), modelos aplicando lógica, lógica difusa o redes neuronales. En el caso de los ACs las reglas de transición son las equivalentes a parámetros de los modelos tradicionales. Algunas investigaciones (Gar-On y Li, 2004) plantean el uso de sistemas expertos como la minería de datos, que envuelve el descubrimiento y captura del conocimiento con base en una serie de datos, y de esta

manera se solucionan los problemas de incertidumbre en la definición de las reglas de transición.

Las técnicas de construcción de las reglas de transición del AC deben ser lo más automatizadas como sea posible. Dentro de las ventajas de uso de los sistemas de aprendizaje se encuentran:

- El árbol de decisión es la manera más eficiente para el aprendizaje inductivo (Huang y Jensen, 1997).
- Los sistemas expertos pueden identificar el umbral de los valores y crear conocimiento a partir de los datos observados.
- Son de fácil acople con los SIG en cuanto al uso de datos espaciales.
- Los AC pueden ser calibrados simultáneamente durante la construcción de las reglas de transición mediante el proceso de minería de datos.
- Las reglas de transición son explícitas lo que facilita su aplicación y comprensión.

1.1.2.5 Condiciones de Frontera o de Borde

A pesar de ser una herramienta de modelación matemáticamente simple, los ACs presentan algunos problemas, como son las celdas límites del espacio a modelar. Estas celdas alteran las reglas de transición al presentar una vecindad diferente a las celdas interiores del sistema. Para solucionar esta situación, Weimar, 1997 y Chen 2004 proponen varias condiciones de borde:

- Frontera periódica. En este caso los extremos se conectan entre sí. De esta forma una malla bidimensional se convierte en un toroide.
- Frontera abierta. Se considera que fuera de la *lattice* residen células, todas con un valor fijo.
- Frontera reflectora. Se considera que las células fuera de la *lattice* reflejan los valores de aquellas dentro de la *lattice*. Así, una célula que estuviera junto al borde de la *lattice* (fuera de ella) tomaría como valor el de la célula que este junto al borde de la *lattice*, dentro de ella.
- Sin frontera. Haciendo uso de implementaciones que hagan crecer dinámicamente el uso de memoria de la *lattice* implementada, se puede asumir que cada vez que las células deben interactuar con células fuera de la *lattice*, esta se hace más grande para dar cabida a estas interacciones. Obviamente, existe un límite (impuesto por la memoria disponible) para esta condición.



Figura 1-3. Posibles condiciones de frontera en la configuración de un autómata celular. a) Periódica, b) Fija 1, c) Adiabática, d) Reflectiva. Las celdas sombreadas son virtuales.

Fuente: Chen 2004.

1.1.2.6 Comportamiento

Se han identificado comportamientos característicos de los AC, tanto en una como en dos dimensiones. De acuerdo con Wolfram (1986a), existen cuatro grupos básicos para los autómatas unidimensionales según su comportamiento espacio-temporal (Ver Figura 1-4):

- Clase I: El estado final es homogéneo (todas las celdas en estado '0' ó '1', sólo puede tener dos estados).
- Clase II: El estado final está formado por un conjunto de estructuras estables o periódicas.
- Clase III: La evolución conduce a un conjunto de patrones bien definidos pero organizados en forma caótica (Estado final caótico).
- Clase IV: Aparecen estructuras localizadas y complejas que perduran a lo largo del tiempo. Estructuras organizadas.

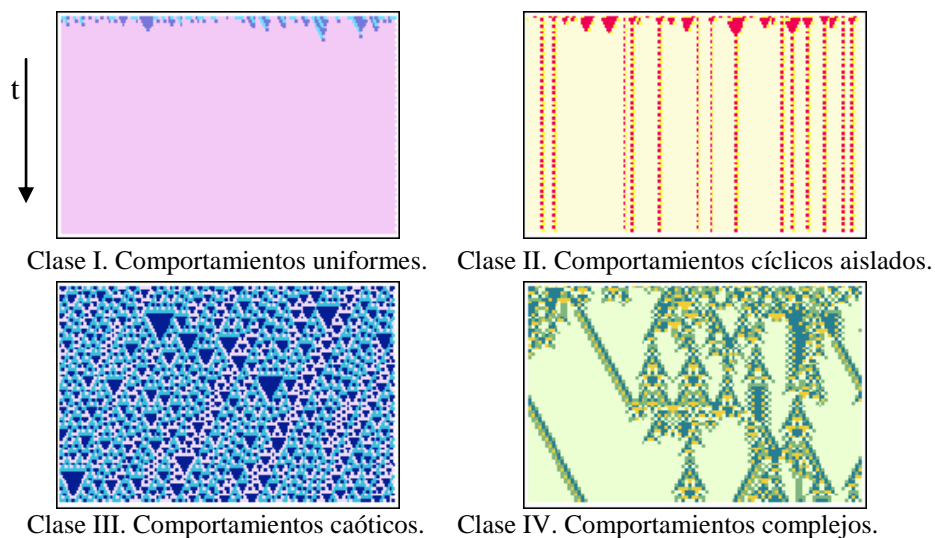


Figura 1-4. Posibles comportamientos de los autómatas celulares unidimensionales.
Esta figura debe leerse de arriba hacia abajo, ya que el eje vertical representa la evolución del AC en el tiempo.

Fuente: Chen 2004.

1.2 MINERÍA DE DATOS Y EXTRACCIÓN DE CONOCIMIENTO

En el estudio del pronóstico y su aplicabilidad a las políticas de toma de decisiones es de gran importancia identificar los criterios o variables para darle solución a los problemas de interés. Un sistema experto es un sistema informático que simula el proceso de aprendizaje, memorización, razonamiento, comunicación y acción de un experto humano en una determinada rama de la ciencia. Estas características le permiten almacenar datos y conocimiento, sacar conclusiones lógicas, tomar decisiones, aprender de la experiencia y de los datos existentes, comunicarse con otros expertos humanos o sistemas expertos, explicar el porqué de las decisiones tomadas y realizar acciones como consecuencia de todo lo

anterior. Los sistemas expertos se consideran una rama de la inteligencia artificial. El uso de sistemas expertos es especialmente recomendado en las siguientes situaciones:

- Cuando los expertos humanos en una determinada materia son escasos.
- En situaciones complejas, donde la subjetividad humana puede llevar a conclusiones erróneas.
- Cuando es muy elevado el volumen de datos que ha de considerarse para obtener una conclusión.

En busca del conocimiento y la comprensión de la información en algunas ocasiones los sistemas expertos se unen con otras disciplinas como la minería de datos. La minería de datos es el proceso de “descubrimiento de conocimiento en bases” (KDD por sus siglas en inglés, *Knowledge Discovery in Databases*), mediante un proceso de extracción de información, previamente desconocida y potencialmente útil (Hernandez, 2008). El hecho de descubrir conocimiento implica buscar patrones de comportamiento aún no conocidos en los datos. El conocimiento se puede representar como: patrones, reglas de conocimiento, asociaciones, grupos, restricciones, tendencias, etc. El proceso de descubrimiento de conocimiento en bases de datos (KDD) involucra varios pasos (Ver Figura 1-5):

- Determinar las fuentes de información: que pueden ser útiles y dónde conseguirlas.
- Diseñar el esquema del almacenamiento de datos (Data Warehouse): con el fin de consignar y unificar de manera operativa toda la información recogida.
- Implantación del almacén de datos: que permita la navegación y visualización previa de sus datos, para discernir qué aspectos puede interesar que sean estudiados.
- Selección, limpieza y transformación de los datos que se van a analizar: Equivalente al análisis exploratorio de datos (AED) con el fin de diseñar una estrategia adecuada para manejar ruido, valores incompletos, secuencias de tiempo, casos extremos entre otros.
- Seleccionar y aplicar el método de minería de datos apropiado: esto incluye la selección de la tarea de descubrimiento a realizar, por ejemplo, clasificación, agrupamiento o *clustering*, regresión, etc. La selección de él o de los algoritmos a utilizar. La transformación de los datos al formato requerido por el algoritmo específico de minería de datos, y llevar a cabo el proceso de minería de datos. Se buscan patrones que puedan expresarse como un modelo o simplemente que expresen dependencias de los datos. El modelo encontrado depende de su función (clasificación) y de su forma de representarlo (árboles de decisión, reglas, etc.). Se tiene que especificar un criterio de preferencia para seleccionar un modelo dentro de un conjunto posible de modelos. Igualmente se tiene que especificar la estrategia de búsqueda a utilizar (normalmente está predeterminada en el algoritmo de minería)
- Evaluación, interpretación, transformación y representación de los patrones extraídos: Los resultados se interpretan y de acuerdo con los resultados es posiblemente regresar a los pasos anteriores. Esto puede involucrar repetir el proceso, quizás con otros datos, otros algoritmos, otras metas y otras estrategias
- Difusión y uso del nuevo conocimiento.

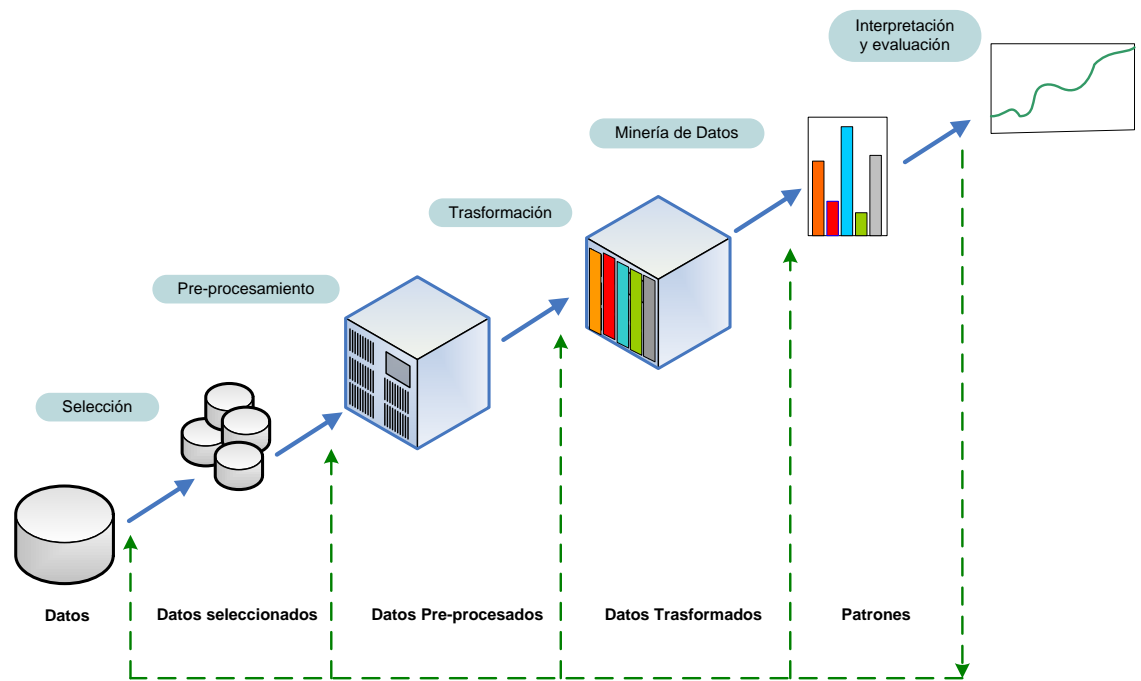


Figura 1-5. Proceso del descubrimiento del conocimiento de bases de datos (KDD).

Fuente: Hernández et. al 2008

Las técnicas usadas por la minería de datos son: redes neuronales, algoritmos genéticos, reglas de inducción, árboles de decisión y el vecino más cercano. Los árboles de decisión son un método de aprendizaje automático frecuentemente utilizado debido a su fácil construcción e interpretación. Es de gran versatilidad pues los valores que pueden tomar las entradas y las salidas pueden ser valores discretos o continuos. Se han utilizado con éxito en diferentes campos que van desde el diagnóstico médico, juegos, sistemas de predicción meteorológica y en la hidrología (Schärer et al. 2006, Preis y Ostfeld 2008, Wei y Hsu 2009).

De manera general, los árboles de decisión son un sistema de predicción basado en reglas. El modelo se constituye de un diagrama de construcciones lógicas que son construidas a partir de la información de una base de datos. En ese sentido sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva y así dar solución a un problema. Pueden ser representados como grupos de reglas SI-ENTONCES para mejorar la interpretación humana. Los principales algoritmos para la construcción de árboles de decisión son: Algoritmo de Hunt (CLS 60's), ID3 (Quinlin 70's-80's), C4.5 (Quinlin 90's), SLIQ y SPRINT.

Un árbol de decisión clasifica variables de entrada ordenándolas hacia abajo desde el nodo inicial o raíz hasta alguna hoja. Cada nodo en el árbol especifica una prueba de alguna etiqueta de la variable anterior, y cada rama descendiente desde éste nodo corresponde a una de las posibles etiquetas para esta variable. Se clasifica una variable en el nodo inicial

del árbol, y se generan ramas hacia abajo examinando las etiquetas específicas para la raíz. Estas ramas llegan hasta nuevos nodos correspondientes a otras variables de entrada. Este proceso se repite para los subárboles, generando otros nodos. (Mitchell, 1997)

En la Figura 1-6 se presenta un ejemplo de datos o patrones de entrenamiento que describen el nivel del Caudal de Escorrentía en función de algunas variables topográficas e hidrológicas. Se observa que se tienen en cuenta cuatro variables de entrada (Tipo o Uso del Suelo, Pendiente del Terreno, Área de Drenaje e Intensidad de la Lluvia) y una variable de salida (Caudal de Escorrentía). Las variables de entrada Tipo o Uso del Suelo y Pendiente del Terreno tienen tres posibles clasificaciones o etiquetas (pavimento asfáltico, zona verde y casas con jardines, y baja, media y alta, respectivamente), las variables Área de Drenaje e Intensidad de la Lluvia poseen 2 etiquetas cada una (grande y pequeña, y llovizna leve y aguacero, respectivamente). La variable de salida Caudal de Escorrentía puede tomar los valores bajo y alto.

Patrón	Tipo o Uso del Suelo	Pendiente del Terreno	Área de Drenaje	Intensidad de la Lluvia	Caudal de Escorrentía
1	Pavimento asfáltico	Baja	Grande	Llovizna leve	Bajo
2	Pavimento asfáltico	Baja	Grande	Aguacero	Alto
3	Zona verde	Baja	Grande	Llovizna leve	Bajo
4	Casas con jardines	Media	Grande	Llovizna leve	Bajo
5	Casas con jardines	Alta	Pequeña	Llovizna leve	Bajo
6	Casas con jardines	Alta	Pequeña	Aguacero	Alto
7	Zona verde	Alta	Pequeña	Aguacero	Bajo
8	Pavimento asfáltico	Media	Grande	Llovizna leve	Bajo
9	Pavimento asfáltico	Alta	Pequeña	Llovizna leve	Bajo
10	Casas con jardines	Media	Pequeña	Llovizna leve	Bajo
11	Pavimento asfáltico	Media	Pequeña	Aguacero	Alto
12	Zona verde	Media	Grande	Aguacero	Alto
13	Zona verde	Baja	Pequeña	Llovizna leve	Bajo
14	Casas con jardines	Media	Grande	Aguacero	Alto

Figura 1-6. Conjunto de patrones que describen el nivel del caudal de escorrentía en función de variables topográficas e hidrológicas.

A partir de la tabla de patrones (ver Figura 1-6) es posible generar el árbol de decisión de la Figura 1-7. Este árbol de decisión clasifica las variables que afectan el Caudal de Escorrentía definiendo si el nivel del caudal es bajo o alto. Se puede observar que el nodo inicial o raíz del árbol corresponde a una de las variables de entrada y será la que mejor clasifique los datos de entrenamiento. Los demás nodos serán las otras variables de entrada. Las ramas del árbol son las posibles etiquetas de la variable del nodo inmediatamente anterior y las hojas del árbol corresponden a los posibles valores que puede tomar la variable de salida.

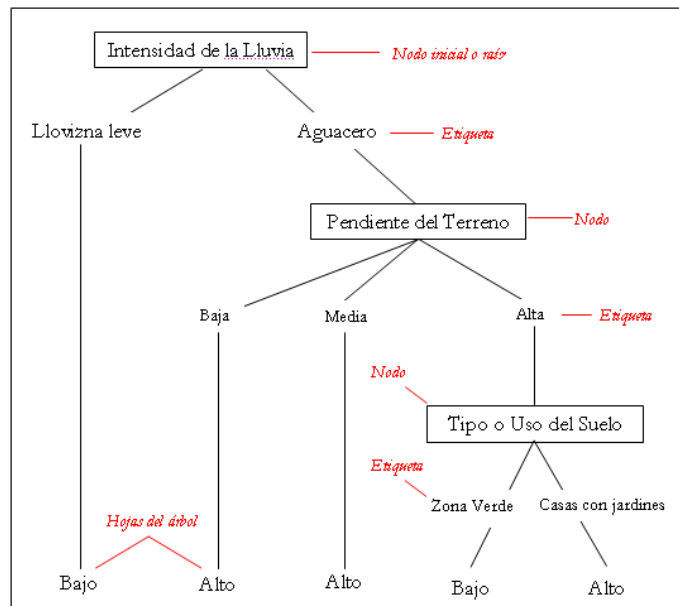


Figura 1-7. Árbol de decisión que define el nivel del caudal de escorrentía en función de variables topográficas e hidrológicas.

La construcción de los árboles de decisión comienza con la pregunta: ¿Qué variable de entrada debe ser la raíz del árbol? Con el fin de dar solución a esta incógnita cada variable de entrada es evaluada usando un examen estadístico para determinar qué tan bien esta variable clasifica los patrones de entrenamiento. La mejor variable de entrada se selecciona y usa como el nodo inicial del árbol. Entonces, se crea una rama descendiente del nodo inicial para cada posible etiqueta de esta variable y los patrones de entrenamiento son clasificados en los nodos descendientes. El proceso completo se repite usando los datos de entrenamiento asociados con cada nodo descendiente para seleccionar la mejor variable en este punto del árbol. Esto forma una búsqueda gradual por árboles de decisión aceptables, en el cual el algoritmo nunca retrocede para reconsiderar elecciones tempranas. (Mitchell, 1997). Existen diferentes criterios de selección o pruebas estadísticas para definir los nodos hijos en el proceso de construcción del árbol de decisión: Índice Gini, Entropía (Ganancia de información), Test Chi-cuadrado y proporción de Ganancia de Información

Comúnmente se utilizan dos de estos criterios: la ganancia de información y la entropía. La ganancia de información es una propiedad estadística que mide qué tan bien una variable dada separa los patrones de entrenamiento para seleccionar las variables candidatas en cada paso del crecimiento del árbol y se puede expresar de la siguiente forma:

$$G(M, A) = S(M) - \sum_{v \in \text{Val}(A)} \frac{|M_v|}{M} * S(M_v) \quad \text{Ecuación 1-1}$$

donde M es el conjunto de entrenamiento, A es el atributo a evaluar, $G(M,A)$ es la ganancia de información del atributo A sobre el conjunto M , $Val(A)$ son los posibles valores que puede asumir el atributo A , M_v es el subconjunto de M para el cual A toma el valor de v , $S(M)$ es la entropía total del conjunto de datos de entrenamiento y $S(M_v)$ es la entropía del subconjunto de datos de entrenamiento M_v . En resumen la ganancia de la información es la cantidad de información de un atributo brindada por conocimiento de otro atributo.

La entropía es una medida comúnmente usada en la teoría de la información que caracteriza la pureza o impureza de una colección arbitraria de patrones o en otras palabras es la medida de la cantidad de información necesaria para describir un sistema. Dada una colección S , la entropía de S se define como

$$Entropía(S) = \sum_{i=1}^c - p_i \log_2(p_i) \quad \text{Ecuación 1-2}$$

donde p_i es la proporción de S perteneciente a la clase i . El logaritmo es en base 2 porque la entropía es una medida de la longitud codificada en bits. (Mitchell, 1997).

Para el caso mostrado en la tabla de patrones de la Figura 1-6, primero se calcula la entropía de la información del sistema para cada una de las variables de entrada. La variable de entrada que proporcione la mayor ganancia de información al sistema será el nodo inicial. Para la variable *Caudal de Escorrentía* con etiqueta *bajo* y *alto* los valores 9 y 5 corresponden a la cantidad de patrones para los cuales los niveles del caudal cumplen las condiciones de bajo y alto respectivamente, por lo tanto la entropía es:

$$Entropía(S) = \sum_{i=1}^2 - p_i \log_2(p_i) = - \left[\frac{9}{14} \log_2 \left(\frac{9}{14} \right) + \frac{5}{14} \log_2 \left(\frac{5}{14} \right) \right] = 0.94 \quad \text{Ecuación 1-3}$$

La ganancia de información se calcula como la resta de la entropía del sistema y la entropía promedio de la variable de entrada. Cuando se presenta la condición Tipo o Uso del Suelo con la etiqueta *pavimento asfáltico*, existen 3 patrones para los cuales el nivel del *Caudal de Escorrentía* es *bajo* y 2 para los cuales es *alto*, entonces la ganancia de información tanto para el pavimento como para las otras dos etiquetas es:

$$Entropía_{\text{Pavimento asfáltico}}(S) = - \left[\frac{3}{5} \log_2 \left(\frac{3}{5} \right) + \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right] = 0.97$$

$$Entropía_{\text{Zona Verde}}(S) = - \left[\frac{3}{4} \log_2 \left(\frac{3}{4} \right) + \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] = 0.81$$

$$Entropía(S)_{\text{Casas con jardín}} = - \left[\frac{3}{5} \log_2 \left(\frac{3}{5} \right) + \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right] = 0.97$$

El promedio de la entropía:

$$\bar{E} = \frac{5 \cdot 0.97 + 4 \cdot 0.81 + 5 \cdot 0.97}{14} = 0.93$$

Por lo tanto la ganancia de información para estas variables es:

$$GI_{S|A} = 0.94 - 0.93 = 0,01bits$$

1.3 PRECIPITACIÓN

En la presente sección no se busca dar una descripción detallada de los fenómenos meteorológicos que contribuyen a la formación y ocurrencia de la precipitación, si no que se muestran los principales componentes que caracterizan la precipitación como un fenómeno complejo que muestra variaciones en el espacio y el tiempo.

1.3.1 Complejidad y Patrones Espaciales en la Precipitación

Para el análisis de la precipitación se puede partir de varias fuentes de información: la medición en estaciones en superficie o la realizada por percepción remota. Otra alternativa es contar con los resultados de modelos meteorológicos. Con la información recopilada se pueden observar las variaciones del campo de lluvia en el tiempo y el espacio. Para el análisis espacial (en el caso de estaciones puntuales en superficie) se generan mapas de isoyetas, con los cuales se observan los lugares de mayor o menor pluviosidad. En cuanto al análisis temporal, se busca entonces analizar la evolución de la precipitación en cada celda del mapa o la trayectoria que lleva el centro de tormenta a través del tiempo.

A partir de una red de estaciones de medición (entre las que se cuentan las pluviográficas), se pueden realizar estudios de la distribución espacial de la precipitación. Para ello se requiere extender la información desde los puntos de toma de datos a la totalidad del área de interés, para así disponer de un campo continuo. Por lo general este procedimiento se lleva a cabo mediante el uso de algoritmos de interpolación espacial como el *kriging* o las superficies de *splines* o el empleo de modelos estadísticos de correlación con otra variable conocida, como la altitud.

Una breve revisión de algunos de los últimos estudios de investigación de complejidad en los procesos hidrológicos se encuentra en Sivakumar (2000), donde se cita el caso de la precipitación, estudiado por: a) Rodríguez-Iturbe. et. al. (1989), mediante la estimación de correlaciones y el exponente de Lyapunov de los datos históricos, concluyen que existe una fuerte evidencia de la dinámica caótica de la lluvia; y b) Sivakumar et. al. (1998), analizan

registros de precipitación diaria y confirma que para este caso existe un comportamiento caótico en los datos de precipitación diaria.

En el caso específico de la precipitación en el trópico, Georgakakos et. al. (1994) hace énfasis en la extrema variabilidad de las precipitaciones sobre los océanos tropicales. Austin y Houze (1972), con mediciones de radar y estaciones en superficie, examinaron varias tormentas que se encontraban en diferente temporada (estación), escala y clasificación sinóptica, concluyendo la existencia de características de forma y comportamiento similar entre las tormentas y de esta manera se identificaban varios patrones de lluvia. El mismo análisis se realizó en varias escalas espaciales, tomando áreas de 10^3 Km^2 hasta 10 Km^2 . En este último caso también se encontraron patrones de intensidad, duración y movimiento de las tormentas. Zawadzki (1973) ha estudiado los patrones de precipitación con base a mediciones de radar, partiendo de funciones de autocorrelación de las funciones de espacio y tiempo, para lo cual que analizó datos estadísticos y la diferencia de la tasa de precipitaciones durante un tiempo secuencia de una lluvia convectiva.

Actualmente, desde la Universidad Nacional de Colombia se están desarrollando investigaciones en las que se ha demostrado la variabilidad de la precipitación para el caso específico de la ciudad de Bogotá (Rodríguez et. al. 2008a, Rodríguez et. al. 2008b y Jiménez et. al. 2008). De acuerdo con Rodríguez et. al. (2008a) el análisis de las curvas de masa para del evento de granizada ocurrido sobre Bogotá el 3 de noviembre de 2007 a diferentes escalas: a nivel de mesoescala (ciudad de Bogotá) y a nivel de microescala (red de estaciones del campus de la Universidad Nacional de Colombia) indican una importante variabilidad espacio-temporal no solo en términos de valores totales de precipitación, sino también en términos de intensidad a lo largo del evento.

También para el caso de Bogotá, Nauffal y Obregón (2008) analizaron el comportamiento de la dinámica espacio temporal de la precipitación para una tormenta de duración de 13 horas. En la Figura 1-8 se observan los mapas de isoyetas. Se observa cómo la tormenta se desplaza en diferentes direcciones y vuelve a concentrarse a los alrededores de donde inicialmente ésta estaba concentrada. Concluyen que es claro el papel que juegan los cerros orientales en la dinámica de vientos que desplazan las masas de agua.

Respecto a la trayectoria de los centros de tormenta, no existe un movimiento que indique una sola dirección predominante; por el contrario, la tormenta se mueve en varias direcciones (Figura 1-9).

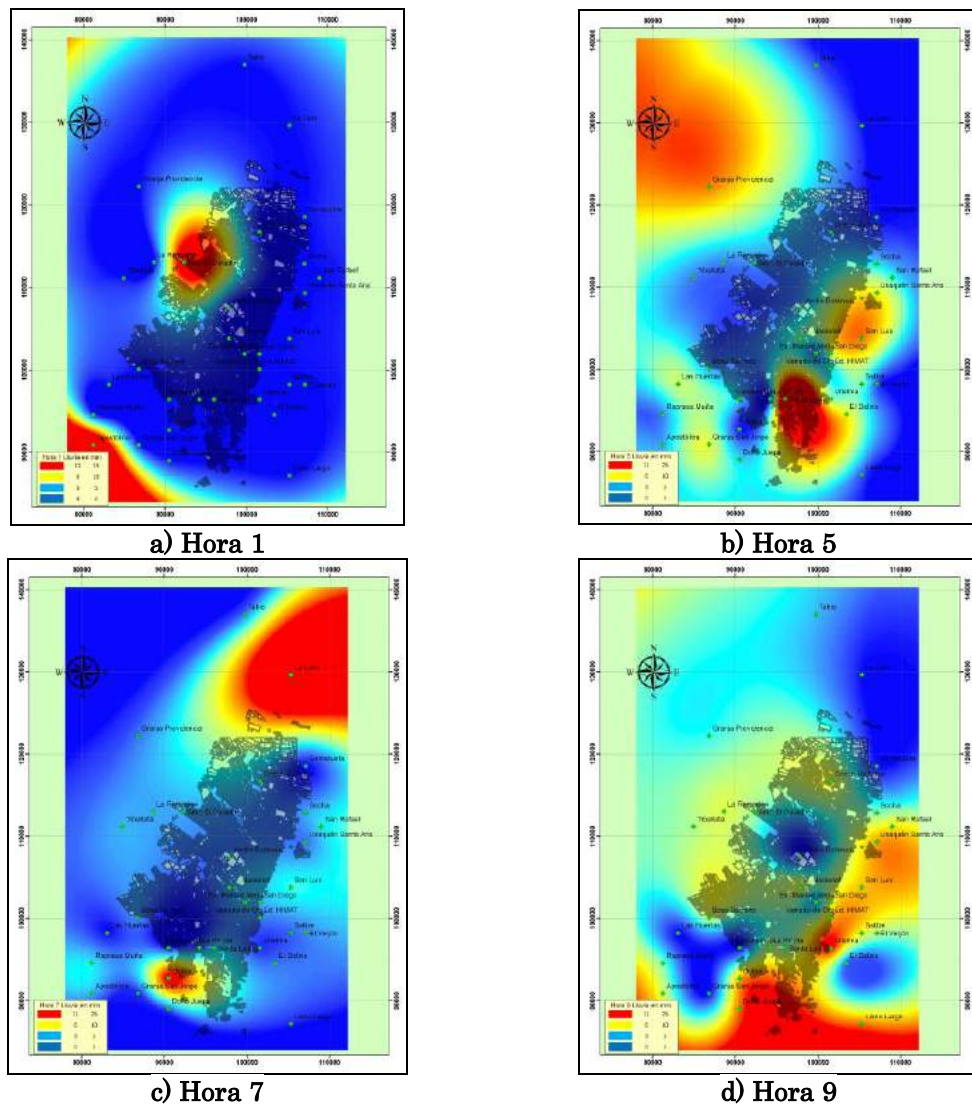
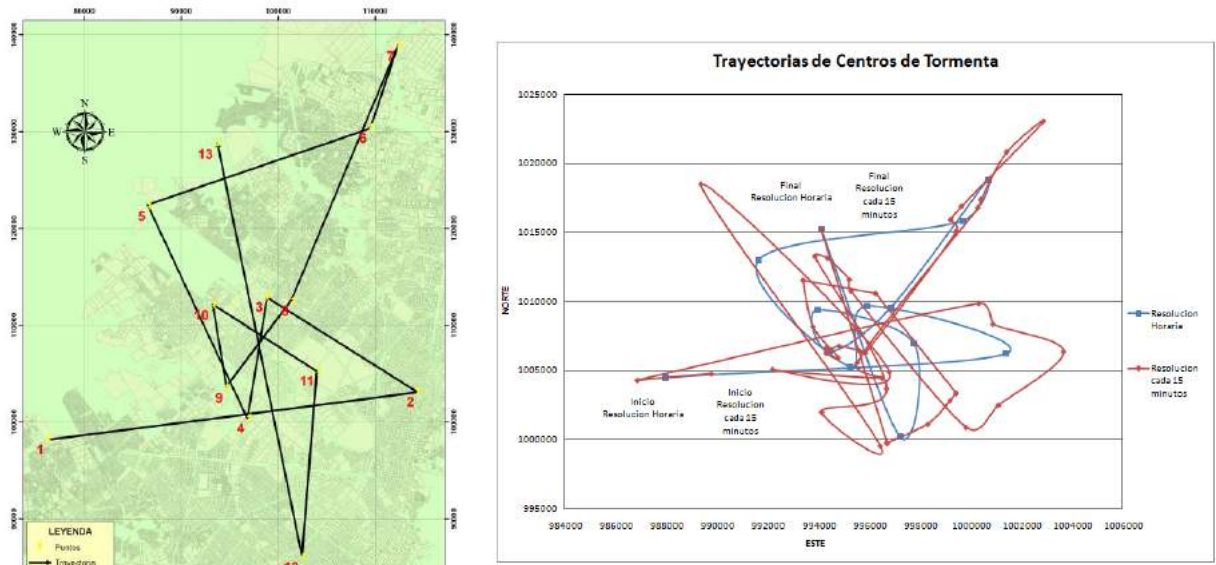


Figura 1-8. Mapas de isoyetas para la tormenta del 13 de enero de 1993 en la ciudad de Bogotá.
Fuente: Nauffal, M. S, Obregón, N. N, (2008)

Al realizar análisis como los citados anteriormente es común cuestionarse acerca de las resoluciones temporales y espaciales de estudio. Por ejemplo, la trayectoria de los centros de tormenta de la Figura 1-9 muestra la gran diferencia que hay entre la trayectoria de resolución horaria a la trayectoria de resolución cada 15 minutos. Por lo tanto se puede inferir que 1 hora no es un lapso de tiempo representativo cuando se trata de tormentas, ya que la duración de éstas en un punto determinado puede ser corta y su variabilidad espacial está fuertemente condicionada por la resolución temporal en la cual se desarrolla el análisis.



a) Trayectoria precipitación horaria

a) Trayectoria precipitación horaria y quinceminutal

Figura 1-9. Trayectoria de los centros de tormenta, para el evento registrado el 13 de enero de 1993 en la ciudad de Bogotá.

Nota: entiéndase centro de tormenta como el primer momento estadístico bidimensional.

Fuente: Nauffal, M. S, Obregón, N. N, (2008)

En cuanto a las resoluciones espaciales el problema se enfoca en la generación de los campos de precipitación cuando la fuente de información son las estaciones meteorológicas en superficie, donde la selección del método de interpolación espacial es de gran interés (Rodríguez *et. al.* 2008b).

1.3.2 Modelación de la Precipitación con AC

Debido a la naturaleza espacial de la precipitación, y todos los fenómenos meteorológicos y físicos que la producen, es útil la integración con los Sistemas de Información Geográfica (SIG). Ello facilita el manejo y análisis espacialmente de los datos asociado al modelo (Qi y Wu, 1996; De Koning et al., 1999). De acuerdo con la base teórica de AC ya expuesta, en la simulación del fenómeno de la precipitación con AC los componentes del modelo parten del mapa de interpolación espacial de la precipitación. Éste representa el campo de precipitación de un evento dado o de una serie de tiempo en el que se representa el total de la lámina de agua precipitada en un instante de tiempo, en otros términos, corresponde a una malla bidimensional, donde la altura o atributo de cada celda representa el estado actual de cada célula. De esta manera se tiene un mapa en el que se asocian tres electos básicos: 1) Valor de la precipitación, que corresponde al estado de la celda, 2) Coordenada de la celda (i,j) y 3) Tiempo en el que se produce el mapa.

El estado de cada celda depende de su estado actual y del estado de las células vecinas. En el proceso de simulación, a medida que el tiempo avanza el modelo de AC actualiza el estado de todas las células que componen el enmallado, debido a que el sistema está sincronizado. A diferencia del AC del Juego de la vida en la presente investigación se toman varios estados no solo el binario de ‘vivo’ o ‘muerto’ o el equivalente en la precipitación: ‘lluvia’ o ‘no lluvia’, sino que por el contrario se agrupan unos valores de precipitación para etiquetarlos y luego tener estados no dicotómicos, por ejemplo los estados de precipitación serán: no precipitación, precipitación baja, precipitación media y alta precipitación.

Partiendo de la premisa que los ACs representan la evolución de un fenómeno en el espacio y en el tiempo, involucrar en el proceso de modelación un SIG permite: a) La identificación de patrones espaciales en los datos, que representan las platillas del AC; b) La cuantificación de los cambios temporales observados o predichos, y; c) La visualización de los resultados.

2 CARACTERIZACIÓN DE LA ZONA DE ESTUDIO E INFORMACIÓN DISPONIBLE

Bogotá se encuentra localizada en el centro del país, entre los 4°28' a 4°47' Norte y 74° 06' a 74° 13' Este, con una altura promedio de 2600 msnm, donde sobresale la cordillera de los Andes localizada en el oriente de la ciudad.

Respecto al clima, Bogotá tiene un clima frío, subhúmedo y con tendencia a la sequía hacia el sur y sur oeste del área urbana; cuenta con un régimen de lluvias bimodal con épocas lluviosas en los meses de abril, mayo, octubre y noviembre. La precipitación media multianual varía espacialmente, con valores de aproximadamente 600 mm/año en las zonas sur occidental y centro occidental y valores de más de 1400 mm/año en el costado oriental. En la zona urbana la precipitación disminuye desde los cerros de oriente y sur oriente hacia el occidente y sur occidente, este último donde alcanza los valores más bajos. La mayor parte de la ciudad se encuentra en el intervalo de 700 y 1000 mm, pero al sur occidente en los sectores de Tunjuelito - Sierra Morena – Venecia y al sur de Bosa, la precipitación alcanza valores inferiores a 600 mm, constituyendo la zona más seca de la ciudad (Ingeocim Ltda & UPES. 1998, citado por Ortiz et. al. ,1995).

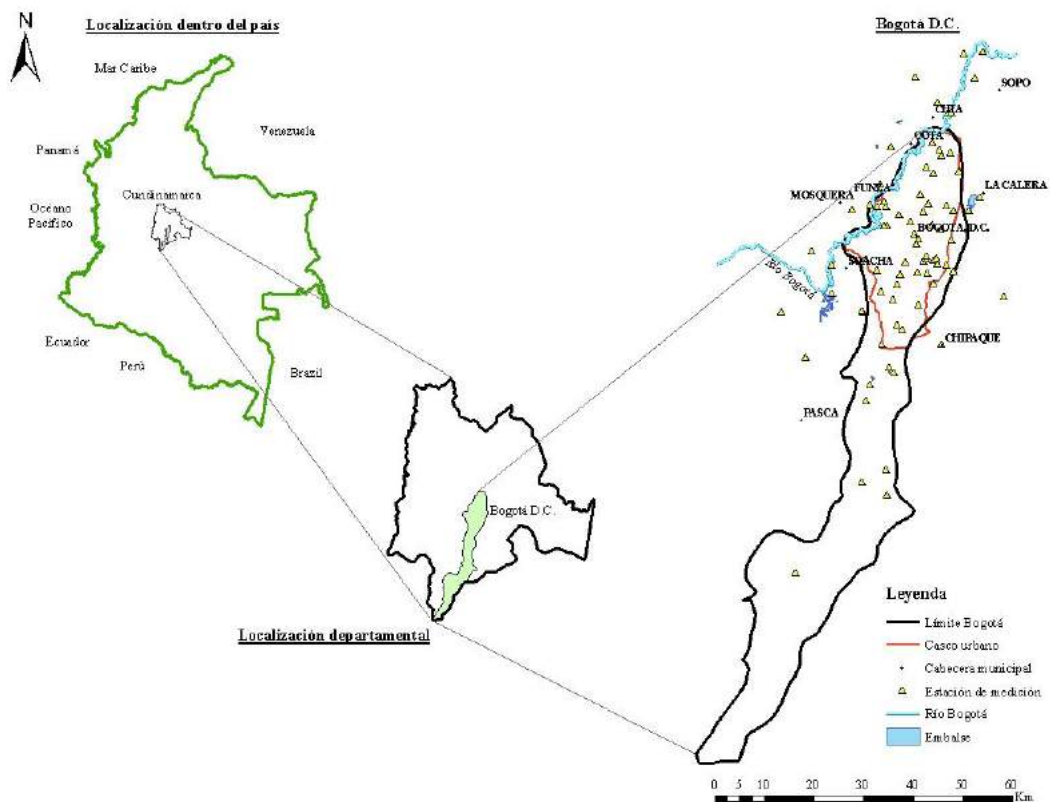


Figura 2-1. Localización de la zona de estudio

De manera general, los vientos se mueven en sentido oriente-occidente en horas de la mañana mientras que en la tarde los vientos predominantes viajan hacia el norte de la ciudad, con una velocidad media de $1 \text{ m}\cdot\text{s}^{-1}$. Durante el mes de agosto se presentan los vientos más fuertes en la ciudad. La temperatura media es de 14°C , en horas de la noche la ésta puede descender hasta los 4°C mientras que en el día se pueden registrar temperaturas de hasta 25°C (DAMA, 2006)

La altura y la temperatura afectan la humedad relativa y esta guarda una estrecha relación con la precipitación, ya que los valores más altos (83%) se presentan en los meses lluviosos y los más bajos (75%) en los meses secos. De igual forma ocurre con el brillo solar, el que alcanza valores de número de horas del brillo solar más bajos en los meses lluviosos y nublados en abril, mayo, junio, octubre y noviembre, entre 99 y 114 horas por mes (Ingeocim Ltda & UPES. 1998, citado por Ortiz et. al. ,1995).

La principal corriente hídrica que atraviesa la ciudad es el río Bogotá. Dentro del casco urbano se encuentran los ríos: Juan Amarillo, Fucha y Tunjuelo. Estas corrientes alimentan el conjunto de humedales comprendido por: Burro, Capellanía, Conejera y el Conejito, Córdoba, Guaymaral y Torca, Jaboque, Juan Amarillo, Santa María del Lago, Techo, Tibanica y Vaca.

2.1 INFORMACIÓN PLUVIOGRÁFICA DISPONIBLE

Se dispone con la información de precipitación a nivel diario y subhorario de estaciones de tipo meteorológica especial (ME), sinóptica principal (SP), pluviométrica (PM), climatológica principal (CP), pluviográfica (PG) y climatológica ordinaria (CO). En la Tabla 2-1 se muestran la cantidad de estaciones utilizadas para la modelación con resolución diaria y subhoraria. Los registros hacen parte de la base de datos de la Corporación Autónoma Regional (CAR), de la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB) y del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM).

El periodo de simulación fue elegido de acuerdo con la disponibilidad y calidad de la información por parte de las entidades, y por sugerencia de la revisión de las bases de datos de las cartas pluviográficas hechas por Bernal (2007). Para la precipitación diaria el período en estudio que comprende desde enero 01 de 1995 al 31 de diciembre de 1999 y para la precipitación horaria desde enero 01 de 1995 al 31 de diciembre de 1996.

De la Tabla 2-1 es claro verificar que la cobertura espacial de las estaciones supera la sugerida a nivel internacional, que esta alrededor de los 5 Km^2 por estación, ahora si para efectos de análisis de eventos de tormenta la cobertura oscila en los 22 Km^2 es de suma importancia que las redes de medición se densifiquen y se busque la manera de tecnificar la captura de la información y de esta manera disminuir considerablemente la incertidumbre espacial generada al estimar campos continuos de la precipitación.

Tabla 2-1. Resumen de estaciones con información de precipitación a nivel diario y subhoraria, para el periodo comprendido entre enero 1 de 1995 al 31 de diciembre de 1996, en la ciudad de Bogotá

Resolución temporal	Número de estaciones	Densidad de estaciones en el casco urbano
Diaria	CAR 18 EAAB 28 IDEAM 31 Total 77	12 Km ² /estación
Subhoraria	EAAB 27 IDEAM 10 Total 37	22 Km ² /estación

Nota: Área del casco urbano= 560,79 Km², estaciones con precipitación diaria= 47 y estaciones con precipitación subhoraria= 25.

El formato de adquisición de los registros de precipitación total diaria es en archivos planos de formato *.txt, mientras que para los registros subhorarios se partió de las cartas pluviográficas (ver Figura 2-2).

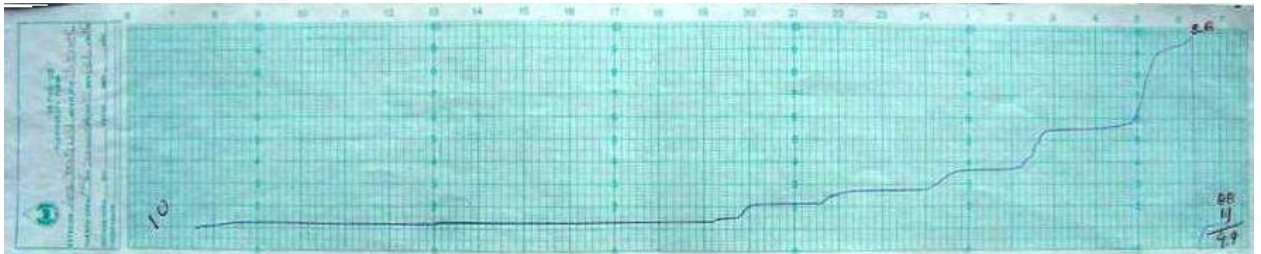


Figura 2-2. Carta pluviográfica de la estación LOS TAQUES

El procesamiento de la información se efectuó digitalizando la línea registrada por la pluma del pluviógrafo, marcando un punto de quiebre en los valores de cada cambio de pendiente del registro o cada 30 minutos, de acuerdo con el tipo de registro y la variabilidad en la forma de la línea. La lectura de los datos para los registros diarios se efectuó con una escala temporal mínima de cinco minutos, cuando la forma del registro o los cambios de pendiente lo sugerían conveniente. En el caso de los registros semanales se permitió la lectura hasta intervalos de 30 minutos. Esta operación se llevó a cabo para 10 estaciones operadas por el IDEAM. Las 27 estaciones de la EAAB ya fueron procesadas y facilitadas de la investigación desarrollada por Bernal (2008). Como resultado del proceso de digitalización se tienen los registros continuos de las 37 estaciones de precipitación con incrementos de tiempo de 30 minutos para el periodo 01-ene-1995 al 31-dic-1999.

La incertidumbre asociada con la información proveniente de los registros pluviográficos, así como las causas de error que se pudieron identificar son: 1) Falla en el registro (por tinta, falta de papel, falta de operario, falla en el reloj, falla en el sifón, poca visibilidad del registro, etc.); y 2) Ausencia de registros individuales (semanales o diarios).

Con el proceso de digitalización descrito anteriormente se procedió a estimar el registro continuo de la precipitación haciendo una interpolación lineal entre los puntos de quiebre de la línea del registro. A continuación se enlistan los nombres y entidades que operan las

estaciones con resolución subhoraria y diaria (Tabla 2-2 y Tabla 2-3 respectivamente), y en la Figura 2-3 y en la Figura 2-4 se observa su localización espacial.

Tabla 2-2. Listado de estaciones disponibles con registro de precipitación subhoraria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá.

ID	ENTIDAD	ESTACIÓN	CODIGO	CODIGO ENTIDAD	ESTE	NORTE	COTA
1	EAAB	ARRAYAN SAN FRANCISCO	2120008	P-29	1004500	999200	3047
2	EAAB	SAN RAFAEL	2120011	P-05	1011100	1012902	2750
3	EAAB	EL DELIRIO	2120013	P-35	1001802	995615	3000
4	EAAB	EL HATO	2120020	P-52	988998	975335	3150
5	EAAB	EL VERJON	2120024	P-37	1005995	998002	3250
6	EAAB	CERROS DE SUBA	2120031	P-01	999208	1013410	2691
7	EAAB	EL GRANIZO	2120032	P-36	1002516	1000781	3125
8	EAAB	SAN LUIS	2120040	P-41	1003254	1006379	2300
9	EAAB	SANTA LUCIA	2120052	P-42	995096	997414	2630
10	EAAB	CHIPAQUE	3502005	P-33	1003431	983518	2850
11	EAAB	USAQUEN SANTA ANA	2120111	P-09	1005805	1010244	2647
12	EAAB	BOSA BARRENO No 2	2120653	P-51	988257	1000417	2550
13	EAAB	SALITRE CASA DE BOMBAS	2120196	P-04	999780	1010008	2580
14	EAAB	CASABLANCA	2120197	P-31	990520	998273	2665
15	EAAB	SERREZUELA	2120202	P-78	1006941	1018259	2800
16	EAAB	QUIBA	2120205	P-90	991300	994082	3000
17	EAAB	LA CONEJERA	2120208	P-95	1000443	1018986	2500
18	EAAB	LAS HUERTAS	2120211	P-92	981377	999344	2572
19	EAAB	LA REGADERA	2120509	P-54	992841	978850	3050
20	EAAB	VITELMA	2120524	P-47	1000641	997681	2800
21	EAAB	LA CARO	2120531	P-79	1004903	1029918	2560
22	EAAB	TIBITOC	2120545	P-08	1011894	1042150	2698
23	EAAB	FONTIBON	2120547	P-48	992431	1007175	2518
24	EAAB	CAMAVIEJA	2120569	P-30	998447	1003528	2680
25	EAAB	PLANTA WIESNER	2120642	P-76	1009006	1010268	2795
26	EAAB	UNE	3502042	P-46	1004911	980136	2430
27	EAAB	SOCHA	2120199	P-74	1012307	1007795	2750
28	IDEAM	APTO ELDORADO P1-2	2120579	2120579	991883	1012096	2547
29	IDEAM	CASD	2120651	2120651	999732	1000105	2600
30	IDEAM	ED.IDEAM CRA 10	2120160	2120160	1000513	1001207	2685
31	IDEAM	ELDORADO DIDACTICA	2120552	2120552	992335	1011164	2546
32	IDEAM	ESC COL INGENIERIA	2120605	2120605	1003545	1021246	2650
33	IDEAM	INST MERANI- ABRA	2120650	2120650	1001767	1017742	2570
34	IDEAM	TAQUES LOS	3502507	3502507	987406	955820	3150
35	IDEAM	UNIV NACIONAL	2120622	2120622	998716	1004631	2556
36	IDEAM	VEGAS LAS HDA	2120623	2120623	991799	1007240	2543
37	IDEAM	VENADO ORO VIVERO	2120558	2120558	1001770	1000239	2725

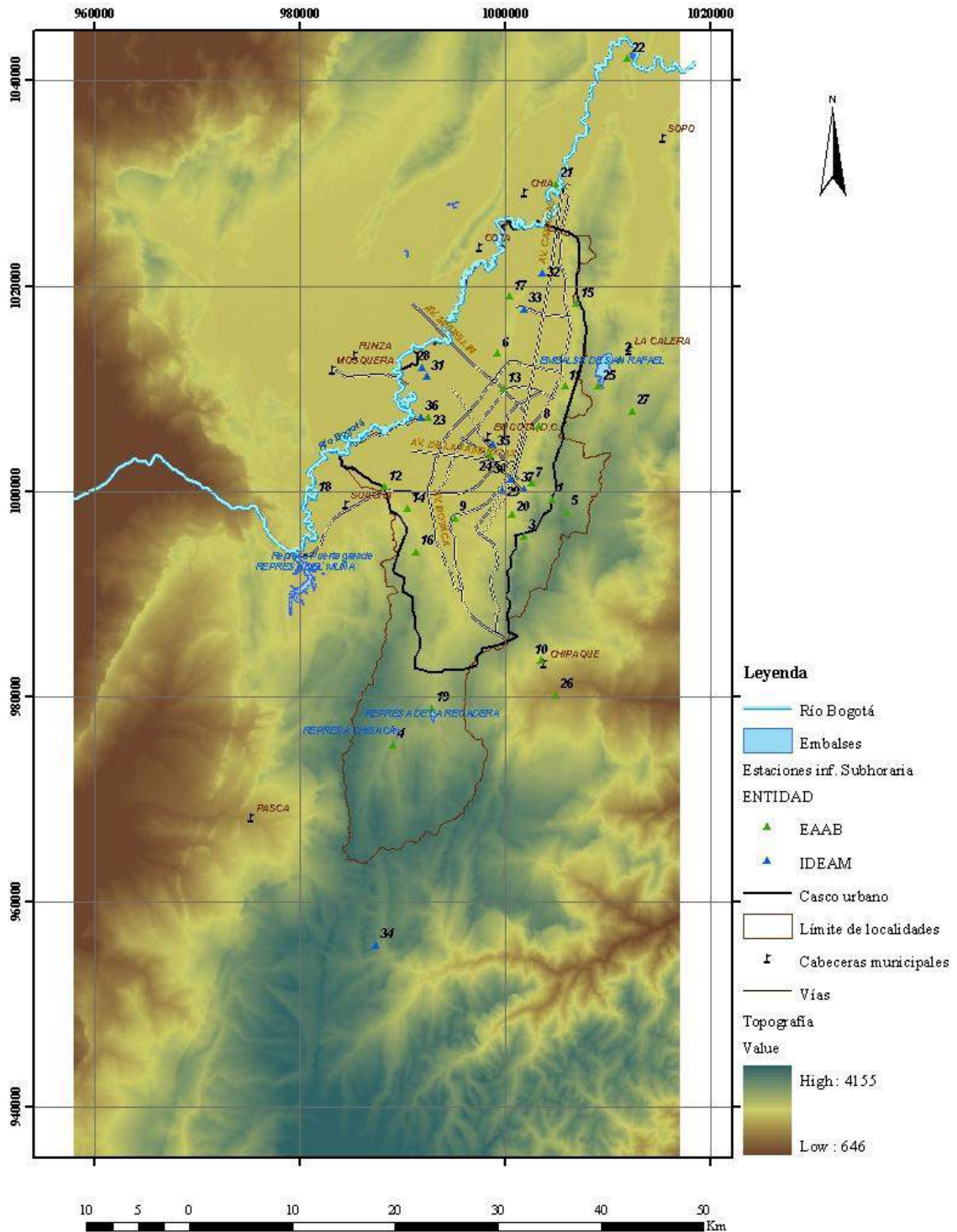


Figura 2-3. Estaciones disponibles con registro de precipitación subhoraria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá.

Tabla 2-3. Listado de estaciones disponibles con registro de precipitación diaria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá.

ID	ENTIDAD	ESTACION	CODIGO	CODIGO ENTIDAD	ESTE	NORTE	COTA
1	CAR	ALCO	2120159	2120159	1008000	1041700	2590
2	CAR	ALMAVIVA	2120113	2120113	1005600	1029965	2595
3	CAR	APOSTOLICA	2120051	2120051	971174	989870	2618
4	CAR	LA CASITA	2120112	2120112	1005400	1004320	3045
5	CAR	EL BOSQUE	2120085	2120085	995560	986200	2880
6	CAR	DOÑA JUANA	2120630	2120630	993700	992300	2700
7	CAR	EDIFICIO M. MEJIA	2120115	2120115	1000500	1000500	2580
8	CAR	EL FUTE	2120166	2120166	977280	1002150	2607
9	CAR	GUARANI EL PEÑON	2120060	2120060	976070	980870	2800
10	CAR	APTO GUAYMAR	2120559	2120559	1001550	1023950	2560
11	CAR	LA PICOTA	2120156	2120156	994600	995500	2580
12	CAR	EL MUÑA	2120561	2120561	981280	993600	2565
13	CAR	PARQUE SOPO	2120134	2120134	1010170	1036950	2540
14	CAR	LA RAMADA	2120516	2120516	989110	1011430	2545
15	CAR	SAN JORGE	2120172	2120172	987380	990160	2890
16	CAR	SANTA INES	2120136	2120136	993400	1023000	2550
17	CAR	VILLA PAULA	2120188	2120188	998200	1036980	2750
18	CAR	TORCA	2120077	2120077	1005320	1021800	2579
19	EAAB	ARRAYAN SAN FRANCISCO	2120008	P-29	1004500	999200	3047
20	EAAB	SAN RAFAEL	2120011	P-05	1011100	1012902	2750
21	EAAB	EL DELIRIO	2120013	P-35	1001802	995615	3000
22	EAAB	EL HATO	2120020	P-52	988998	975335	3150
23	EAAB	EL VERJON	2120024	P-37	1005995	998002	3250
24	EAAB	CERROS DE SUBA	2120031	P-01	999208	1013410	2691
25	EAAB	EL GRANIZO	2120032	P-36	1002516	1000781	3125
26	EAAB	SAN LUIS	2120040	P-41	1003254	1006379	2300
27	EAAB	SANTA LUCIA	2120052	P-42	995096	997414	2630
28	EAAB	CHIPAQUE	3502005	P-33	1003431	983518	2850
29	EAAB	USAQUEN SANTA ANA	2120111	P-09	1005805	1010244	2647
30	EAAB	BOSA BARRENO No 2	2120653	P-51	988257	1000417	2550
31	EAAB	SALITRE CASA DE BOMBAS	2120196	P-04	999780	1010008	2580
32	EAAB	CASABLANCA	2120197	P-31	990520	998273	2665
33	EAAB	GUADALUPE	2120198	P-49	1002583	999527	3316
34	EAAB	SERREZUELA	2120202	P-78	1006941	1018259	2800
35	EAAB	JUAN REY	2120204	P-81	998762	991429	2985
36	EAAB	QUIBA	2120205	P-90	991300	994082	3000
37	EAAB	LA CONEJERA	2120208	P-95	1000443	1018986	2500
38	EAAB	LAS HUERTAS	2120211	P-92	981377	999344	2572
39	EAAB	CHOACHI - EL UVAL	2120053	P-34	1016011	993046	1950
40	EAAB	LA REGADERA	2120509	P-54	992841	978850	3050
41	EAAB	VITELMA	2120524	P-47	1000641	997681	2800
42	EAAB	LA CARO	2120531	P-79	1004903	1029918	2560
43	EAAB	TIBITOC	2120545	P-08	1011894	1042150	2698
44	EAAB	FONTIBON	2120547	P-48	992431	1007175	2518
45	EAAB	CAMAVIEJA	2120569	P-30	998447	1003528	2680
46	EAAB	PLANTA WIESNER	2120642	P-76	1009006	1010268	2795
47	IDEAM	ADPOSTAL	2120600	2120600	994882	1009350	2550
48	IDEAM	APTO ELDORADO P1-2	2120579	2120579	991883	1012096	2547
49	IDEAM	AUSTRALIA	2120130	2120130	993952	977668	3050

ID	ENTIDAD	ESTACION	CODIGO	CODIGO_ENTIDAD	ESTE	NORTE	COTA
50	IDEAM	BETANIA	3502035	3502035	992300	958277	3150
51	IDEAM	CASD	2120651	2120651	999732	1000105	2600
52	IDEAM	CENTRO MED ANDES	2120573	2120573	1004513	1011276	2575
53	IDEAM	CLINICA SAN RAFAEL	2120576	2120576	998554	997786	2600
54	IDEAM	ED.IDEAM CRA 10	2120160	2120160	1000513	1001207	2685
55	IDEAM	ELDORADO DIDACTICA	2120552	2120552	992335	1011164	2546
56	IDEAM	ENMANUEL D ALZON	2120123	2120123	1000799	1011603	2520
57	IDEAM	ESAP	2120604	2120604	997908	1005593	2553
58	IDEAM	ESC COL INGENIERIA	2120605	2120605	1003545	1021246	2650
59	IDEAM	ESC LA UNION	2120120	2120120	988191	971996	3320
60	IDEAM	FLORES COLOMBIANAS	2120621	2120621	991144	1015487	2560
61	IDEAM	IDEAM FONTIBON HB	2120210	2120210	990486	1011164	2511
62	IDEAM	INST MERANI- ABRA	2120650	2120650	1001767	1017742	2570
63	IDEAM	JARDIN BOTANICO	2120571	2120571	997209	1008087	2552
64	IDEAM	NAZARETH	3502031	3502031	992396	953120	2800
65	IDEAM	PASQUILLA	2120158	2120158	991418	983446	3000
66	IDEAM	SAN JUAN	2119027	2119027	974053	937507	2900
67	IDEAM	SENA K 30	2120584	2120584	996191	999907	2553
68	IDEAM	STA MARIA DE USME	2120124	2120124	994587	987295	2800
69	IDEAM	TAQUES LOS	3502507	3502507	987406	955820	3150
70	IDEAM	UNIV NACIONAL	2120622	2120622	998716	1004631	2556
71	IDEAM	UNIV PEDAGOGICA NA	2120619	2120619	1001581	1007478	2570
72	IDEAM	VEGAS LAS HDA	2120623	2120623	991799	1007240	2543
73	IDEAM	VENADO ORO VIVERO	2120558	2120558	1001770	1000239	2725
74	IDEAM	TIBAITATA	2120542	2120542	985411	1010531	2543
75	IDEAM	SAN JORGE GJA	2120572	2120572	987596	989999	2900
76	IDEAM	GUANATA	2120589	2120589	1002570	1032040	2550
77	IDEAM	CORP UNIV AGROPEC	2120626	2120626	1003082	1022386	2570

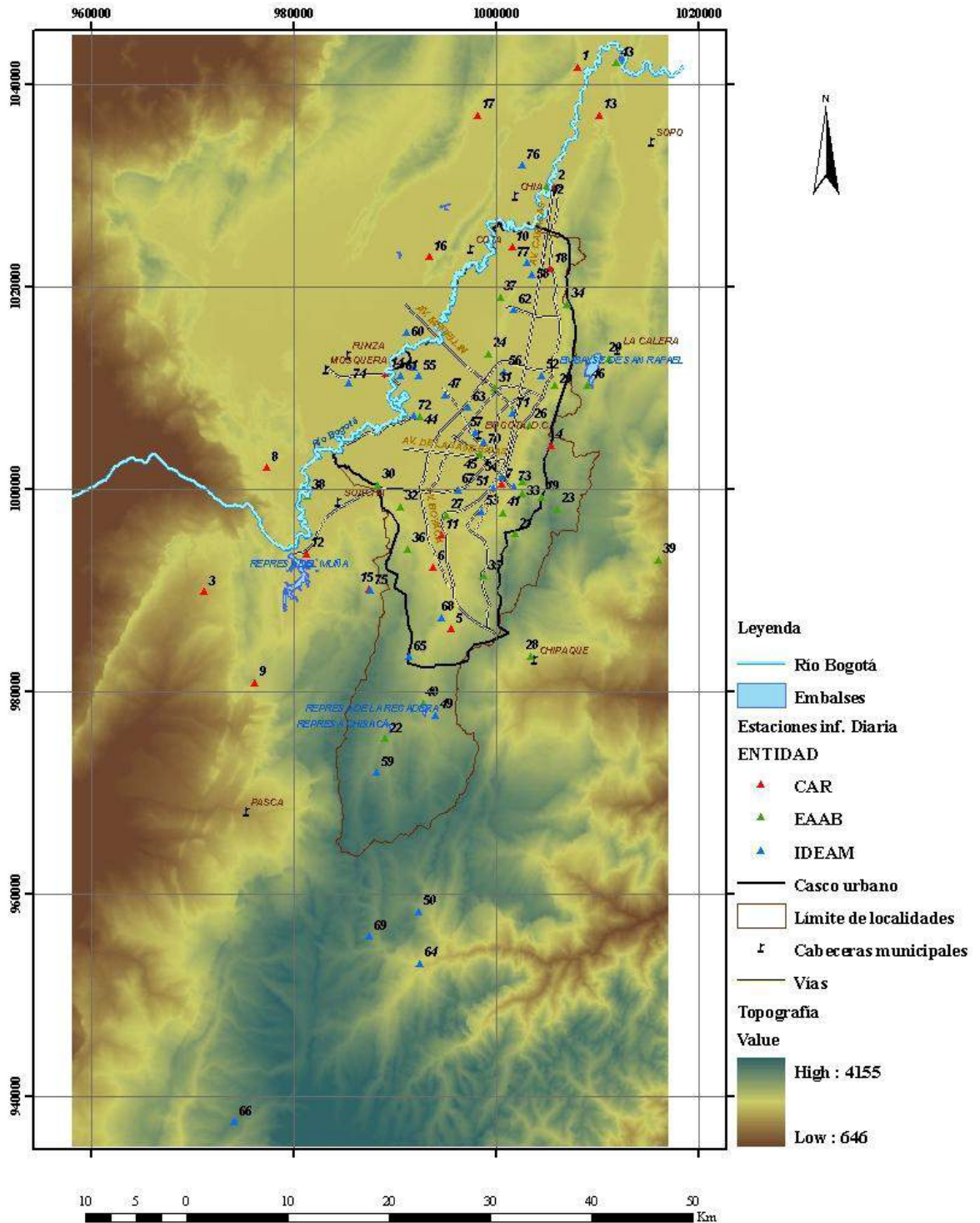


Figura 2-4. Estaciones disponibles con registro de precipitación diaria, para el período comprendido entre el 01-01-1995 al 31-12-1996. Ciudad de Bogotá.

2.2 INFORMACIÓN DE VIENTO

Con el fin de identificar las principales fuentes de contaminación atmosférica en la ciudad de Bogotá, la secretaría distrital del ambiente cuenta con estaciones de monitoreo localizadas en diferentes puntos de la ciudad. Estas estaciones pertenecen a la “Red de monitoreo de Calidad del Aire (RCA) y miden de forma continua las concentraciones de los principales contaminantes y en algunos casos variables meteorológicas como la dirección y magnitud de los vientos. Tiene una resolución temporal de 1 hora o de cada 10 minutos. Dentro del casco urbano se cuentan con 9 estaciones (ver Tabla 2-4 y Figura 2-5).

Con el fin de encontrar patrones en el comportamiento de los vientos en Bogotá que contribuya al entendimiento de la precipitación, se estudiaron los registros de las estaciones citadas anteriormente. El período disponible para los registros de la información de viento para la presente investigación está comprendido desde el año 1997 a 2002 con valores de dirección y magnitud de viento cada 10 minutos. A pesar que el período de registro de las estaciones de viento no coincide con el de precipitación, la información se utilizó para verificar si existía algún comportamiento predominante en los patrones de viento, y de esta manera sugerir una alineación para la malla o *lattice* del autómatas celulares.

Tabla 2-4. Listado de estaciones disponibles con registro de dirección y magnitud de viento, para el período comprendido entre el 01-09-1997 al 01-01-2002. Ciudad de Bogotá.

ID	NOMBRE	LOCALIDAD	ZONA	ESTE	NORTE
1	BOSQUE	USAQUÉN	NORTE	1005075	1012524
2	SONY	KENNEDY	SUROCCIDENTE	992102	999967
3	CARREFOUR	ENGATIVA	NOROCCIDENTE	999439	1010457
4	CAZUCA	CIUDAD BOLIVA	SUROCCIDENTE	988095	981586
5	ESCUELA	SUBA	NORTE	1003676	1020778
6	SANTO TOMAS	CHAPINERO	CENTRO	1002351	1006678
7	CORPAS	SUBA	NOROCCIDENTE	998228	1018363
8	PUENTE ARANDA	PUENTE ARANDA	CENTROCCIDENTE	995570	1003939
9	FONTIBON	FONTIBON	NOROCCIDENTE	992883	1008187

En la Figura 2-6 se visualiza la totalidad de la distribución de la red pluviográfica y de estaciones de viento en la zona de estudio.

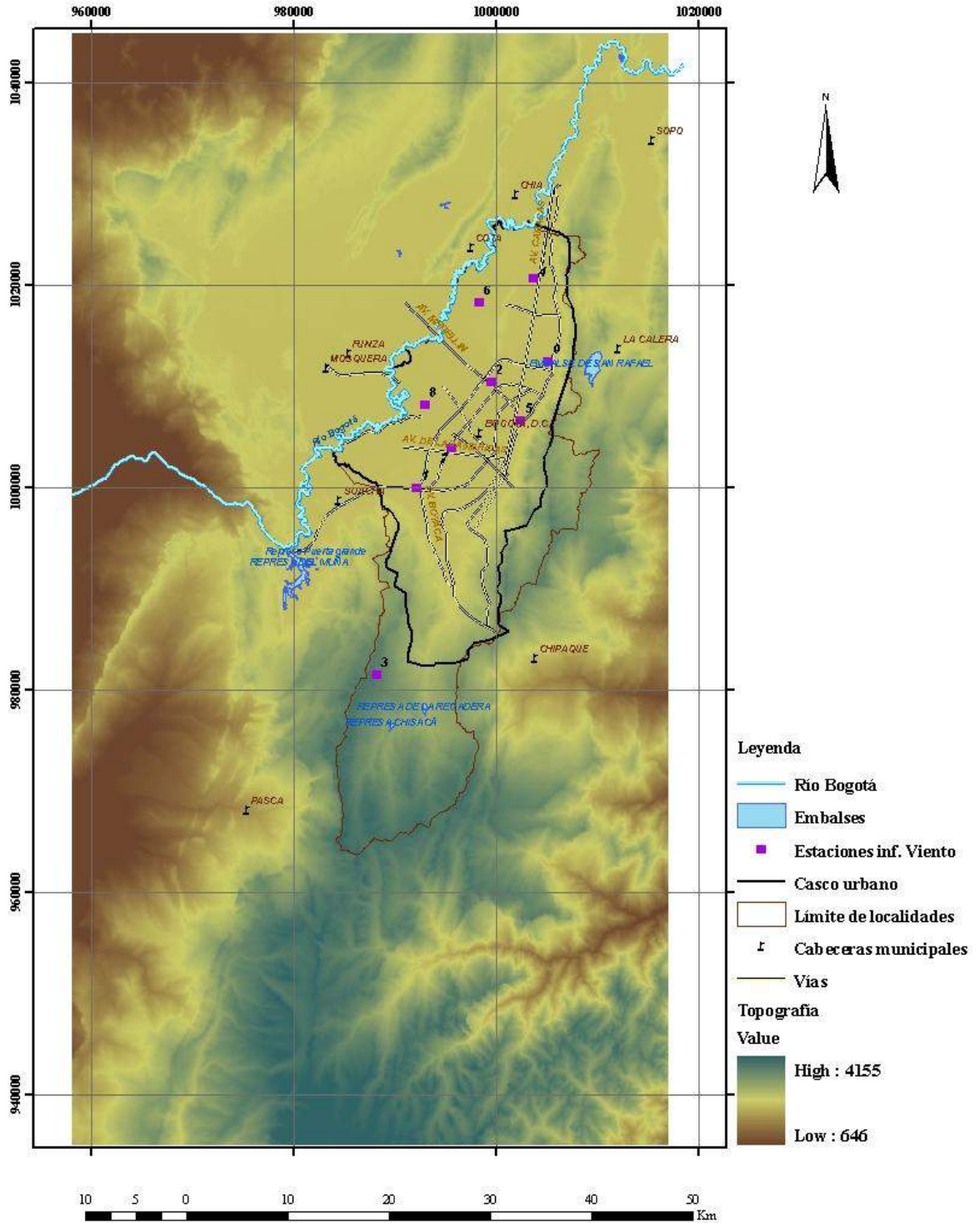
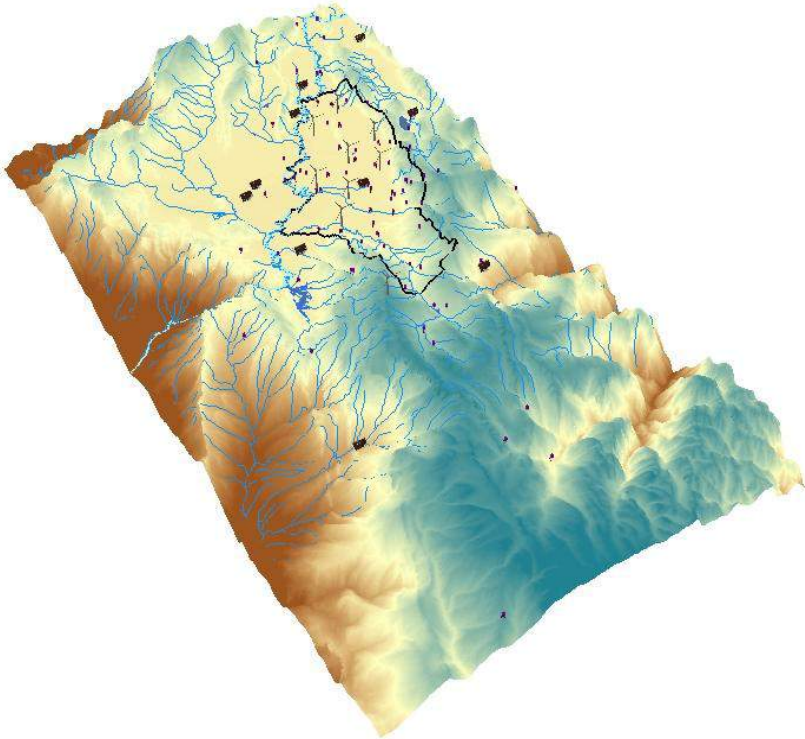
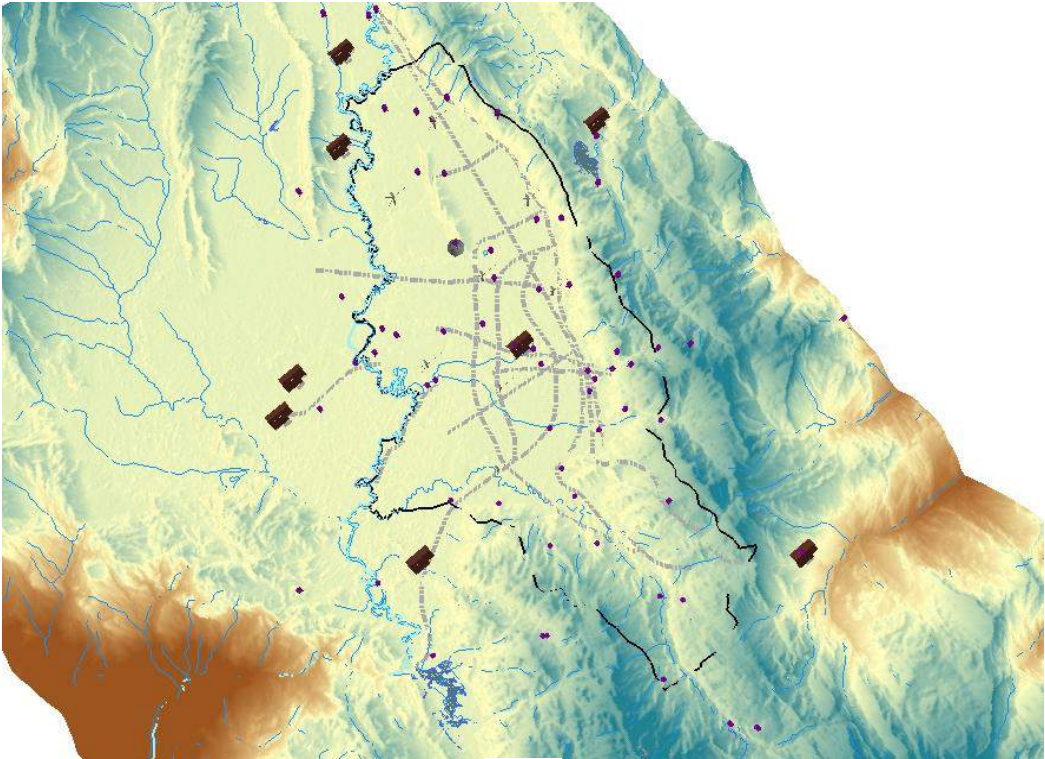


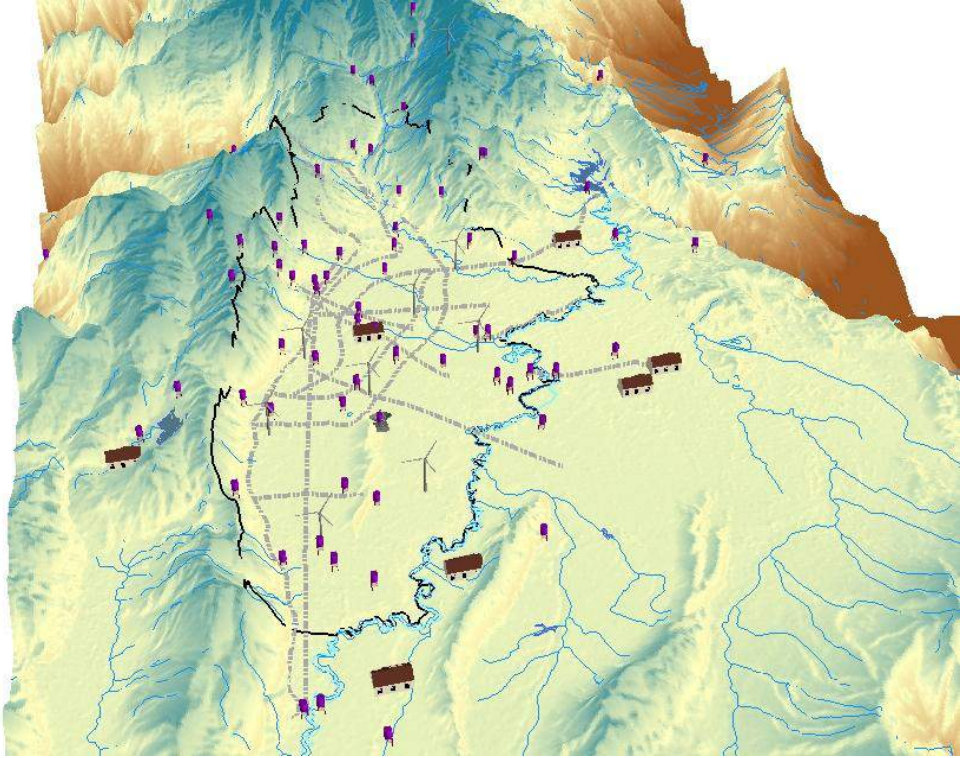
Figura 2-5. Estaciones disponibles con registro de viento con resolución subhoraria, para el período comprendido entre el 01-09-1997 al 01-01-2002. Ciudad de Bogotá.



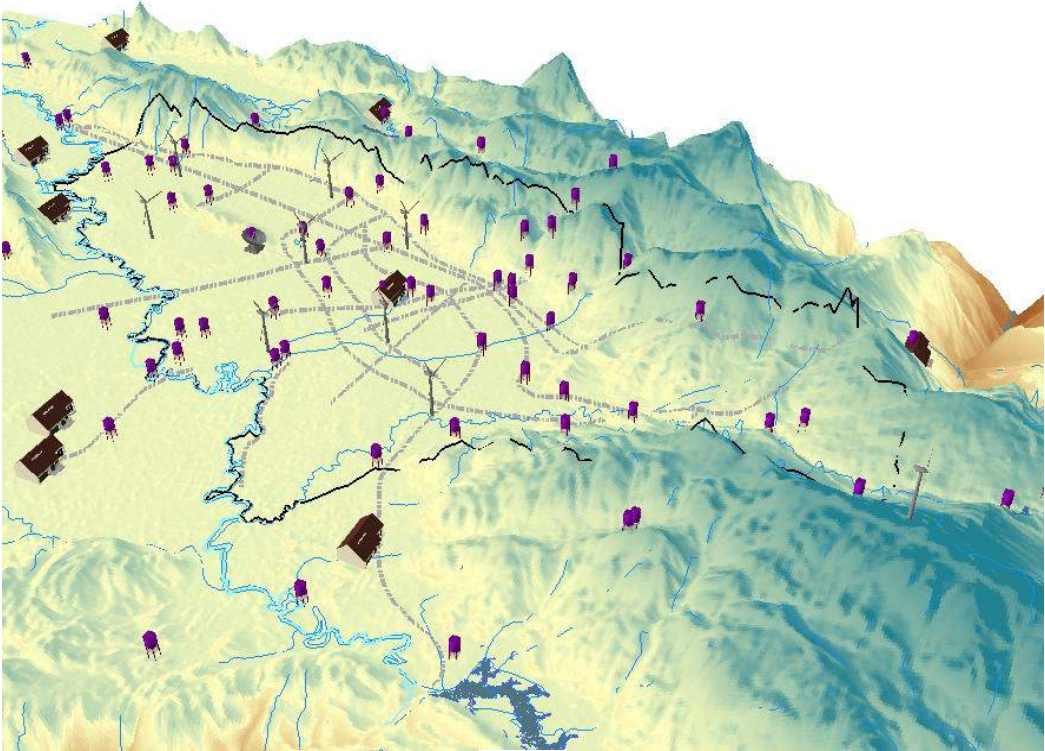
a) Sur-Norte



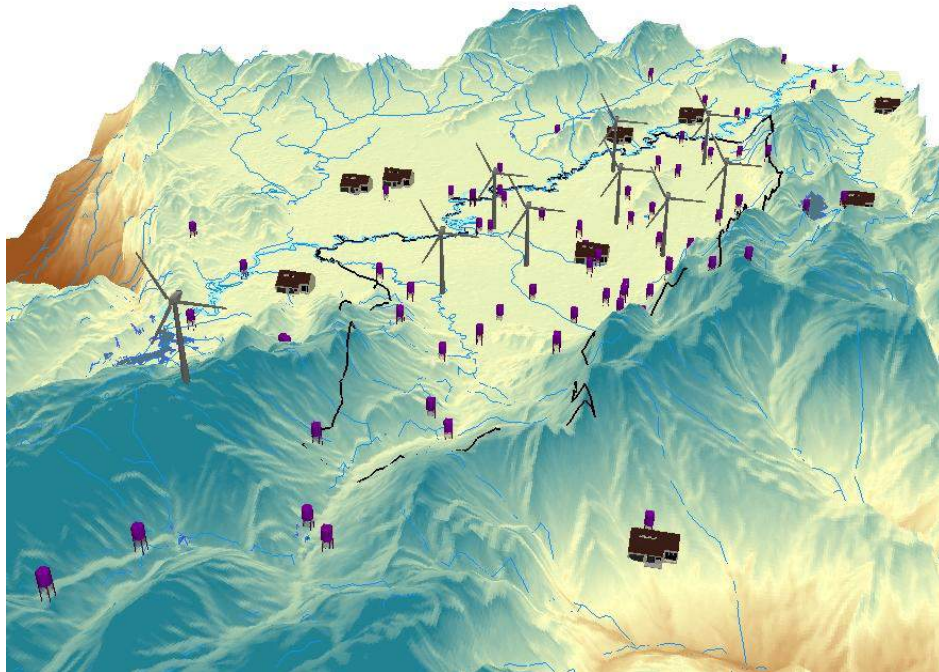
b) Planta



c) Sur



d) Oriente



e) Noroccidente



f) Occidente

Figura 2-6. Localización de las estaciones con información pluviométrica y dirección de viento. Ciudad de Bogotá.

Nota: En la figura se visualizan las estaciones con información de precipitación diaria (cilindros color morado), dirección y velocidad del viento (molinos color gris), cabeceras municipales (edificaciones en forma de casa), ríos y embalses. La topografía se tomó a partir del DEM de 30 metros.

2.3 PRE-PROCESAMIENTO DE LOS DATOS

Con el fin de caracterizar estadística y preliminarmente las series de tiempo de precipitación, se realizó un análisis exploratorio de los datos mediante el desarrollo y presentación de simples gráficas y evaluaciones numéricas, e.j., diagramas de cajas, funciones de correlación cruzada, funciones de autocorrelación que permitan descubrir algunas propiedades importantes de la información disponible. De igual forma se realizó el análisis exploratorio de datos espaciales con el fin de encontrar relaciones o influencia entre estaciones. A continuación se describen las herramientas gráficas empleadas para cada uno de los dos análisis.

2.3.1 *Análisis Exploratorio de Datos (AED)*

De manera individual, a los registros de cada una de las estaciones pluviográficas seleccionadas se les realizó un Análisis Exploratorio de Datos (AED), con el fin de identificar la consistencia y calidad de la información. El AED consiste en organizar y preparar los datos, detectar fallas en la toma o almacenamiento, tratamiento y evaluación de datos ausentes (missing), la identificación de casos atípicos (outliers).

Cada técnica utilizada en el AED posee sus propios atributos que son útiles para descubrir las características de algunos datos. Debido a que ningún método puede representar con claridad todos los componentes de los datos, es conveniente examinar la serie de tiempo mediante el uso de una serie de herramientas gráficas y numéricas útiles de investigación. A continuación se describen las principales de estas herramientas con ejemplos de algunas de las estaciones con las que se trabajaron en el presente estudio. En los apéndices se mostraran los resúmenes del AED de cada una de las 77 estaciones con registros diarios y las de registro horario.

2.3.1.1 *Gráfica de la Serie de Tiempo*

Es de gran interés observar en las series de tiempo el comportamiento de la variable de estudio en el tiempo. El grafico de series de tiempo reporta los datos cronológicamente de manera que el observador puede identificar características de los datos como son: la detección de valores extremos, texturas, tendencias, dependencias entre observaciones y la estacionalidad de los datos (Hipel y Ian McLeod, 1994).

En la Figura 2-7 se muestra la serie de tiempo de precipitación total diaria de la estación ALCO para el período comprendido entre el 01-01-1995 al 31-12-1999. Se observa que en esa estación el año en el que se presentaron mayores precipitaciones fue 1998 y que estas superan los 35 mm.

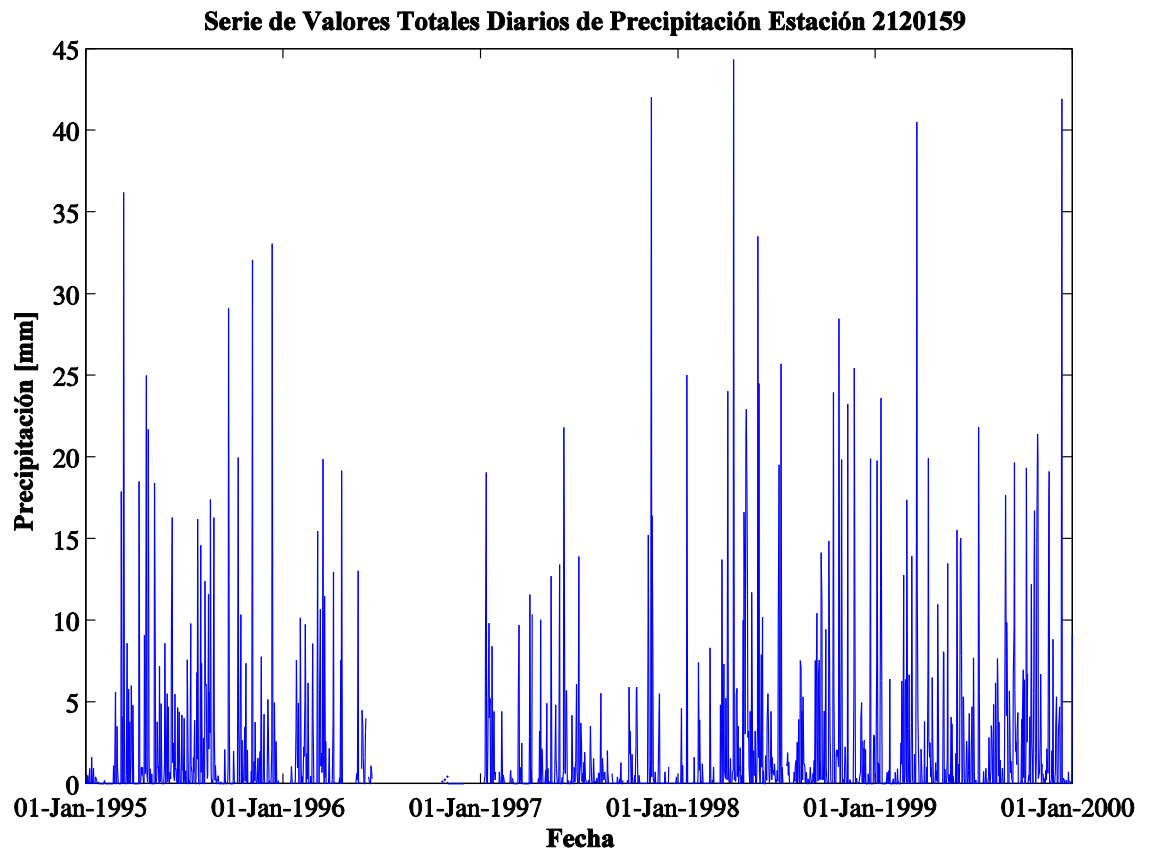


Figura 2-7. Serie de tiempo de registros de precipitación diaria de la estación pluviográfica ALCO para el período comprendido entre 01-01-1995 y 31-12-1999.

2.3.1.2 Diagrama de Cajas o Gráficos de Cuantiles

El diagrama de caja está basado en lo que es llamado el “resumen de cinco números” (Tukey, 1997). Para un conjunto de datos, el resumen de cinco números está compuesto de: 1) los valores más pequeños, 2) los valores más grandes, 3) la mediana, 4) cuantil 0.25 y 5) cuantil 0.75.

Para calcular los valores de los cuantiles, es conveniente emplear la definición de cuantiles dada por Chambers et al. (1983). En un grupo de datos dados, que son representados por x_i para $i = 1, 2, \dots, n$, son organizados de menor a mayor y denotados por $x_{(i)}$ para $i = 1, 2, \dots, n$. Si p representa cualquier fracción entre 0 y 1, el correspondiente cuantil es representado por $Q(p)$, donde cualquier p es una de las fracciones definidas por:

$$p_i = (i - 0.5) / n \quad \text{para } i = 1, 2, \dots, n \quad \text{Ecuación 2-1}$$

La mediana, la cual es dada por $Q(0.5)$ divide los datos en dos grupos de igual tamaño. Si n es impar la mediana es $x_{((n+1)/2)}$. Cuando n es par, $Q(0.5)$ es calculado como el promedio entre $x_{(n/2)}$ y $x_{((n+1)/2)}$ los cuales son los dos valores ordenados más cercanos al punto medio entre los datos.

Los cuantiles inferior y superior definidos como $Q(0.25)$ y $Q(0.75)$ respectivamente, y en ocasiones son llamados articulaciones. La distancia entre el primer y el tercer cuartil, esta dado por $Q(0.75) - Q(0.25)$, es denominado rango intercuartil el cual es empleado para medir la dispersión de los datos. Así mismo, para caracterizar los valores extremos, Tukey (1977) ha sugerido un paso de 1.5 veces del rango intercuartil, por fuera de las articulaciones. La Figura 2-8 describe un ejemplo de diagrama de cajas para registros de precipitación diaria para cada mes, durante cinco años en la estación ALCO.

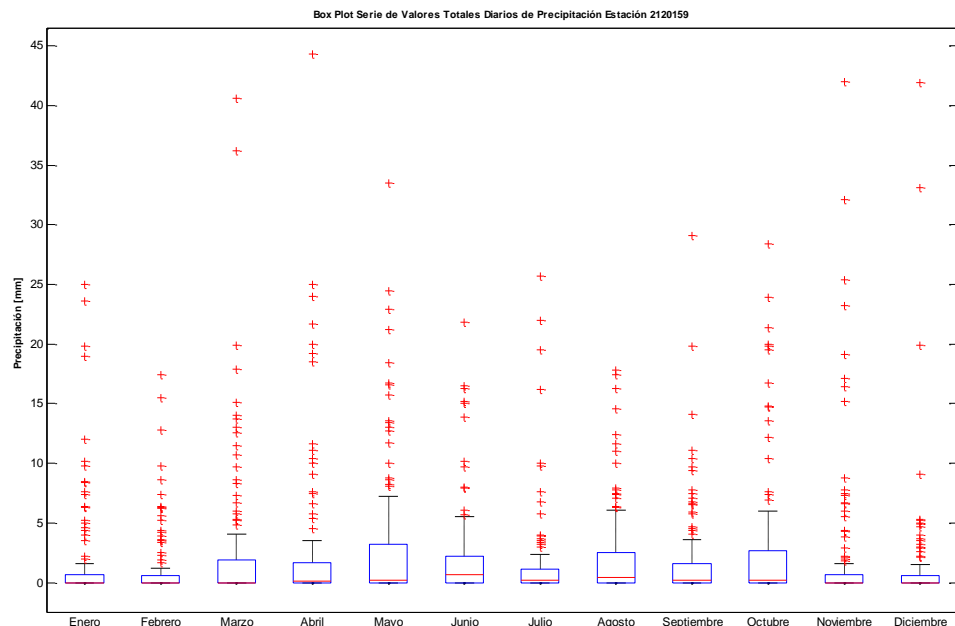


Figura 2-8. Diagrama de cajas para el análisis de registros de precipitación diaria para la estación ALCO. Período de registro 01-01-1995 y 31-12-1999.

2.3.1.3 Función de Auto-Correlación

La función de auto-correlación mide la relación existente entre los valores de la serie temporal discreta de un proceso y los correspondientes a la misma serie rezagada k unidades de tiempo o desfase. Esta función es de gran utilidad para encontrar patrones repetitivos dentro de una señal. Para estimar el coeficiente de auto-correlación entre la serie

original y la misma serie pero rezagadas k unidades de tiempo, se usa la siguiente expresión:

$$r_k = \frac{\frac{1}{n-k} \left[\sum_{t=1}^{n-k} (x_t - \bar{x}) \cdot (x_{t+k} - \bar{x}) \right]}{\frac{1}{n} \left[\sum_{t=1}^n (x_t - \bar{x})^2 \right]}, k > 0 \quad \text{Ecuación 2-2}$$

Donde \bar{x} es la media estimada de la serie. Dado que la función de autocorrelación es simétrica respecto al rezago cero, únicamente se dibujan los rezagos positivos hasta un valor máximo de rezago igual a $n/4$ (Hipel y Ian McLeod, 1994). Cuando la función tiene un valor de cero y además la serie es ruido blanco, r_k es asintótico independiente normalmente distribuido con una media de cero y varianza de $1/n$.

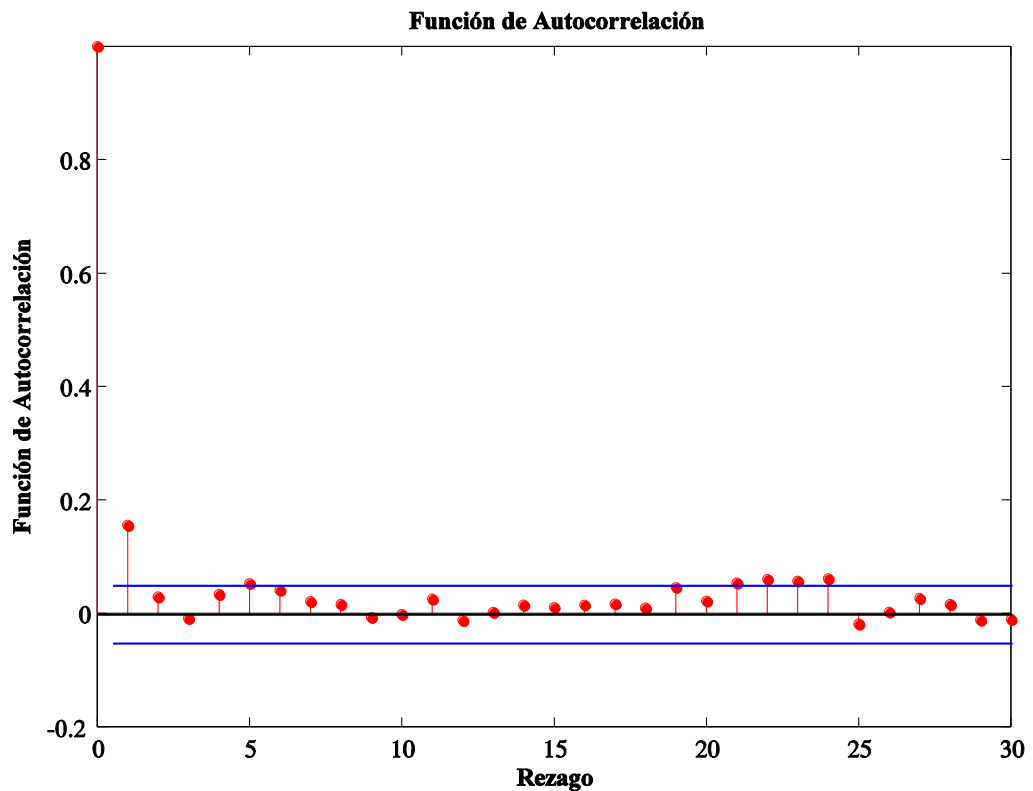


Figura 2-9. Función de autocorrelación para el análisis de registros de precipitación diaria para la estación ALCO. Período de registro 01-01-1995 y 31-12-1999.

2.3.1.4 Análisis Espectral

El espectro de potencia es la transformada de Fourier de la función de auto-covarianza (Jenkins y Watts, 1968). El espectro no provee una información diferente a la contenida en la función de autocovarianza (o equivalentemente la ACF). Sin embargo, ofrece una interpretación diferente de las propiedades estadísticas de la serie de tiempo desde que se exprese la distribución de la varianza de la serie en un dominio de frecuencias, dentro de un rango entre 0 y $\frac{1}{2}$ de acuerdo a las sugerencias de Jenkins y Watts (1968). Un análisis espectral puede suministrar información importante de la periodicidad de un proceso físico observado.

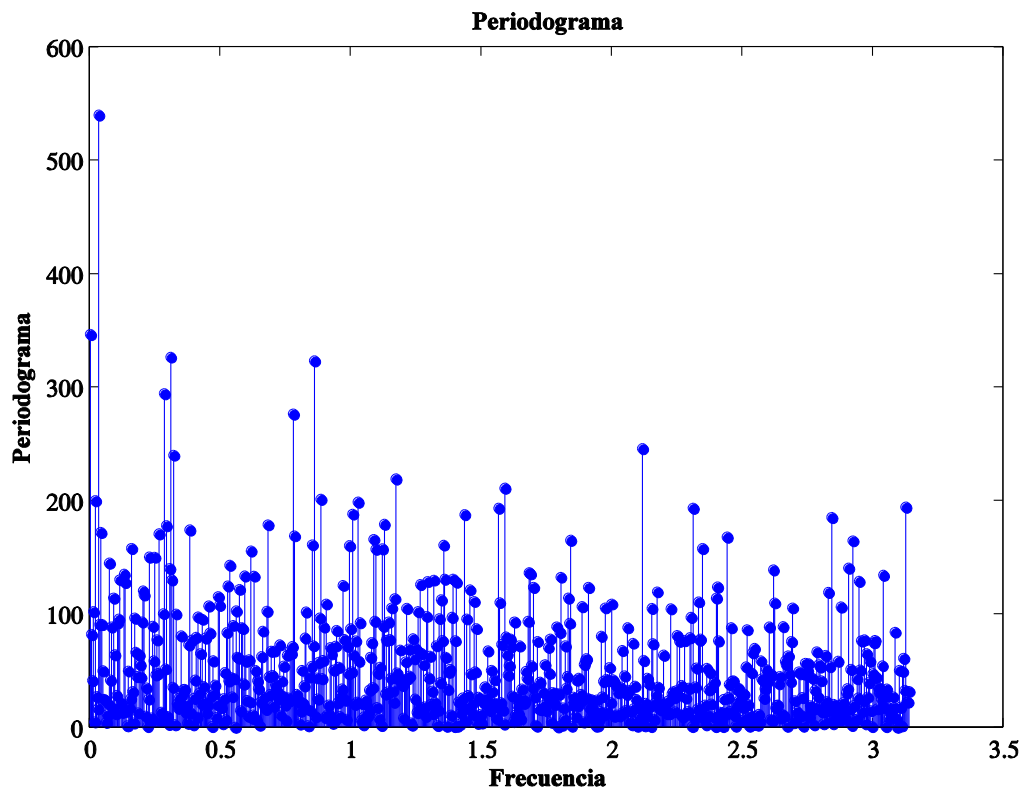


Figura 2-10. Espectro de potencia de registros de precipitación diaria de la estación pluviográfica ALCO para el periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.

2.3.1.5 Histograma de Frecuencias

El histograma es una gráfica que describe la distribución univariada de un conjunto de datos y permite observar la dispersión, el sesgo y el centro de los datos de estudio. Esta gráfica se construye subdividiendo el conjunto de datos en la mayoría de los casos de igual extensión llamados clases para los cuales se cuantifica el número de elementos integrantes (frecuencias). Este gráfico permite al investigador una visión inmediata de la amplitud de los datos, los valores que más se repiten o de mayor frecuencia absoluta y el grado de dispersión alrededor de valores centrales o típicos.

Para fijar de una manera aproximada la amplitud del intervalo se puede hacer uso de las siguientes consideraciones matemáticas:

$$k = 1 + 3,3 \cdot \log_{10} n \quad \text{Ecuación 2-3}$$

$$k = 2 \cdot \sqrt[3]{n} \quad \text{Ecuación 2-4}$$

$$k = \sqrt{n} \quad \text{Ecuación 2-5}$$

Donde k es la cantidad de intervalos y n el número de datos

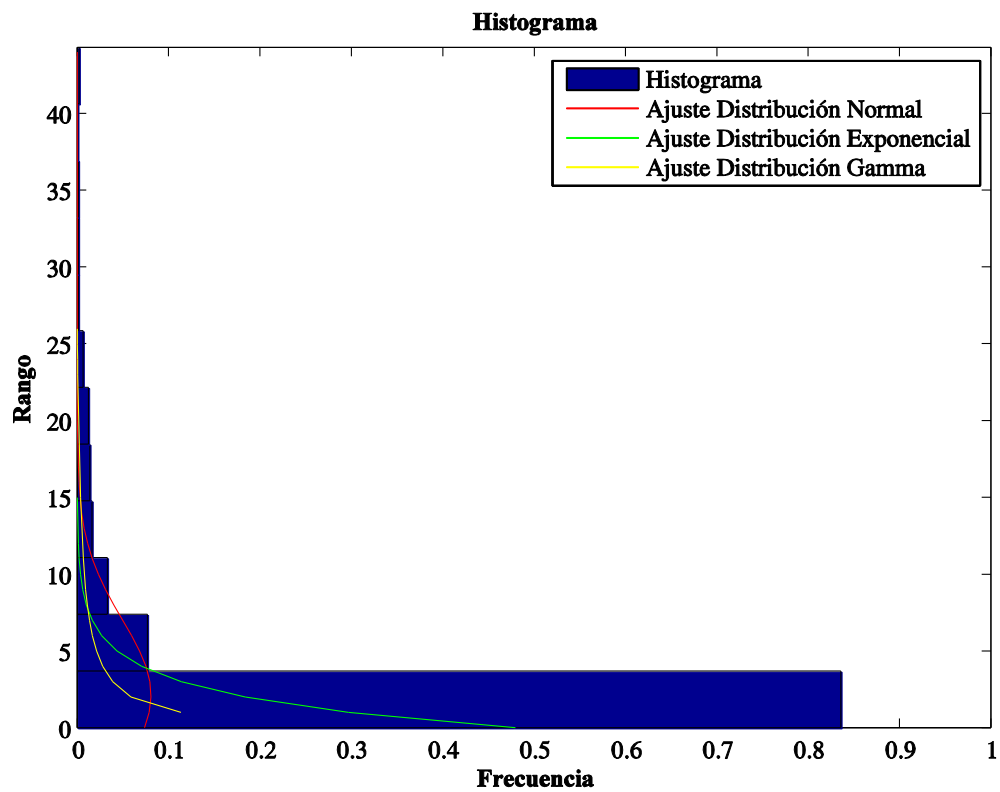


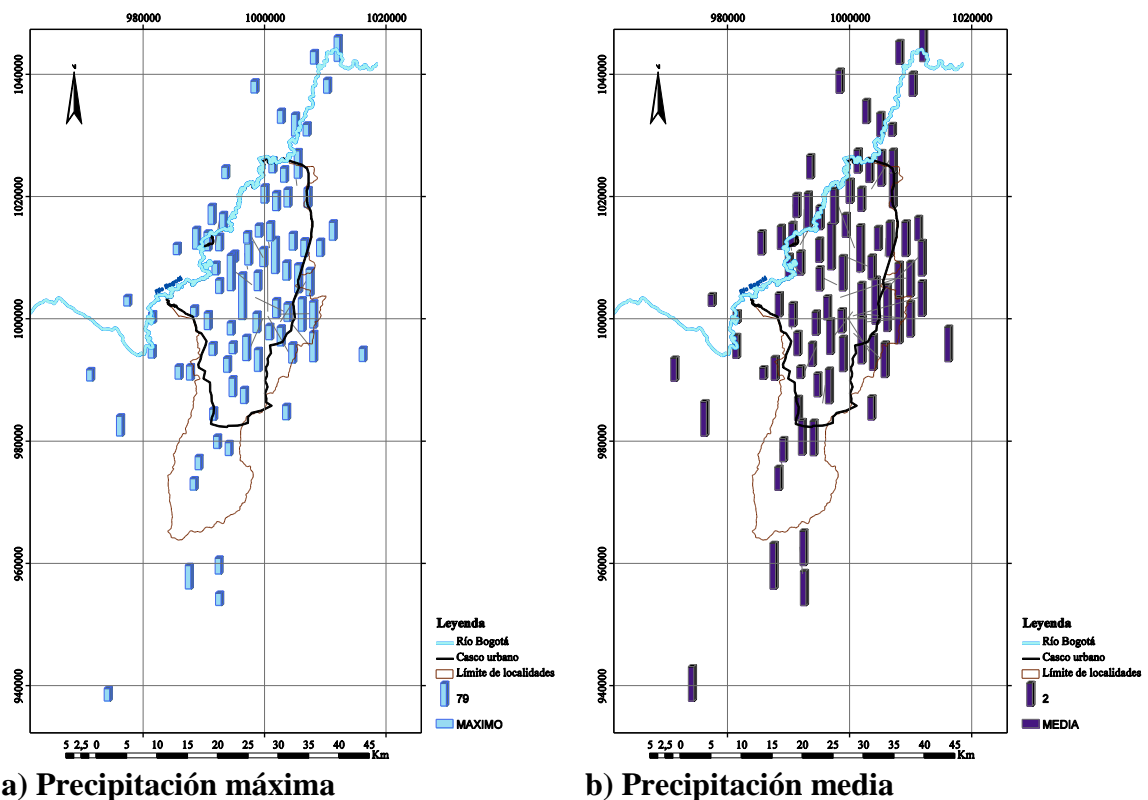
Figura 2-11. Histograma de frecuencias y ajuste a distribuciones de probabilidad para los registros de precipitación diaria de la estación pluviográfica ALCO para el periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.

2.3.2 Análisis Exploratorio de Datos Espaciales (AEDE)

Tomando la información en conjunto de todas las estaciones seleccionadas sobre un mismo plano o *lattice*, el AED aborda el componente geográfico de los datos, haciendo uso de técnicas que permitirán visualizar las distribuciones y patrones espaciales de precipitación, identificar “atípicos espaciales” (spatial outliers), agrupamientos (clusters) o puntos calientes (hot spots).

El AEDE combina el análisis estadístico con el gráfico, dando lugar a lo que podría denominarse una “visualización científica” que hace contrastes estadísticos sobre los efectos espaciales de dependencia y heterogeneidad, a la vez que une un amplio marco de gráficos o “vistas” múltiples y dinámicas sobre la información geográfica.

En el caso de los registros de precipitación diaria en la Figura 2-12 se observa cómo los registros de las máximas precipitaciones se concentran en el costado oriental de la ciudad, efecto de la topografía de los cerros. Por el contrario, la distribución de la precipitación media no presenta un patrón predominante, posiblemente por la influencia de la gran cantidad de valores con precipitación igual a cero.



a) Precipitación máxima **b) Precipitación media**
Figura 2-12. Distribución espacial de la precipitación diaria para periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.

La distribución de la precipitación total multianual para el período de cinco años se puede ver en la Figura 2-13. Es claro que la orografía juega un papel importante en el comportamiento de las precipitaciones. Tanto en los cerros orientales, los cerros de suba y los cerros sur occidentales se presentan las máximas precipitaciones totales.

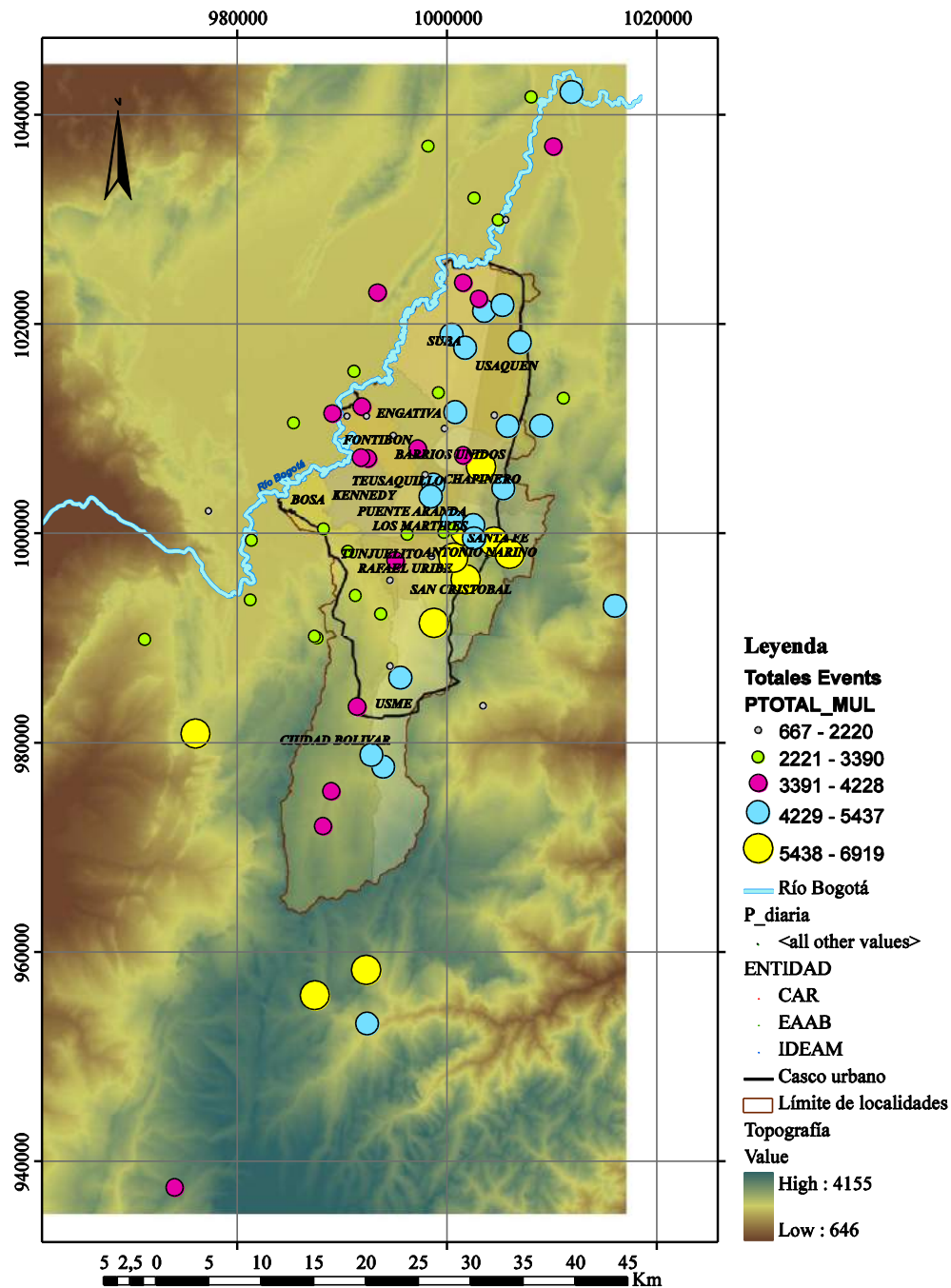


Figura 2-13. Distribución espacial de la precipitación total diaria multianual [mm] para periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.

A partir de la observación de la Figura 2-13, se infiere que el comportamiento de las precipitaciones son similares en algunos sectores específicos. Para hacerlo más gráfico se identificaron por las localidades a las que corresponden: 1) Zona de precipitaciones muy altas: Santa Fe, San Cristobal y Candelaria; 2) Zona de precipitaciones altas: Suba, Usaquen, Chapinero, Teusaquillo y Puente Aranda; 3) Zona de precipitaciones medias: Engativa, Fontibón, Barrios Unidos y Kennedy; 4) Zona de precipitaciones bajas: Ciudad Bolívar.

Dentro de las bases de datos de las entidades es común encontrar períodos donde no se registraron los valores de precipitación. De igual forma para las épocas de poca lluvia el valor de cero es frecuente. Estos componentes de la serie de tiempo son de gran importancia en el momento de realizar la modelación con el autómatas celulares. En la Figura 2-14 se observan cómo en las estaciones localizadas en la zona sur occidental tiene gran número de datos con valor cero (color naranja), gran cantidad de datos sin registro (color rojo) en las estaciones del centro y occidente y precipitaciones mayores a cero en la línea de los cerros orientales (color azul).

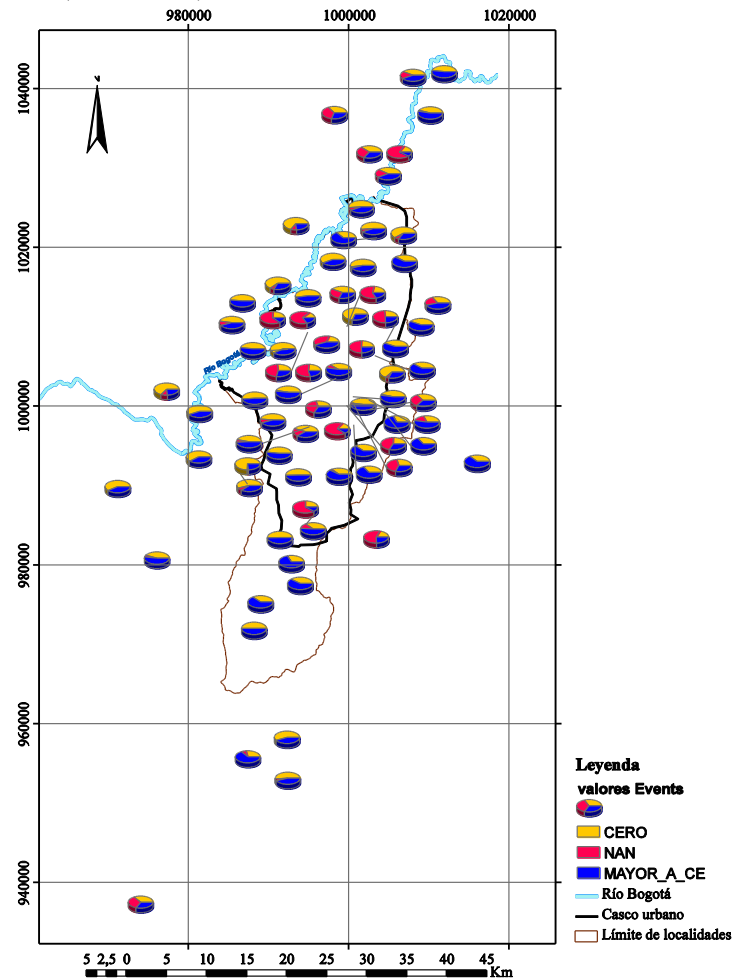


Figura 2-14. Número de datos registrados, no registrados y con valor cero por estación, para el periodo de estudio comprendido entre 01-01-1995 y 31-12-1999.

2.4 SELECCIÓN DEL PERÍODO DE SIMULACIÓN Y EVENTOS DE PRECIPITACIÓN

Luego de tener las series de precipitación tanto diarias como los eventos “depuradas” se procede a la selección del período de simulación y selección de los eventos de tormenta. Con el fin de estudiar el desempeño del pronóstico del modelo de un AC en el pronóstico de la precipitación, el criterio de selección estará dirigido en tres líneas: 1) resolución temporal de la serie de datos; 2) período de modelación y 3) estaciones disponibles.

La resolución temporal de modelación, hace referencia al nivel de agregación de los registros de precipitación. Después de la digitalización de las cartas pluviográficas para los eventos de tormenta, los registros se encuentran con resolución de 30 minutos. Para efectos de la modelación del AC se eligieron agregaciones temporales de 30 minutos, 1 hora y de 2 horas. El criterio de utilizado en esta selección de niveles de agregación se basa en la observación de la duración de las tormentas encontradas en la base de datos. Para los registros diarios sólo se tomó en cuenta esa resolución temporal.

El horizonte de modelación corresponde al período del tiempo o ventana de tiempo sobre la cual se realiza el modelo. En el presente estudio se analizará el pronóstico de: a) eventos de tormenta con resolución de 30 minutos; b) eventos de tormenta con resolución de 1 hora; c) eventos de tormenta con resolución de 2 horas y d) Temporada húmeda comprendida por los meses de marzo, abril y mayo para registros con resolución diaria.

La identificación de los eventos de tormenta se basa en la metodología empleada por Bernal (2007). De igual forma, en el mismo estudio se encontró que en el periodo comprendido entre los años 1995 a 1996 se presentaron 17 tormentas. En la presente investigación solo se analizaron 3 eventos. La selección de estos está definida por los eventos en los que la duración superara las 2 horas y que por lo menos mas de diez estaciones registraran el evento. Los eventos elegidos son los siguientes:

Tabla 2-5. Eventos seleccionados para el estudio de pronóstico.

EVENTO		AÑO	MES	DIA	HORA	MIN
Evento 1	inicio	1995	4	12	10	00
	final	1995	4	13	4	40
Evento2	inicio	1995	5	6	9	30
	final	1995	5	6	22	0
Evento 3	inicio	1996	11	15	10	0
	final	1996	11	15	20	30

La selección de la temporada húmeda para registros diarios se realizó mediante la observación del número de registros con valores de precipitación mayores a cero, con el fin de encontrar valores extremos de precipitación que favorecen la presencia de condiciones que representan riesgo en la ciudad. Dentro de las dos temporadas húmedas que se encuentran en el año, se observó que los meses de marzo, abril y mayo en el periodo disponible presentan precipitaciones diarias mayores a las de los meses de septiembre, octubre y noviembre. Además de los valores altos de precipitación también es importante la cantidad de estaciones que tengan registro. De esta manera se tendrán tres meses en cada año con información disponible. En la Figura 2-15 se muestra la distribución de la precipitación para los meses de marzo, abril y mayo para la estación ALCO en los años 1995 a 1999.

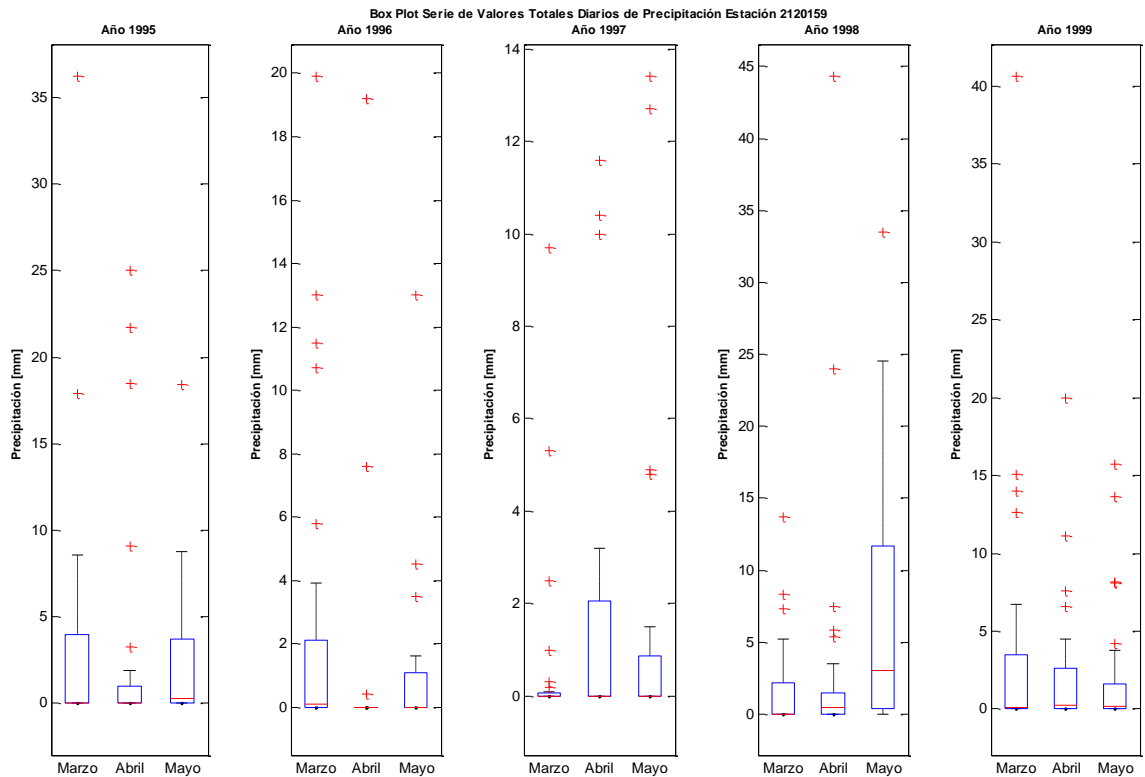


Figura 2-15. Diagrama de cajas para el análisis de registros de precipitación diaria para la estación ALCO. Período de registro 01-01-1995 y 31-12-1999.

3 ESTIMACION DE LOS CAMPOS DE PRECIPITACIÓN

La precipitación es un fenómeno que tiene un fuerte componente geográfico. El estudio de la existencia de patrones espaciales puede contribuir al asentamiento de bases que contribuyan al estudio de la variabilidad de la lluvia. En localidades como Bogotá, donde la medición de la precipitación se hace de manera puntual, es necesario conocer la precipitación como si fuera un campo continuo en el que para cualquier punto se tenga un valor de precipitación.

La interpolación espacial es un procedimiento matemático utilizado para estimar el valor de un atributo, en este caso, la precipitación, en una localidad o ‘celda’ a partir de valores obtenidos en puntos de medición. La interpolación espacial transforma un número finito de observaciones de las estaciones pluviográficas, en un espacio continuo de manera que su patrón espacial sea comparable con aquel presentado por las observaciones puntuales de origen.

Localmente Rodríguez *et. al* (2008b) y Vargas *et. al* (2010) han estudiado para el caso de Bogotá bien sea en un área reducida o en una extensión más amplia los efectos del uso de diferentes técnicas de interpolación espacial. Es importante aclarar que la calidad de los resultados de la interpolación espacial del campo de precipitación depende fundamentalmente de la calidad de los datos, del número y distribución de los pluviógrafos utilizados y del método a utilizar.

El objetivo de esta etapa de la modelación es evaluar la bondad de varios métodos de interpolación espacial para la generación de los campos de precipitación de los eventos seleccionados en el paso anterior o de la serie de tiempo elegida. Esto con el fin de obtener un único mapa que determine los patrones observados para el desarrollo del modelo de AC.

La evaluación de los métodos de interpolación se realizará mediante la comparación de los resultados obtenidos por cada uno de ellos, siguiendo dos criterios:

- Estadística global: donde se verificará el grado de ajuste al modelo real empleando e validaciones cruzadas.
- Evaluación visual: de manera cualitativa se observarán los patrones de comportamiento de sus curvas de nivel y de las superficies.

Dentro de la literatura se encuentran diferentes métodos de interpolación. Éstos son clasificados en globales y locales, de acuerdo a la cantidad de información que utilizan, y en exactos e inexactos, de acuerdo al valor de las estimaciones respecto a los datos de origen (Tabla 3-1). En la presente investigación se evaluarán: a) Inverso de la distancia, b) V4 (Método MATLAB ®) y c) Kriging.

Los métodos globales se caracterizan porque las estimaciones son realizadas utilizando todos los datos disponibles en el área de interés, en cambio los métodos locales operan dentro de una pequeña vecindad, alrededor de la ubicación donde se desea obtener un valor interpolado. Por otra parte, los métodos exactos se diferencian de los inexactos debido a que en los primeros el valor interpolado en una localidad donde existe un dato de muestreo no presenta diferencia respecto al valor utilizado de base para la interpolación.

Tabla 3-1. Clasificación de diferentes métodos de interpolación espacial

Método	Global/Local	Tipo de interpolación
Clasificación	Global	Inexacta
Superficies de tendencia	Global	Inexacta
Modelos de regresión	Global con refinamiento local	Inexacta
Polígonos de Thiessen	Local	Exacta
Interpolación lineal	Local	Exacta
Inverso de la distancia y medias móviles	Local	Inexacta cuando se aplica un suavizamiento de la superficie interpolada
Splines	Local	Exacta, con un refinamiento limitado
Kriging	Local, pero con un variograma global	Exacta
Simulación condicional	Local, pero con un variograma global	Inexacta

Fuente: Burrough y McDonnell (1998), Lam (1983).

En el campo de la ingeniería y la hidrología, el método de interpolación con Kriging es el más usado (Bras y Rodríguez-Iturbe, 1994). Esto se debe a que se basa en la estructura de correlación espacial de los datos, mientras que todos los demás imponen una estructura espacial esencialmente arbitraria. Los métodos de la geoestadística requieren que se asigne unos límites físicos para la estimación del campo de precipitación, el tamaño de la celda y la alineación de la malla.

A continuación se hace algunas consideraciones que se tuvieron en cuenta para la estimación de los campos de precipitación, junto con una breve descripción de los métodos utilizados.

Para la serie de datos de precipitación diaria se realizó el análisis estructural de algunos días en los que se presentaron altos valores de precipitación diaria. Primero se tomaron 4 días del mes de septiembre de 1995, analizando la precipitación en toda la región, es decir el casco urbano y sus alrededores. El criterio de selección de estos días esta en función de la densidad de estaciones que reportaron valores de precipitación total diaria diferente de cero en un periodo continuo de días.

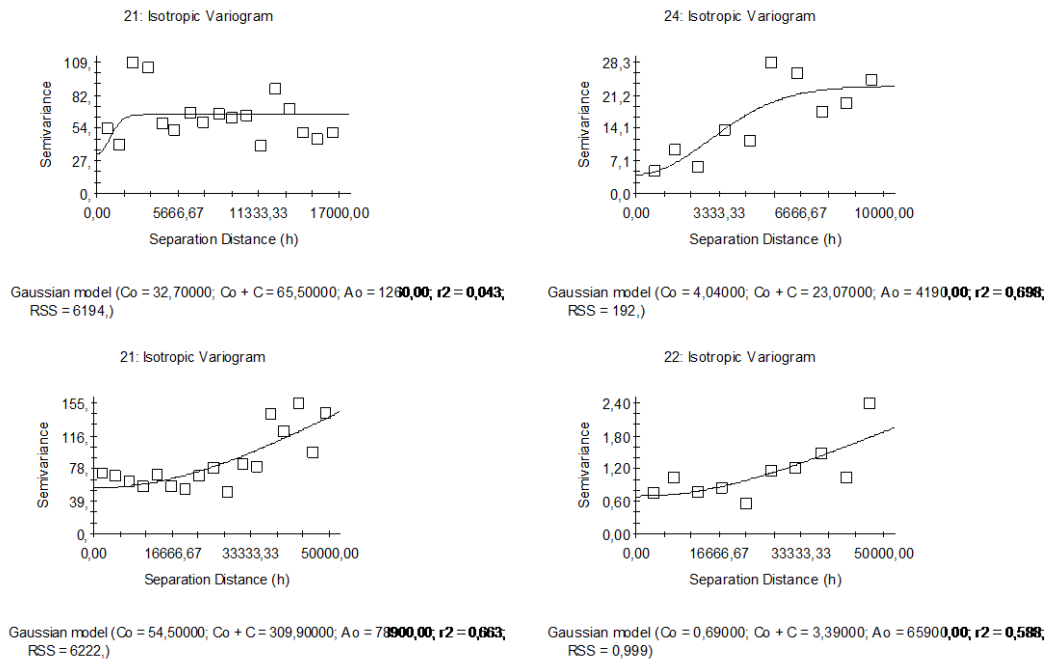
En la Tabla 3-2 se muestran las fechas de los días seleccionados y el modelo que más se ajustó a las condiciones encontradas. Se hace la diferenciación entre el análisis para todas las estaciones y solo para las estaciones localizadas en el casco urbano; debido a que la densidad de estaciones en el último caso es mayor.

El software utilizado es el GS+, que se distribuye de manera gratuita en la red en el sitio <http://www.gammadesign.com/>. Se estimaron los semivariogramas experimentales con el fin de elegir el modelo teórico que mejor representara el fenómeno.

Tabla 3-2. Parámetros del semivariograma para precipitación diaria. Caso 1: Estaciones localizadas en el casco urbano, Caso 2: Estaciones localizadas en el casco urbano y alrededores

Localización	Fecha	Modelo	Alcance [m]	R ²
Estaciones en el casco urbano (50 estaciones)	1995-09-21	Gaussiano	1260	0.043
	1995-09-22	Gaussiano	65900	0.588
	1995-09-24	Gaussiano	4190	0.698
	1995-09-30	Esférico	1680	0.182
Todas las estaciones (77 estaciones)	1995-09-21	Gaussiano	78900	0.663
	1995-09-22	Gaussiano	13920	0.973
	1995-09-24	Esférico	7800	0.506
	1995-09-30	Exponencial	7500	0.457

En la estimación de los semivariogramas del análisis estructural se encontró que no existe una dirección en la que se encuentren mejores ajustes a los modelos teóricos. En la Figura 3-1 se muestran algunos de los semivariogramas ajustados a los modelos teóricos.



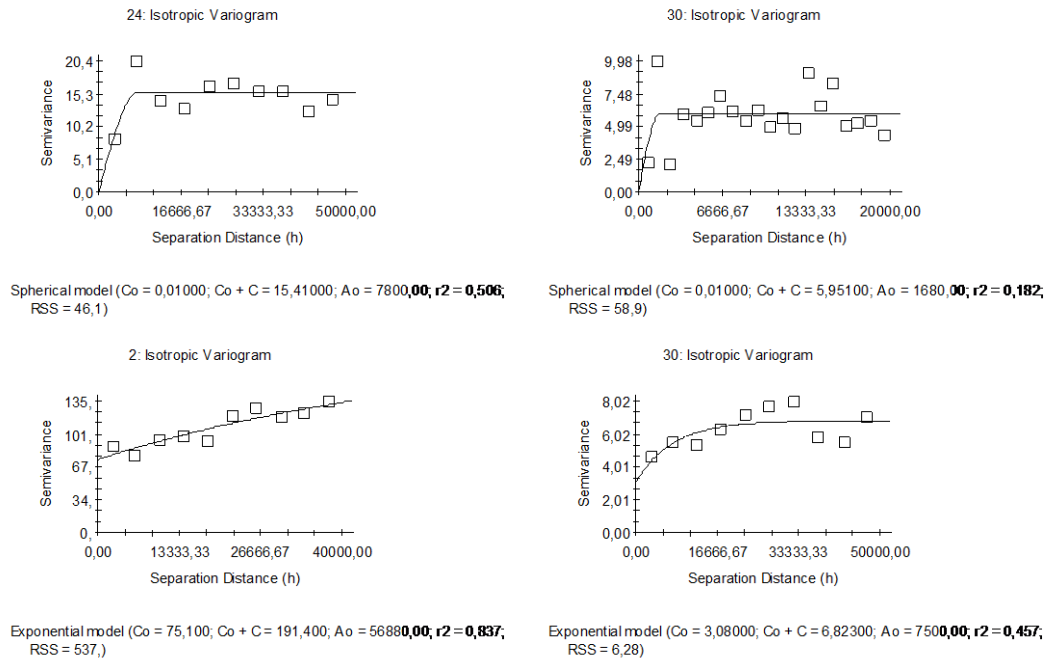


Figura 3-1. Representación grafica de los modelos de semivariograma para algunos registros de precipitación diaria.

Nota: las unidades de longitud como h están en metros.

3.1 CONSIDERACIONES PARA LA CONSTRUCCIÓN DE LOS CAMPOS DE PRECIPITACIÓN

3.1.1 Selección del Tamaño de Celda

Para definir el ancho de celda con el que se visualizan los mapas de precipitación (formato “raster”), se sugirió que el tamaño de celda no debería ser menor al mínimo desplazamiento del centro de tormenta. Ello con el fin de lograr identificar el movimiento de la tormenta con mayor facilidad. Para ello, en un evento de tormenta se estimó la posición (coordenada norte y este) para cada instante de tiempo en la que se presentó el mayor valor de precipitación. El centro de tormenta se definió como el centro de masa, o lugar donde se concentró el mayor volumen de agua precipitada en cada intervalo de tiempo,

$$R_{CM} = \frac{\sum_i p_i \cdot r_i}{\sum_i p_i} = \frac{1}{P} \cdot \sum_i p_i \cdot r_i \quad \text{Ecuación 3-1}$$

donde R_{CM} es la coordenada norte o este del centro de tormenta, p_i es la precipitación registrada en la estación i , P es la precipitación total registrada en un intervalo de tiempo, y r es la coordenada norte o este de la estación i .

Para este análisis se estudiaron 17 tormentas comprendidas en el periodo de 1995 a 1996. Adicionalmente se analizaron otras dos que no pertenecen al periodo 95-96 pero se estudiaron por considerarse eventos de magnitud considerable en duración y cantidad de agua precipitada. En la Figura 3-2 se muestran las frecuencias de las distancias mínimas en las que se movió el centro de tormenta para los 19 eventos analizados con intervalos de tiempo de 30 minutos.

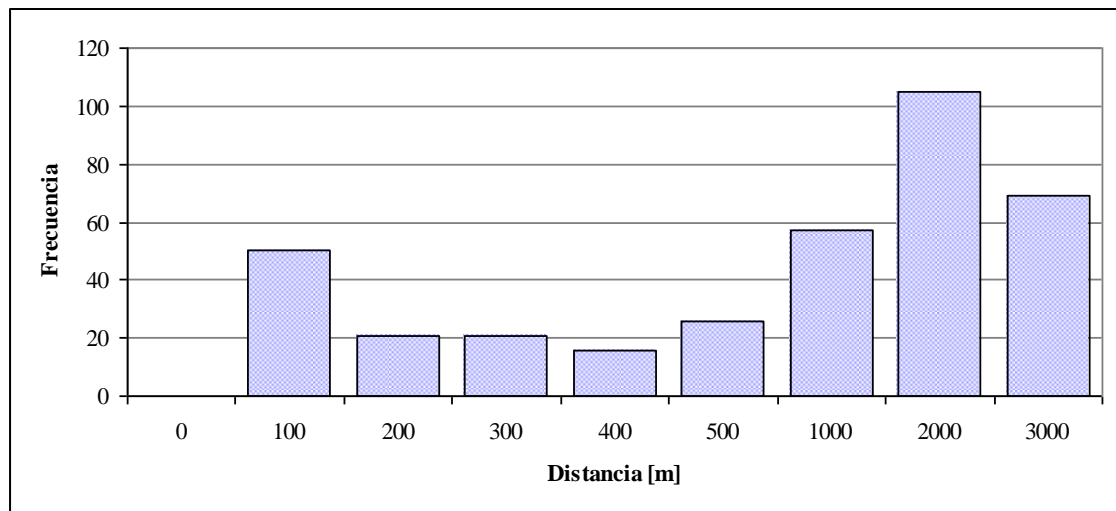
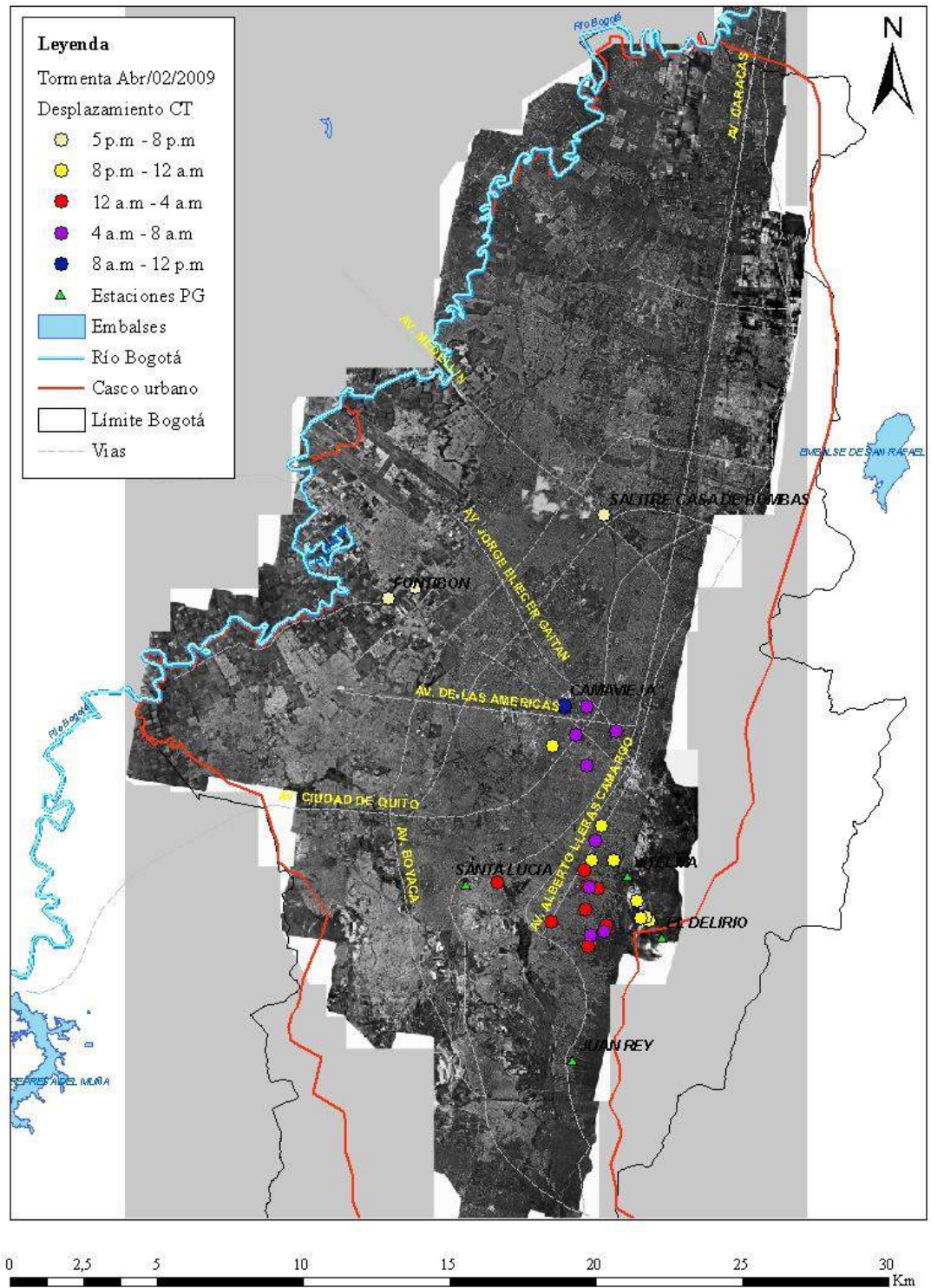


Figura 3-2. Frecuencia de la mínima distancia de desplazamiento del centro de tormenta.

En la Figura 3-3 se observa cómo se desplaza el centro de tormenta sobre la ciudad de Bogotá para dos eventos. Especialmente se observa que no existe una dirección predominante y que por lo general se concentran los centros de tormenta en la zona del centro de la ciudad.

Tomando en cuenta que para escalas espaciales mayores a 1 km^2 varios autores han reportado la importante variabilidad espacial y temporal de la precipitación (Butyaert et al., 2006, Segond et al., 2006, Vaes et al., 2005 citado por Rodríguez 2008a) no se consideró que la resolución de la celda fuese de 1000 metros, como es el caso más frecuente (Figura 3-2). Le siguen las distancias más frecuentes de 500 metros y 100 metros, para el caso de la distancia de 100 metros se identificó que aumenta los tiempos computacionales y que como se mostrará más adelante no representa mejoras en la interpolación espacial. En conclusión, la distancia que se consideró más apropiada es la de 500 metros.



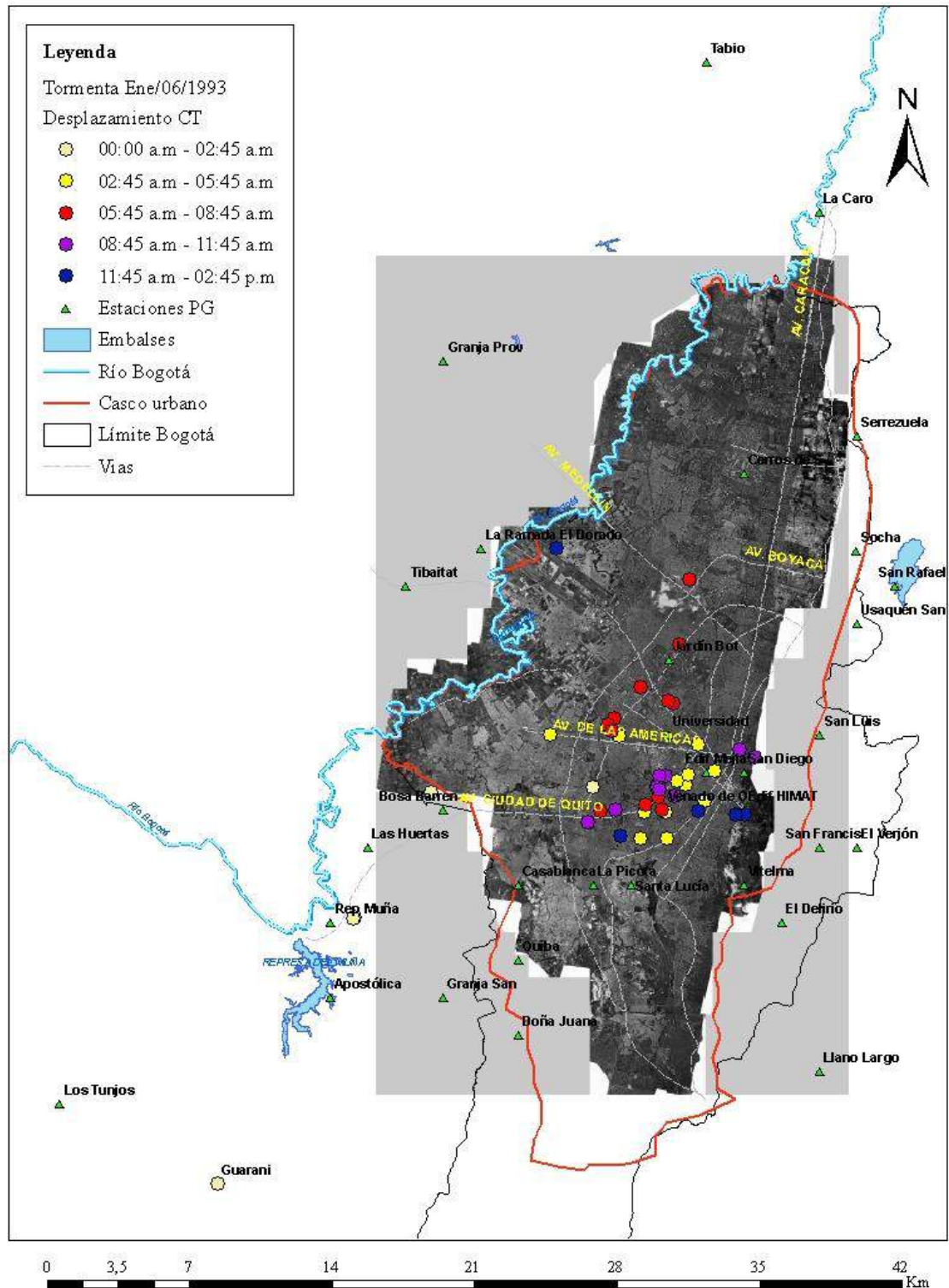


Figura 3-3. Desplazamiento del centro de tormenta en la ciudad de Bogotá. Abril 02 de 2009 (sup.) y 2) Enero 06 de 1993 (inf.)

3.1.2 Alineación de la Malla

Con el fin de obtener una orientación de malla que mejorara la representación de los campos de precipitación, se trató de establecer un patrón espacio-temporal de los vientos en el casco urbano, analizando registros comprendidos entre los años 1997 y 2006, que correlacionara estos dos procesos meteorológicos.

En principio se sugirió que la orografía y la temporada (mes, trimestre, semestre, día, noche, etc) jugarían un papel importante en la definición de un patrón de la dirección de los vientos en Bogotá. Se elaboraron entonces las rosas de los vientos para cada uno de los años donde se tenían registros (Ver algunas en la Figura 3-4), agregando las series de las diferentes estaciones de la Red de Calidad de Aire de Bogotá. No obstante, no se pudo llegar a un patrón claro y concluyente que pudiera vincular dichos procesos en el análisis por año.

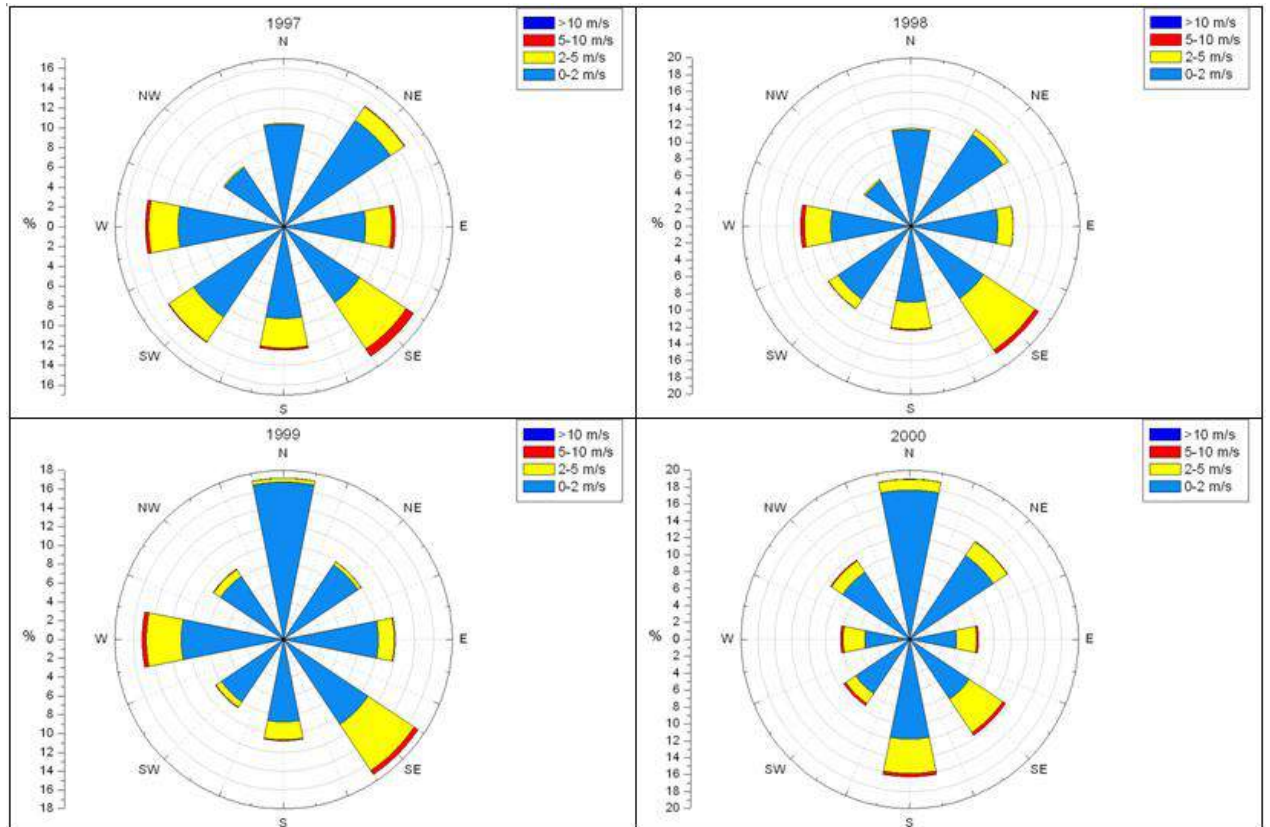


Figura 3-4. Rosa de los vientos anual construida a partir de registros de las estaciones de Calidad de Aire de Bogotá. Año 1997 a 2000.

Realizando el análisis multianual de los registros se observa que el patrón predominante se encuentra en la dirección norte-sur (Figura 3-5), entonces se empleó el sistema de coordenadas planas sin ninguna rotación como el sistema de referencia de los campos de precipitación interpolados.

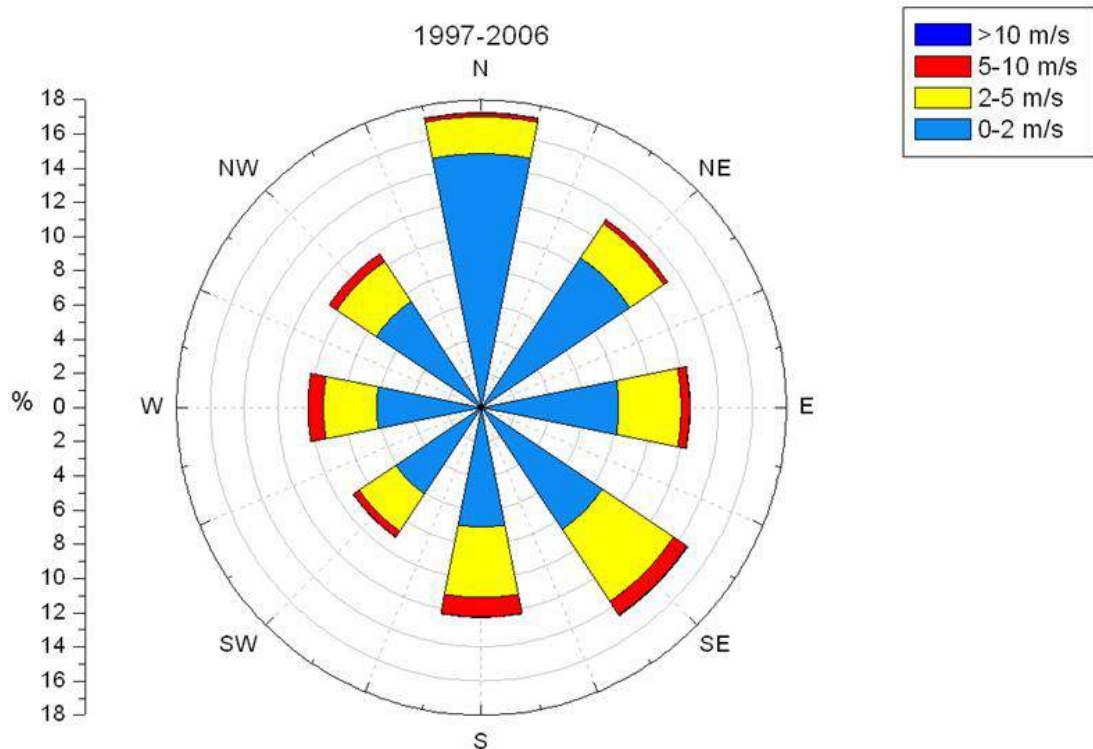


Figura 3-5. Rosa de los vientos construida partir de registros de las estaciones de Calidad de Aire de Bogotá con resolución cada 10 minutos del periodo comprendido entre los años 1997 a 2006.

3.1.3 Cobertura Espacial de la Interpolación

Con el fin de disminuir la incertidumbre de los valores interpolados en las fronteras de un campo de precipitación calculado (interpolado a partir de registros en superficie) se recomienda que se disponga de estaciones localizadas fuera de la zona de interés. De esta manera en el momento de hacer la interpolación se contara con más información, para luego hacer el “recorte” de la zona de interés.

Adicionalmente se sugirió que la dirección de las líneas de coordenadas fuesen lo más cercano posible a la malla del TRMM (Misión de Medición de Lluvias tropicales. *Tropical Rainfall Measuring Mission* TRMM por sus siglas en inglés) con el fin de que en el futuro se pueda combinar los valores pronosticados con los estimados por el radar (ver Figura 3-6).

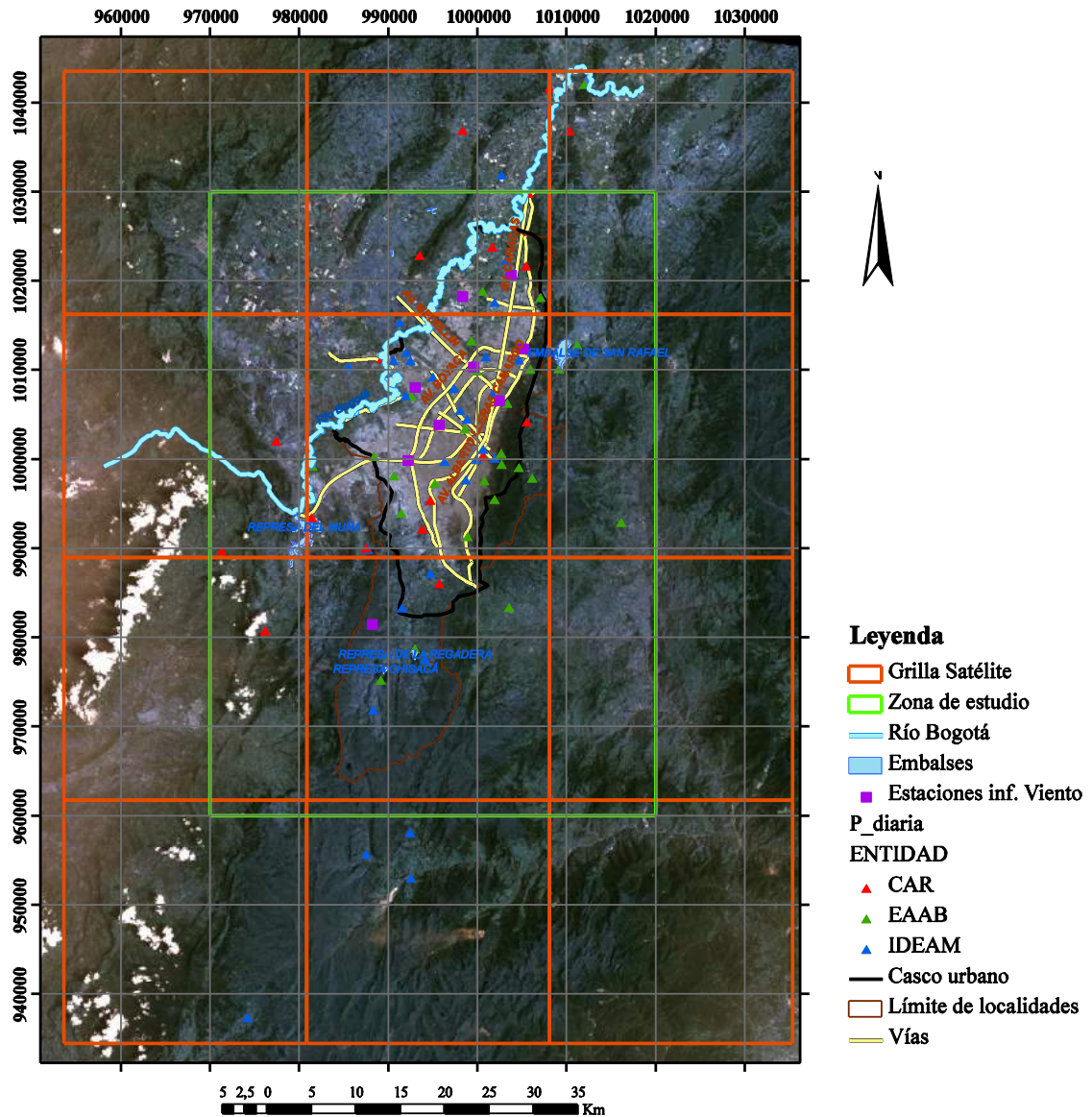


Figura 3-6. Cobertura de la interpolación espacial, de acuerdo con la cercanía de la malla del TRMM. Ciudad de Bogotá.

Nota: Las líneas naranjas muestran las celdas de la estimación obtenida por el TRMM. El rectángulo verde muestra la zona sobre la cual se realizó el “recorte” para así delimitar una zona de estudio específica.

3.2 SELECCIÓN DE MÉTODO DE INTERPOLACIÓN - VALIDACIÓN CRUZADA

Un aspecto importante en la selección de la metodología de interpolación más representativa es la validación de los campos de precipitación generados. Para realizar esta selección se realizó un procedimiento de validación cruzada, empleando los registros de 5 estaciones localizadas dentro de la zona de estudio para tres métodos de interpolación espacial: IDW, V4 y Kriging (En el Apéndice B se encontrarán los fundamentos básicos de cada uno de interpoladores estudiados).

El criterio de selección de las estaciones se basó en la extracción de por lo menos una estación de las siete zonas hidrológicas propuestas en el estudio de IRH (1995). En la Figura 3-7 se muestran los límites de las zonas y las estaciones con registro de precipitación diaria. Las estaciones seleccionadas son: El Granizo (IDEAM), Sta. María de Usme (IDEAM), Fontibón (EAAB), CASD (IDEAM) y UNAL (IDEAM).

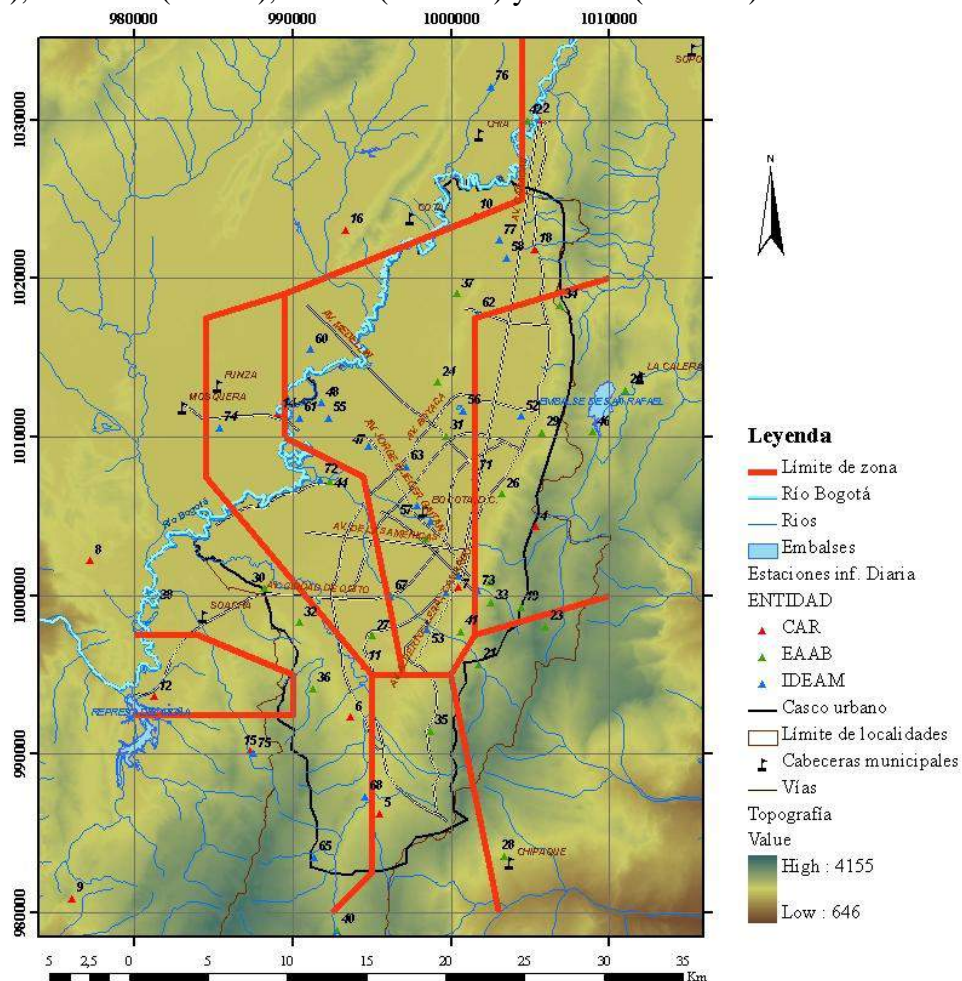


Figura 3-7. Zonificación pluviográfica de las estaciones localizadas en la ciudad de Bogotá y alrededores.

Fuente: Estudio IRH (1995)

Los registros de estas estaciones no se incluyeron en la etapa de interpolación, aspecto que permite identificar las ventajas y desventajas de cada una de las metodologías empleadas.

Para realizar esta validación cruzada se consideraron dos aspectos importantes: la frecuencia de éxito de cada metodología y la precisión en las predicciones realizadas.

El parámetro de comparación empleado en este análisis fue el valor absoluto del error entre el valor estimado y el valor registrado. Considerando la frecuencia de éxito se seleccionó la metodología que generara el mayor número de éxitos y en relación con la precisión de cada metodología se seleccionó la metodología que produjera el menor error acumulado (considerando todos los días registrados). Los resultados de esta validación cruzada se presentan en la Tabla 3-3.

Tamaño de celda [m]	Frecuencias		Error	
	Método	\sum Error	Método	\sum Error
100	IDW	3924.2	IDW	3924.2
200	Kriging *	4112.0	Kriging *	4112.0
250	V4	5869.7	IDW	3992.2
500	IDW	3897.7	IDW	3897.7

* Se empleó un semivariograma lineal por mostrar mejores ajustes al comparar los valores de las celdas de observados y los simulados.

Tabla 3-3. Comparación de metodologías de interpolación en función del tamaño de celda.

Con base en los resultados consignados en anterior tabla, se observa que el método IDW genera la menor diferencia entre los valores estimados y los registrados. En general puede apreciarse que el método V4 genera diferencias sustanciales, aspecto que se evidencia en las altas sumatorias del error obtenidas.

Adicionalmente al análisis explicado anteriormente, se analizaron los diagramas de dispersión para los casos analizados, los cuales se presentan en la Figura 3-8 y la Figura 3-9. En la Figura 3-8 se presentan los diagramas de dispersión entre los valores estimados y los valores registrados comparando los diferentes semivariogramas empleados para el método de Kriging. En esta Figura se observa que el mejor semivariograma para este método es el semivariograma Gaussiano, con parámetros fijos.

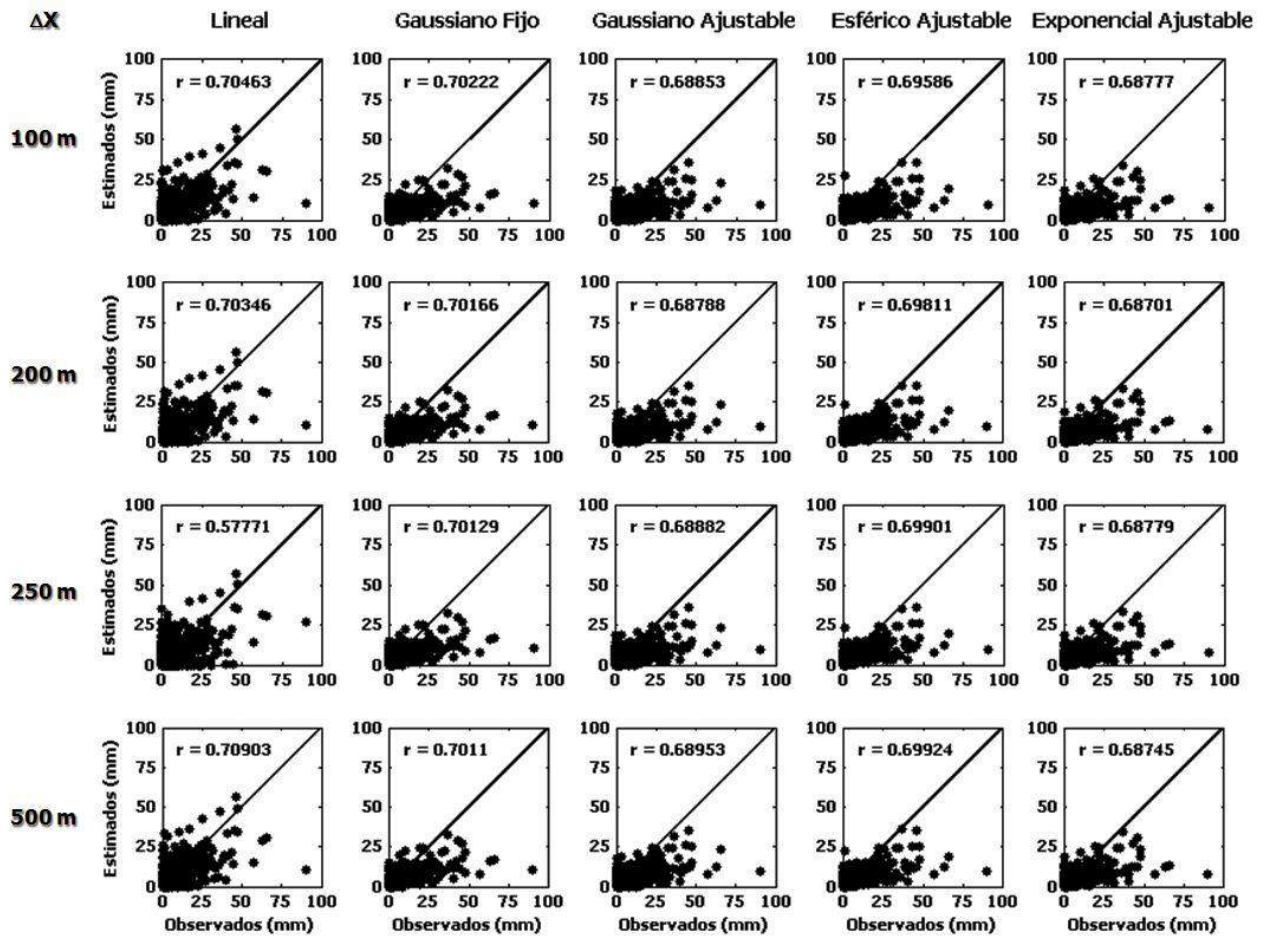


Figura 3-8. Comparación del ajuste de las variaciones del método de Kriging.

Comparando ahora los resultados del mejor semivariograma para el método Kriging y los otros métodos considerados se observa que el método IDW obtiene los mejores resultados, Véase Figura 3-9.

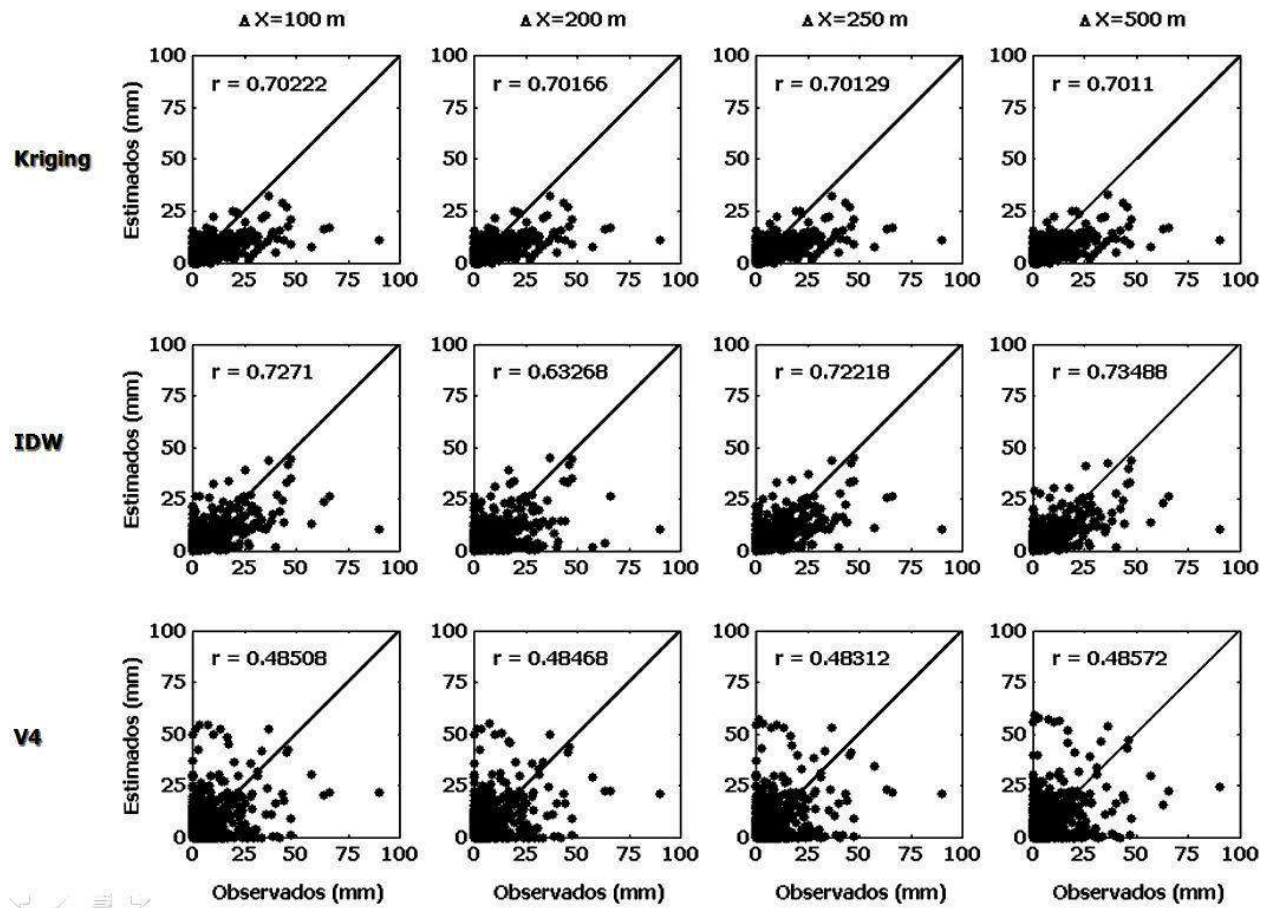


Figura 3-9. Comparación del ajuste obtenido por cada una de las metodologías empleadas.

Del anterior análisis se infiere que el método que mejor representa las características de la precipitación en la estaciones de validación fue el método IDW, y que el variar el tamaño de celda no representa una notable mejoría en los resultados de la validación cruzada por lo que el tamaño de celda recomendado es de 500 metros. El algoritmo implementado para la construcción de los campos de precipitación con el método de IDW se puede consultar en el compendio de algoritmos al final del documento.

4 PROPUESTA METODOLÓGICA PARA LA IMPLEMENTACIÓN DE UN AC EN EL ESTUDIO DE LA DINÁMICA Y PRONÓSTICO DE CAMPOS DE LLUVIA

Aunque se reportan en la literatura modelaciones hidrológicas con diferentes herramientas de sistemas inteligentes, las experiencias con precipitación mediante autómetas celulares son todavía escasas. En este sentido, esta investigación propone una metodología para construir un modelo de pronóstico de precipitación espacial basado en los fundamentos de los autómetas celulares. Tal aproximación involucra estimaciones de campos de precipitación que pueden ser obtenidos con radar o procesos de interpolación a partir de estaciones puntuales pluviométricas. La metodología expuesta en el presente capitulo esta sugerida para el caso en que la adquisición de la información pluviométrica es a partir de estaciones puntuales.

Para lograr implementar el modelo de autómeta, al igual que en cualquier modelo hidrológico se debe realizar previamente el procesamiento de la información para luego implementar el modelo (Figura 4-1). A continuación se describirá la etapa de procesamiento de la información y luego la implementación del modelo.

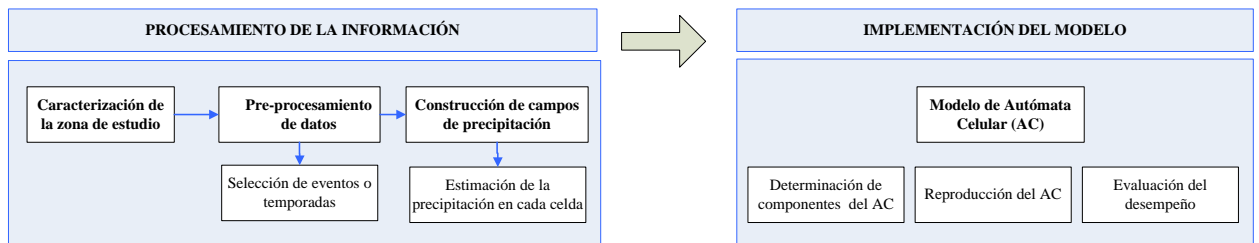


Figura 4-1. Esquema general de la metodología de implementación de u modelo de AC en el estudio de pronóstico de precipitación.

4.1 PROCESAMIENTO DE LA INFORMACIÓN

Esta etapa corresponde al análisis de la información y preparación para la implementación del modelo de autómeta celular (Ver Figura 4-2). Inicialmente se debe definir la zona de estudio de acuerdo con la ubicación de las estaciones meteorológicas disponibles procurando tener una cobertura espacial superior al área de interés para así tener mayor confianza en los valores interpolados en las fronteras. En algunos casos la información suministrada por las entidades operadoras de la red de monitoreo del clima no se encuentra en el mismo formato o debe procesarse para llevarla a formato digital (como

es el caso de las cartas pluviográficas). Para ello debe unificarse el formato de presentación, definir la organización de la base de datos y el software que se utilizará.

Es importante caracterizar la zona de estudio tanto desde el punto de vista geográfico como desde el hidrológico. El componente geográfico debe considerar la influencia de factores como el orográfico, que de alguna manera condicionan el comportamiento de la precipitación y las masas de aire. La caracterización hidrológica permitirá identificar las épocas secas y las húmedas, el orden de magnitud de los valores extremos y sus frecuencias. De acuerdo con la extensión del área de estudio y el interés específico de la modelación (pronósticos a largo o corto plazo, con fines de diseño de estructuras hidráulicas, sistemas de alerta temprana etc.) es posible empezar a inferir la resolución espacial que se desea de la malla del autómatas y la resolución temporal (minutos, horas, días etc.).

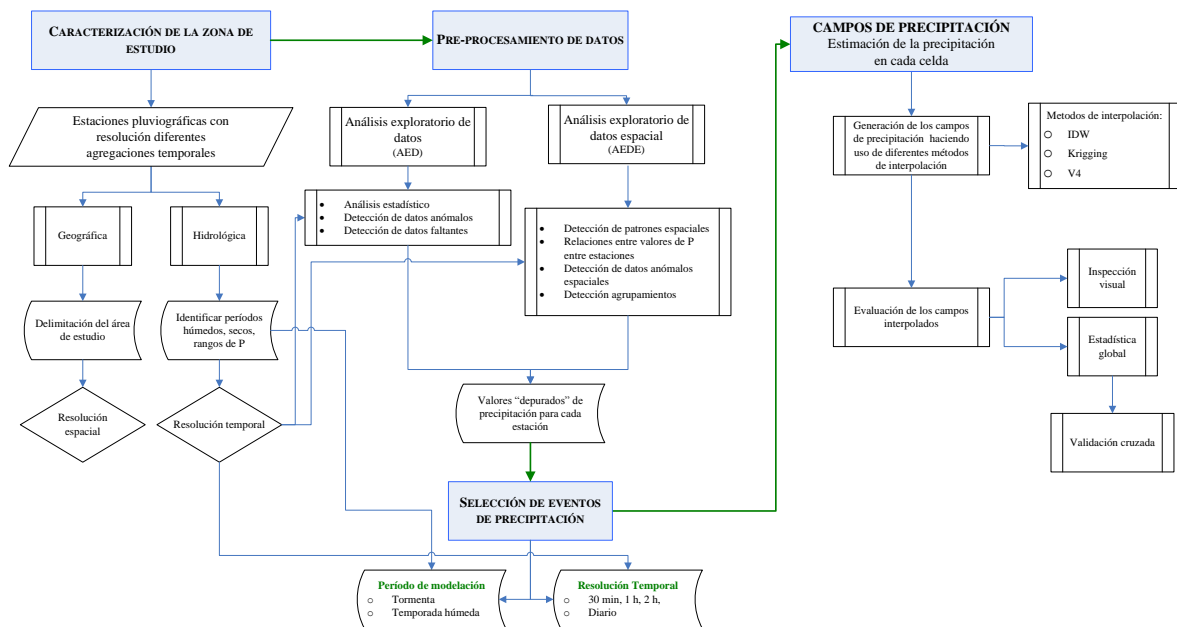


Figura 4-2. Metodología del procesamiento de la información para la implementación del modelo de autómatas celulares.

La calidad y cantidad (en términos de densidad de estaciones y en longitud de la serie de tiempo) de la información pluviométrica es de gran importancia, pues de ello depende la calidad de la interpolación y por ende los resultados del modelo. Con los registros de estaciones meteorológicas a diferentes escalas temporales se realizan análisis a las series de tiempo para así identificar sus principales características estadísticas. Se realiza el análisis exploratorio de los datos, tanto a nivel puntual como el espacial (AED y AEDE, respectivamente), esto con el fin de “depurar” las series y así proceder a una manipulación de la información más confiable. Los análisis exploratorios permiten identificar patrones espaciales, relación entre valores de precipitación entre estaciones, detección de valores anómalos y agrupamiento.

Luego del AED se procede a seleccionar los eventos de interés en el caso de las tormentas o del periodo de interés, bien sea en resolución horaria o diaria.

Debido a que el insumo principal para la aplicación del modelo de autómata celular es un campo en el que los valores sean continuos, es necesario aplicar técnicas de interpolación espacial de los registros puntuales de precipitación. La autora sugiere la utilización de varios interpoladores espaciales para luego mediante inspección visual y validación cruzada elegir la técnica que mejor representa el comportamiento de la precipitación. Se recomienda enfatizar en la resolución espacial de las celdas del campo de precipitación, teniendo en cuenta los siguientes aspectos:

- Requerimientos de pronóstico, es decir el uso que tendrá el producto del pronóstico
- Distancia mínima del movimiento del centro de tormenta, con el fin de evitar resoluciones gruesas que no permitan ver el desplazamiento de la tormenta.
- Referencia del tamaño de celda en estudios previos en la zona de interés.
- Tiempo computacional, pues una celda de menor tamaño implica mayor tiempo computacional.

Finalmente se recomienda que tanto la alineación como la localización de las celdas del campo interpolado sean se fácil integración con la información obtenida de sensores remotos con el objetivo de lograr que en el futuro se diseñe un modelo que combine la bondad de ambas formas de adquisición de la información de precipitación.

En la medida que se tenga más información meteorológica de la zona de interés con la misma resolución temporal y en el mismo período que la información de precipitación, es de utilidad generar otros campos continuos con el fin de alimentar el modelo de autómata con otras variables involucradas con el proceso de precipitación.

4.2 MODELO DE AUTÓMATA CELULAR

El desarrollo de la metodología para la implementación del modelo de autómata celular parte de la necesidad de presentar nuevas propuestas de desarrollo de modelos de pronóstico de precipitación. En la Figura 4-3 se encuentran las cuatro etapas principales del modelo de autómata celular: 1) Definición de los elementos del autómata, 2) Reproducción del modelo de autómata celular, 3) evaluación del desempeño del modelo y 4) análisis de sensibilidad.

4.2.1 Definición de los Elementos del Autómata

Es fundamental la definición de los elementos del autómata celular. Varios de estos elementos fueron establecidos desde la etapa de “Procesamiento de la información” y otros

elementos como la determinación de las reglas de evolución son definidas en esta etapa. Para identificarlos de manera más clara se describirán a continuación:

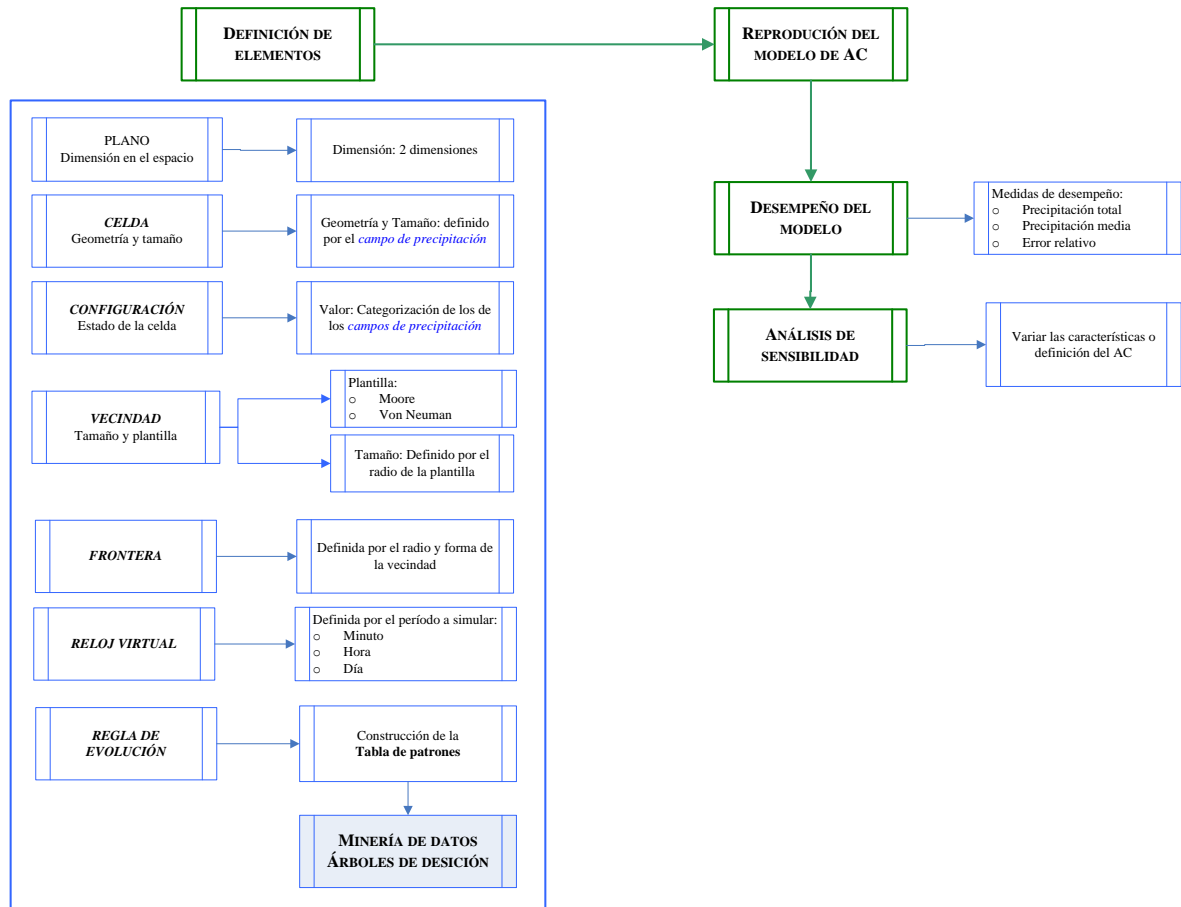


Figura 4-3. Metodología propuesta de la implementación del modelo de autómatas celulares para el pronóstico de la precipitación espacial.

4.2.1.1 Plano y Celda

El estudio del pronóstico espacial de la precipitación se realiza en dos dimensiones y es definido por el campo de precipitación. El campo es un mapa tipo raster en el que las celdas son de forma cuadrada. La extensión del plano está establecido por la delimitación del área de estudio y el tamaño de la celda, por las consideraciones descritas en el numeral anterior.

4.2.1.2 Configuración

La configuración hace referencia al estado de la celda. Con el fin de reducir la variación de tantos valores y poder construir la tabla de patrones en la fase de definición de las reglas de evolución, se realiza una categorización de los datos, agrupando por rangos los valores de las celdas y asignándoles una etiqueta. Los valores pueden variar desde precipitación cero hasta las precipitaciones que superan los 90 mm en el caso de la precipitación total

diaria de la temporada de lluvia de los meses de marzo, abril y mayo entre los años 1995 a 1999 en la ciudad de Bogotá. Si por ejemplo las etiquetas se definieran de acuerdo con los límites de la Tabla 4-1, las celdas con valores entre 0 y 0,2 son precipitaciones con etiqueta “Seco”, precipitaciones con valores entre 0,2 y 0,5 la etiqueta será “Ligeras” y así con toda la serie de datos. En ese caso la configuración del mapa solo tendría siete valores o colores (Ver Figura 4-4).

Tabla 4-1. Ejemplo de asignación de etiquetas a la serie de datos de precipitación diaria.

Clasificación	Limite Inferior	Limite superior	Etiqueta
	[mm]		
1	0	0.2	Seco
2	0.2	0.5	Ligeras
3	0.5	0.7	Lloviznas
4	0.7	1.0	Aguacero
5	1.0	5.0	Lluvias
6	5.0	10.0	Altas
7	10.0	100.0	Fuerte

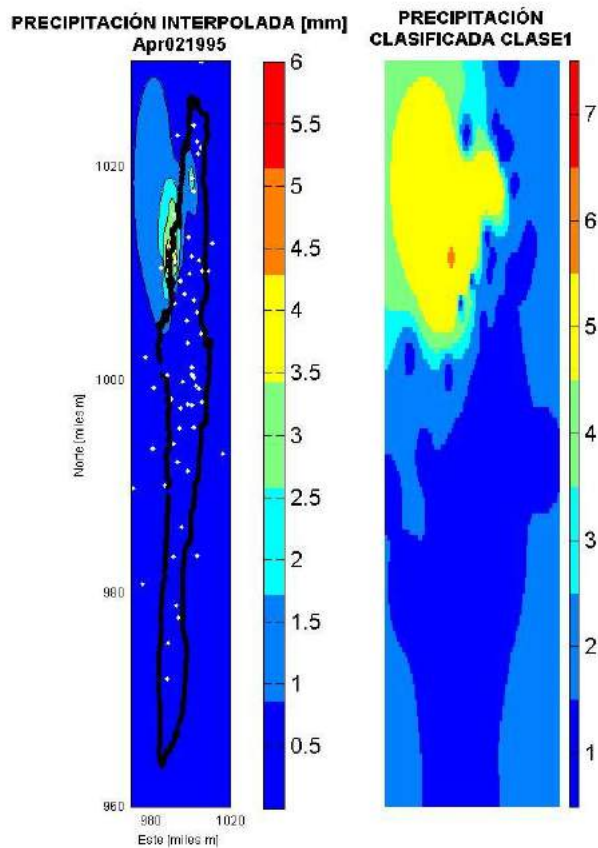


Figura 4-4. Ejemplo de clasificación del mapa interpolado para la serie de datos de precipitación diaria. Nota: Resolución de celda 500 metros. En dirección norte-sur existen 70 kilómetros y en dirección este-oeste 50 kilómetros.

4.2.1.3 Vecindad

Como se explicó en el capítulo de marco teórico, las vecindades más usadas en la aplicación de ACs bidimensionales son las de Von Neumann y Moore. Pero se pueden hacer variaciones a éstas y construir nuevas vecindades como las mostradas en la Figura 4-5, donde la forma ya sea de círculo o rectángulo varía al añadirle un corredor o sector de influencia, que está definido por un radio interno y externo, respecto a la celda central. En la Figura 4-5 es claro que el tamaño de la celda para un mismo espacio puede variar.

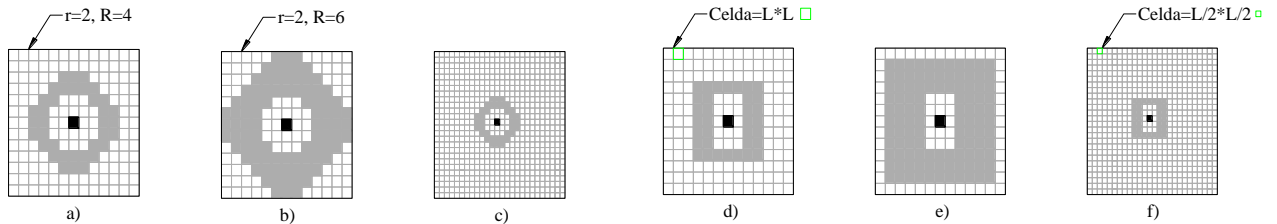


Figura 4-5. Tamaños y formas de vecindad.

a) Círculo radio interior 2, radio exterior 4, tamaño de celda $L*L$; b) Círculo radio interior 2, radio exterior 6, tamaño de celda $L*L$; c) Círculo radio interior 2, radio exterior 4, tamaño de celda $L/2*L/2$; d) Rectángulo radio interior 2, radio exterior 4, tamaño de celda $L*L$; e) Rectángulo radio interior 2, radio exterior 6, tamaño de celda $L*L$; f) Rectángulo radio interior 2, radio exterior 4, tamaño de celda $L/2*L/2$.

4.2.1.4 Frontera

La frontera estará definida de acuerdo con la forma y tamaño de la vecindad. En el caso de seleccionar una frontera adiabática tanto la primera como la última fila y columna serán replicadas en dirección hacia afuera para definir la condición de borde. En la Figura 4-6 se da un ejemplo de la frontera adiabática. El tamaño o el radio de la vecindad definirá cuantas filas y columnas se deberán replicar.

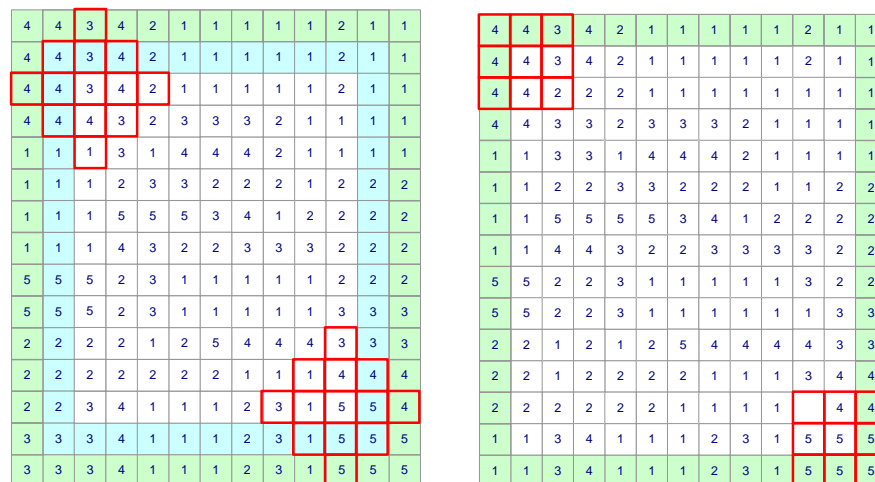


Figura 4-6. Ejemplo de la definición de una frontera adiabática.

Nota: a) Vecindad tipo Von Neuman (izq.) y Moore (der.), b) Las celdas de color azul y verde indican las filas y columnas que se replicaron. Las celdas blancas representan el mapa clasificado.

4.2.1.5 Reloj Virtual

El reloj virtual actualiza cada una de las celdas del espacio en instantes de tiempo que están definidos por la resolución temporal con la que se desee hacer el pronóstico. Por ejemplo si las tormentas están discretizadas cada 30 minutos, el reloj virtual actualizara los estados de las celdas cada 30 minutos.

4.2.1.6 Reglas de Evolución

Las reglas de evolución definen la variación de los estados de las celdas. Básicamente el cambio de estado de una celda de un instante de tiempo a otro está dado en función del estado de sus celdas vecinas y de la misma celda. En ese sentido las reglas son funciones sencillas.

El modelo de autómeta celular propuesto es un modelo guiado por datos, donde el insumo son los registros de precipitación. En ese sentido se deben inferir o detectar las reglas a partir del aprendizaje de la base de datos.

La técnica a usar es la “Minería de datos” (*Data Mining*). Esta herramienta es útil en la identificación de patrones, asociaciones, cambios y anomalías. Entre los algoritmos que se incluye la minería de datos están: a) árboles de decisión, b) redes neuronales, c) modelos estadísticos, y d) agrupamiento o *clustering*, etc.

Se propone el uso de los árboles de decisión para establecer estas reglas. La obtención de las reglas implica la construcción de una tabla de patrones en la que de manera ordenada se registren los valores de los estados de la celda de interés y el de sus celdas vecinas.

La tabla de patrones es una matriz donde el número de columnas es igual al número de celdas que tenga la vecindad y el número de filas es igual al número de celdas del mapa clasificado. El orden de llenado de la matriz lo define la nomenclatura de cada posición de la celda en la plantilla de la vecindad. En la Figura 4-7 esta esquematizada la manera en que se construye la tabla de patrones para el caso de una vecindad tipo Moore.

En el ejemplo de la vecindad de Moore, los valores registrados en las nueve columnas corresponden a las nueve posiciones y sus respectivos valores o estados en el tiempo t . Adicionalmente a la tabla de patrones se le añade una columna en la que se registra el valor de la celda de central de la vecindad en el tiempo $t + 1$. Es posible añadir otras columnas con otras variables meteorológicas, el valor de la precipitación en tiempos posteriores u otras variables que el modelador considere de relevancia en el estudio de la precipitación.

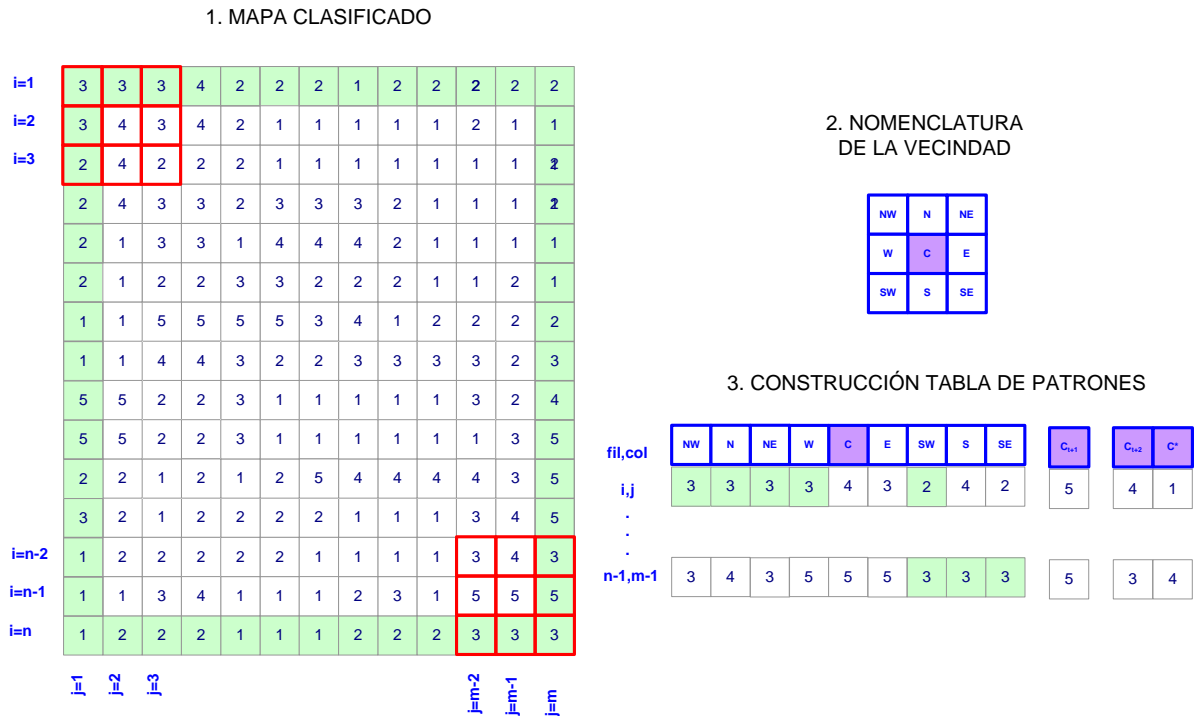


Figura 4-7. Construcción de la tabla de patrones para una vecindad Moore de radio 1.

La construcción del árbol de decisión en el conjunto de entrenamiento se toman en cuenta todos los patrones encontrados en el periodo de interés. El modelo de árbol sugerido es el ID3 debido a que es un algoritmo simple. Sus principales características son:

- Cada nodo corresponde a una variable (posición de la celda por ejemplo N, E, EE...) y cada rama al valor posible de esa variable (estado de la celda por ejemplo Seco, Ligero, Tormenta...). Una hoja del árbol especifica el valor esperado de la decisión de acuerdo con los patrones encontrados. La determinación del valor de la celda en interés viene dado por la trayectoria desde la raíz a la hoja representativa de esa decisión. En la Figura 4-8 se observa un ejemplo de un árbol tipo ID3 para etiquetas categorizadas por los rangos de precipitación y utilizando una vecindad de Moore.
- A cada nodo se le asocia aquella variable más informativa que aún no haya sido considerado en la trayectoria desde la raíz.
- Para medir que tanta información tiene una variable se emplea el concepto de entropía. Cuanto menor sea el valor de la entropía, menor será la incertidumbre y más útil será la variable para la clasificación.

En el árbol de la Figura 4-8 el nodo raíz corresponde a la celda E (Este) de la vecindad de Moore, para definir que ruta seguir en el momento de tomar una decisión existen dos caminos: que la posición “Este” tenga un valor de etiqueta “Aguacero4” o “Llovizna3” y así llegar al siguiente nodo. Este proceso se lleva a cabo sucesivamente hasta llegar a una hoja donde se toma la decisión del estado de la celda C en el tiempo $t + 1$.

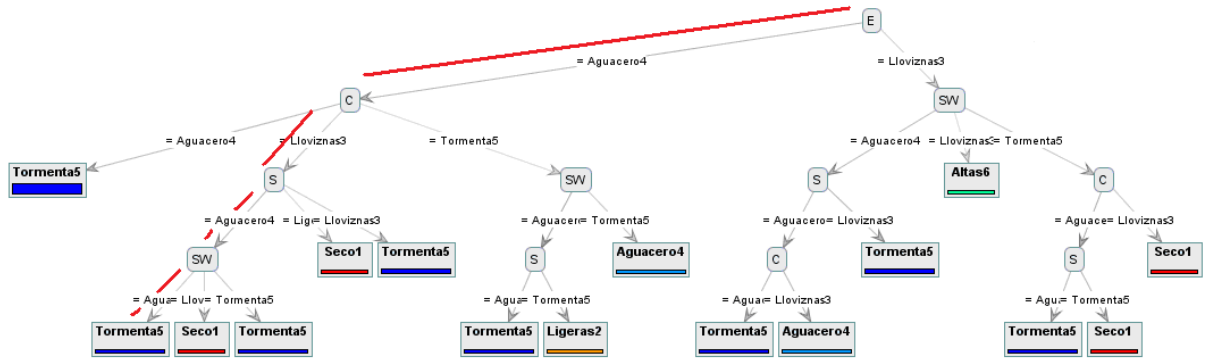


Figura 4-8. Ejemplo de un árbol de decisión con etiquetas que representan la categoría de precipitación.

Nota: Valores de las etiquetas 1=Seco1, 2=Ligeras2, 3=Ligeras3, 4=Aguacero4, 5=Tormenta5, 6=Altas6 y 7=Fuerte7

Tomando como referencia la línea roja que se resalta en el árbol de la Figura 4-8, la definición de la regla se expresa como:

if E = Aguacero4
 and C = Llovizna3
 and S = Aguacero4
 and SW = Aguacero4
 then Ct = Tormenta5

Que dicho en palabras:

- Si la vecindad en la posición **Este** tiene valor de **Aguacero4**
- Y** la vecindad en la posición **Centro** tiene valor de **Llovizna3**
- Y** la vecindad en la posición **Sur** tiene valor de **Aguacero4**
- Y** la vecindad en la posición **Suroeste** tiene valor de **Aguacero4**
- Entonces** el valor de la posición **Centro** en tiempo $t + 1$ es de **Tormenta5**

4.2.2 Reproducción del Modelo de Autómata Celular

Luego de tener definidas las reglas de evolución el paso a seguir es la “reproducción” del modelo de autómeta. A partir de unas condiciones iniciales equivalentes a un mapa clasificado se recorre uno a uno los pixeles o celdas de este con una plantilla o “vecindad” de igual geometría y tamaño con la que se construyó la tabla de patrones del numeral anterior.

En cada celda del mapa clasificado en el tiempo t se buscan las condiciones que cumplan alguna de las ramas del árbol de decisión, recorriendo uno a uno los nodos con sus respectivas etiquetas hasta llegar a una decisión. Para el ejemplo de la

Figura 4-9 se desea conocer el valor de la celda (F_4, C_5) en el tiempo $t + 1$, tomando como referencia el árbol de la Figura 4-8 se inicia por identificar el estado de la celda E, como este corresponde a 4 es decir Aguacero4, el siguiente nodo corresponde a la posición C “Centro” y así hasta llegar a la hoja que indica que la decisión del estado de la celda (F_4, C_5) en el tiempo $t + 1$ es 5 o “Tormenta5”. Este procedimiento se lleva a cabo en todos los pixeles del mapa de condiciones iniciales.

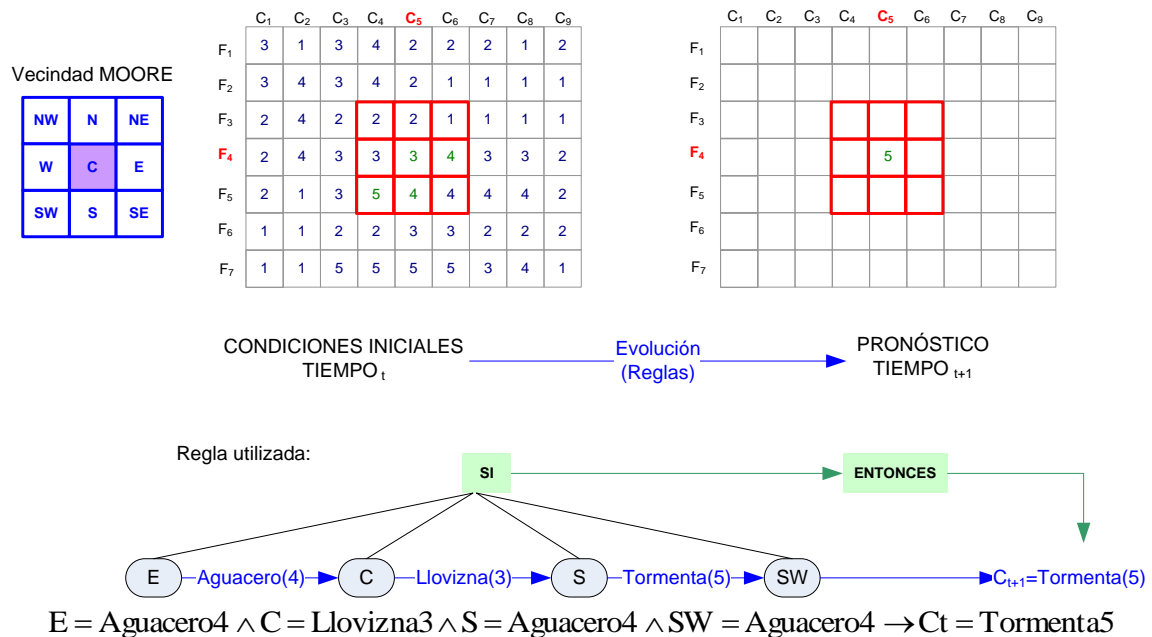


Figura 4-9. Ejemplo del pronóstico del estado de una celda

Para la estimación del pronóstico en cada instante de tiempo las condiciones iniciales pueden variar de dos formas:

- Actualizando las condiciones iniciales para cada instante de tiempo: Cada mapa pronosticado depende del mapa clasificado del tiempo inmediatamente anterior, siendo este la condición inicial (Ver Figura 4-10). De esta manera todas las condiciones iniciales serán los mismos mapas clasificados.
- Sin actualizar las condiciones iniciales para cada instante de tiempo: Las condiciones iniciales para el pronóstico de un tiempo específico, son las dadas por el mapa de pronóstico del tiempo inmediatamente anterior. La única condición inicial de pronóstico diferente es la primera, que parte del mapa clasificado en el tiempo cero (Ver Figura 4-11).

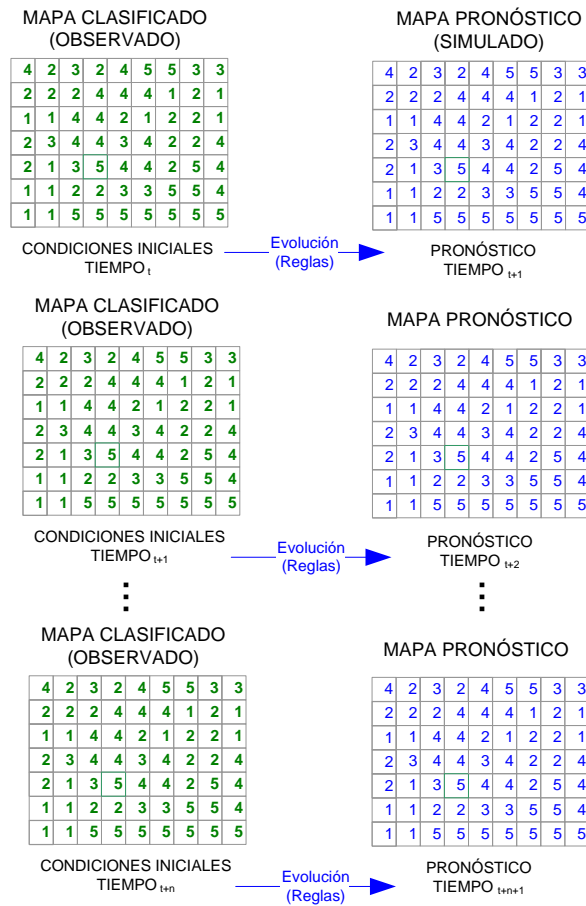


Figura 4-10. Asignación de condiciones iniciales para el caso de pronóstico actualizando las condiciones iniciales.

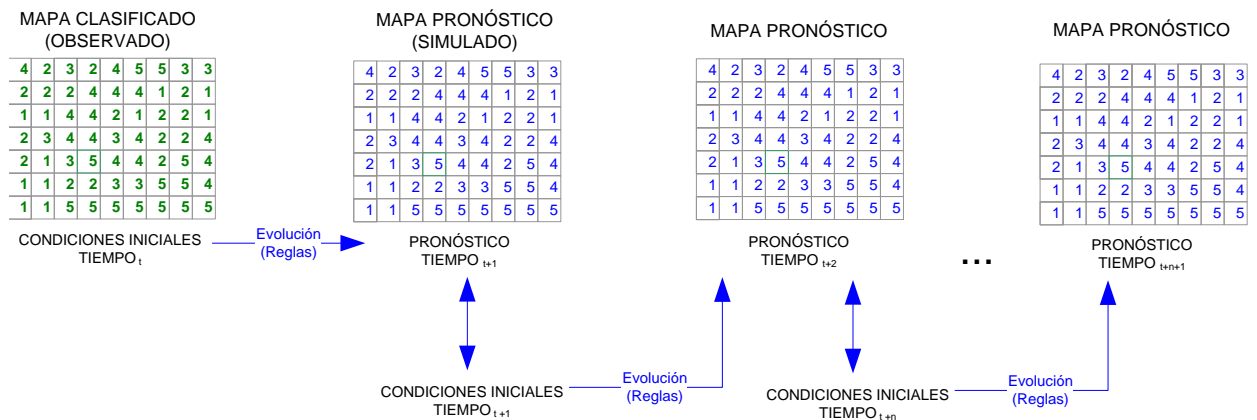


Figura 4-11. Asignación de condiciones iniciales para el caso de pronóstico sin actualizar las condiciones iniciales.

4.2.3 Evaluación del Desempeño del Modelo

La medida de desempeño indica qué tan eficiente es en el pronóstico. Pueden utilizarse varias formas de representarlo. Se propone evaluar el desempeño desde el punto de vista de la teoría de errores y desde indicadores de precipitación. Dentro de la teoría de errores en la Tabla 4-2 se muestran varios tipos de prueba que se pueden utilizar para medir el desempeño del modelo.

Tabla 4-2. Posibles medidas de desempeño del modelo de pronóstico de precipitación.

Medida de desempeño	Ecuación
Error absoluto	$EA = P_t^{obs} - P_t^{sim} $
Error relativo	$ER = \frac{ P_t^{obs} - P_t^{sim} }{P_t^{obs}}$
Raíz cuadrada de error medio (RMSE)	¡Error! No se pueden crear objetos modificando códigos de campo.
Porcentaje de Error Medio (PEM)	$PEM = \frac{1}{n} \sum_{i=1}^n \left(\frac{ P_t^{obs} - P_t^{sim} }{P_t^{obs}} \right) \cdot 100\%$
Coefficiente de Nash y Sutcliffe (NS)	$NS = 1 - \frac{\sum_{i=1}^n (P_t^{sim} - P_t^{obs})^2}{\sum_{i=1}^n (P_t^{obs} - \bar{P}_t)^2}$
Porcentaje BIAS	¡Error! No se pueden crear objetos modificando códigos de campo.

Fuente: Propuestas por la autora

donde n es el número de observaciones de precipitación P_t^{obs} es el valor observado de precipitación en el tiempo t ; P_t^{sim} es la precipitación simulada en el tiempo t ; y \bar{P}_t es la precipitación promedio de las observaciones.

Para cualquiera de las pruebas sugeridas se debe calcular el desempeño celda a celda. Por ejemplo, para el caso de la Figura 4-12 la medida de desempeño seleccionada es el error absoluto. Entonces se resta el valor de la celda del mapa clasificado (equivalente al valor observado) en la posición (i, j) , con el valor celda (i, j) en el mapa simulado por el autómatas celulares. Como resultado se observara un mapa en el que cada celda representa el error absoluto. Entre menor sea el valor de la celda mejor será la estimación del modelo de pronóstico. El error absoluto es la medida de desempeño utilizada en el presente estudio.

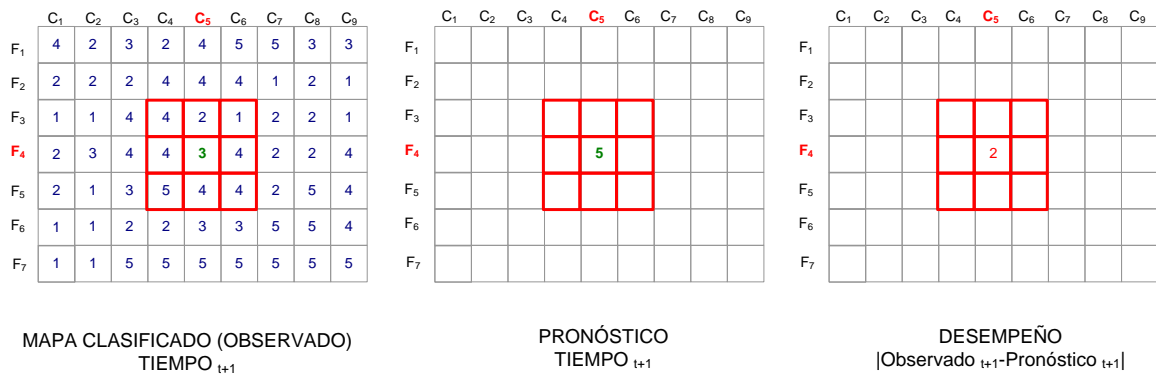


Figura 4-12. Ejemplo de la evaluación del desempeño del modelo de pronóstico de precipitación usando el Error Absoluto.

Otro indicador es la representación mediante un diagrama de dispersión. Este diagrama se construye comparando celda a celda los valores clasificados y pronosticados. En el eje x están los valores de las etiquetas de las celdas del mapa clasificado, en el eje y los valores de las etiquetas de las celdas del mapa de pronóstico y en el eje z están las frecuencias con las que se cumple la condición de valores de la pareja ordenada x_n, y_m (Ver Figura 4-13).

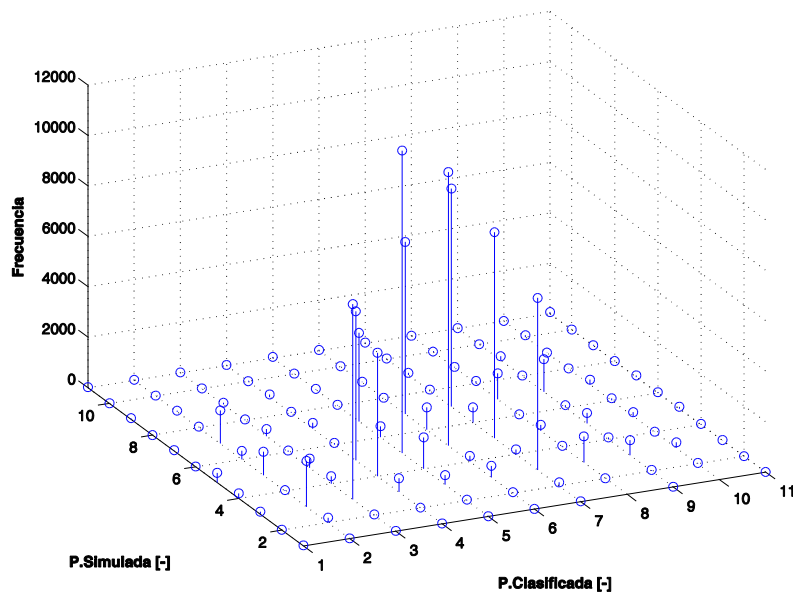


Figura 4-13. Diagrama de dispersión para indicar el desempeño del pronóstico.

Dentro de los indicadores de precipitación están:

- Precipitación total adimensional: este indicador se estima sumando todos los valores de las celdas del mapa clasificado y del mapa de pronóstico en cada uno de los instantes de tiempo en el que se realizó el pronóstico. Este indicador es adimensional pues la clasificación no tiene unidades (Figura 4-14).

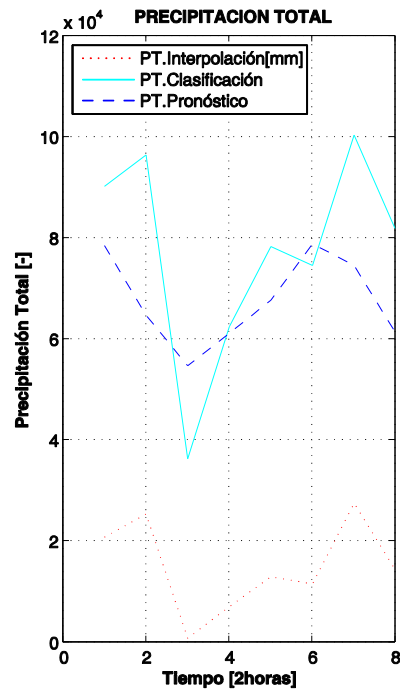


Figura 4-14. Comparación de la precipitación total en el mapa clasificado y el mapa de pronóstico.

- Rango de precipitación en milímetros: a partir de los rangos de precipitación asignados para las etiquetas de las celdas en la etapa de clasificación, se realiza el remplazo de los valores en estas en el mapa producto del pronóstico, teniendo en cuenta los límites superior, medio e inferior de los rangos establecidos. Luego se suman los valores de todas las celdas obteniendo un estimativo de la variación de precipitación máxima, media y mínima, creando así una franja de la variación de la precipitación total. Al encontrarse estos nuevos valores en unidades congruentes con el campo de precipitación se pueden comparar con la precipitación total del mapa interpolado. Tomando como referencia la clasificación de la Tabla 4-1 se muestra un ejemplo para la serie de celdas de la Figura 4-15. Se observa que el valor promedio del mapa pronóstico es de 18.25 mm y se acerca al valor de la precipitación total del campo de precipitación que corresponde a 17mm. Ver otro ejemplo en la Figura 4-16.

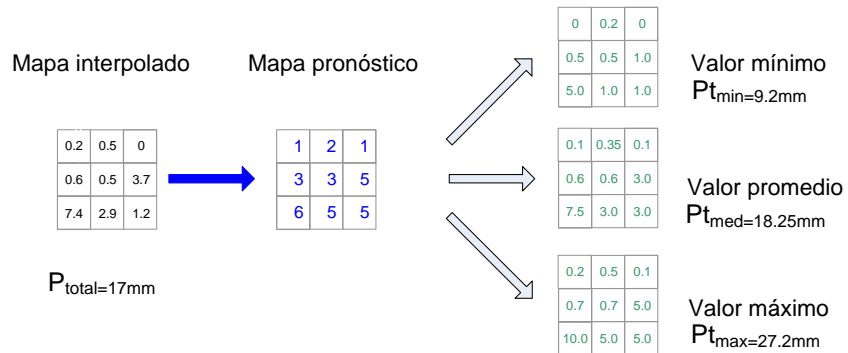


Figura 4-15. Reemplazo de los valores de etiquetas a valores de precipitación en milímetros de acuerdo a la clasificación asignada.

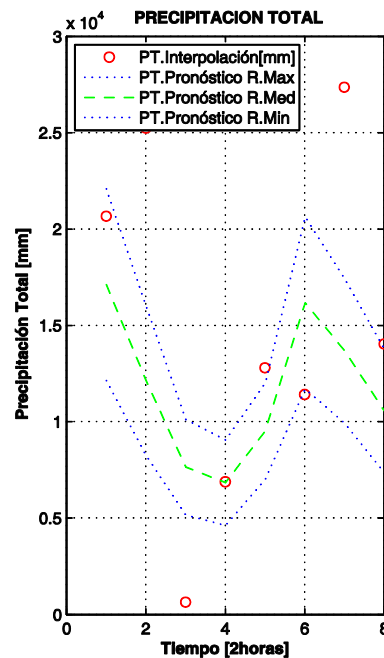


Figura 4-16. Comparación de la precipitación total en milímetros del campo de precipitación y los valores reemplazados en el mapa de pronóstico.

4.2.4 Análisis de Sensibilidad

El análisis de sensibilidad se sugiere realizar a través de la perturbación de “parámetros”, es decir, se varía un parámetro del modelo mientras el resto permanecen constantes, de modo que las variaciones sufridas en las variables de estado reflejan la sensibilidad de la solución al parámetro modificado. Los parámetros o condiciones que se pueden modificar son las siguientes:

- Vecindad: En la literatura son usadas tanto las vecindades pequeñas como las grandes. La resolución espacial de las celdas puede estar alrededor de los 100m, 250m, y 500m. De igual forma se puede variar el tipo de vecindad (Ver Figura 1-2 y Figura 4-5).
- Reglas de evolución: Utilización de otras técnicas de obtención de reglas como algoritmos genéticos o la variación del modelo de árbol de decisión.

Otras modificaciones que se pueden hacer en la modelación pero que representan un cambio de modelo y no específicamente un cambio de parámetros son:

- Reloj virtual: Observar en qué resolución temporal los resultados son mas acertados.
- Configuración: Se pueden probar varios tipos de clasificación, ello debido a que no existe certeza de cuál es la mejor agrupación de los datos.

5 IMPLEMENTACIÓN DE LA METODOLOGÍA PARA EL MODELAMIENTO DE LA PRECIPITACIÓN CON UN AUTÓMATA CELULAR

Quizás una de las metas de la investigación en hidrológica es plantear nuevas alternativas de modelos que reproduzcan el comportamiento de las variables de interés de los fenómenos que se observan en la naturaleza. Algunos investigadores resaltan las bondades de la modelación de los procesos hidrológicos tomando en cuenta el componente espacial. En ese sentido, los denominados modelos distribuidos incorporan con mejor aproximación la variabilidad espacial de las diferentes variables, dando la posibilidad al entendimiento de los procesos que tienen lugar dentro la cuenca.

De acuerdo con Brown et al. (2002), para que los modelos predictivos se transformen en una herramienta útil es necesario que representen de manera eficaz la magnitud de los cambios, la localización de los futuros cambios, y los patrones espaciales de estos cambios. En la presente investigación se implementó una metodología de pronóstico de la precipitación basada en la teoría de autómatas celulares. Este modelo por definición incorpora el componente geográfico y la variabilidad espacial de las variables representadas en un espacio bidimensional. Una de las ventajas de la metodología expuesta es que los resultados serían de fácil acople con modelos hidrológicos distribuidos. Además, la visualización en forma de mapas permite observar con más facilidad la variabilidad espacial y temporal de la precipitación.

La implementación de la metodología expuesta se aplicó tanto a nivel de agregación temporal diaria como para eventos de tormenta con diferentes agregaciones temporales. A continuación se muestra del desarrollo de la implementación de la metodología para el modelo de pronóstico con autómatas celulares en el caso de estudio de la ciudad de Bogotá.

5.1 PROCESAMIENTO DE LA INFORMACIÓN

Como ya se expuso en el capítulo anterior esta etapa de la metodología comprende el procesamiento de la información que incluye varias fases: 1) la caracterización geográfica e hidrológica de las series de datos, 2) la delimitación del área de interés, 3) Análisis exploratorio de datos y 4) la obtención de los campos de precipitación. Estos análisis se llevaron a cabo a partir de la información de precipitación horaria y diaria de estaciones localizadas en el casco urbano y alrededores de la ciudad de Bogotá.

En el capítulo 2 “CARACTERIZACIÓN DE LA ZONA DE ESTUDIO E INFORMACIÓN DISPONIBLE (Pág. 20)”, se encuentra desarrollada la caracterización geográfica e hidrológica junto con los fundamentos del Análisis Exploratorio de Datos. De manera resumida del resultado del AED de las series de precipitación total diaria de las 77

estaciones localizadas en la ciudad de Bogotá para los años 1995 a 1999, se identificó la temporada de lluvias en el que mayor número de estaciones registraron los valores de precipitación. Este período está comprendido entre los meses de marzo, abril y mayo. En el Apéndice A, se puede ver la caracterización de las series, mostrándose el análisis estadístico, la localización y gráficos para visualizar su dispersión. La selección de las tormentas se hizo tomando como referencia el estudio de Bernal (2008).

En el capítulo 3 “ESTIMACION DE LOS CAMPOS DE PRECIPITACIÓN (Pág. 61)” se hacen algunas sugerencias de las consideraciones que se deben tener en cuenta para la construcción de los campos de precipitación. Como resultado de este capítulo se estableció un tamaño de celda igual a 500 metros y una alineación de malla en la dirección norte-sur. Bajo estas especificaciones el área de estudio está dispuesta por un espacio bidimensional de 50 kilómetros en dirección este-oeste y 70 kilómetros en dirección norte-sur. La validación cruzada determinó que el mejor método de interpolación es el IDW.

5.2 MODELO DE AUTÓMATA CELULAR

En los modelos de AC, la transición de una celda de un estado a otro está basada en la regla de evolución y el tipo de vecindad adoptada. En ese sentido, es de suma importancia definir los elementos del autómata, siendo de gran interés identificar la técnica de obtención de reglas de evolución y el tipo de vecindad a utilizar. A continuación se muestran las consideraciones que se tuvieron en cuenta para implementar el modelo de pronóstico tanto a nivel diario como en los eventos de tormenta.

5.2.1 Definición de los Elementos del Autómata Celular

La definición de los elementos del autómata como el plano y el tamaño de celda son definidos desde la etapa de “Procesamiento de la información” cuando se construyen los campos de precipitación. El reloj virtual está en función del nivel de agregación temporal de los datos (día, hora o minuto). Los demás elementos son definidos de la siguiente manera:

5.2.1.1 Vecindad

Se utilizaron dos tipos de vecindad: 1) Vecindad de Moore con radio 1, representada por una plantilla de 3*3 para un total de nueve posiciones y 2) Vecindad de Von Neuman con radio 2 con un total de 13 celdas. (Ver Figura 5-1). Los resultados finales de la investigación muestran el pronóstico de precipitación al haber utilizado la plantilla tipo Moore.

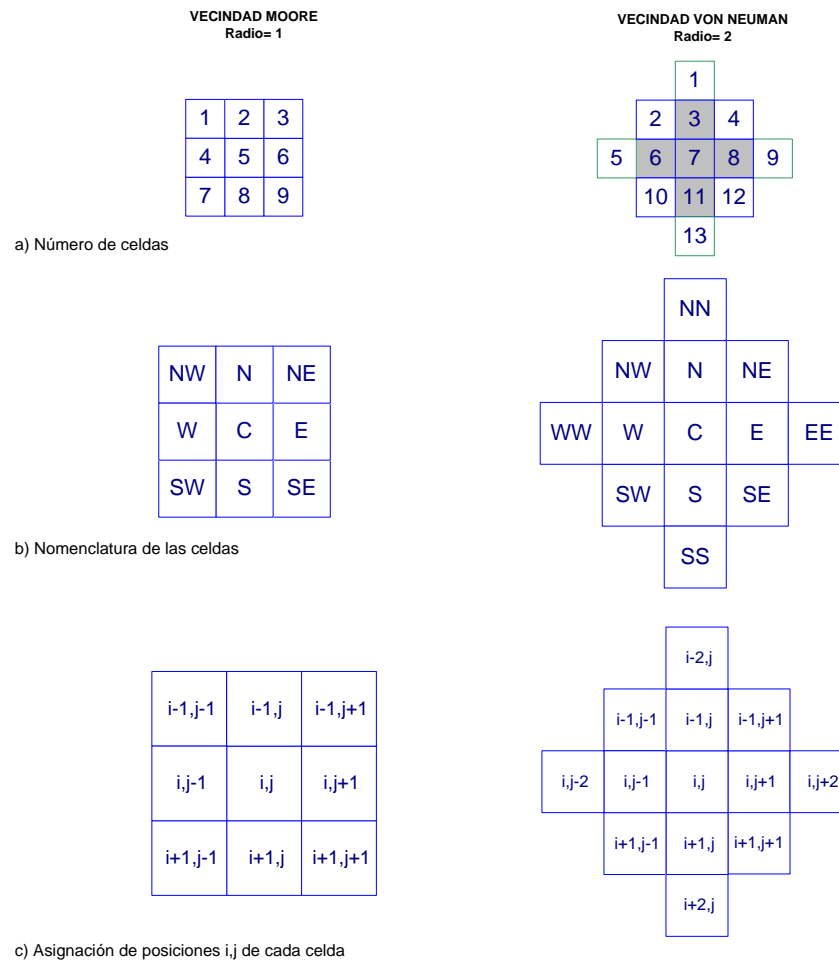


Figura 5-1. Descripción de las plantillas utilizadas en el modelo de autómatas celulares.

5.2.1.2 Configuración

La configuración o estado de las celdas se determina cuando a cada una de las celdas del campo de precipitación se le asigna un valor de etiqueta. Las etiquetas son establecidas por unos rangos de clase. Es difícil establecer un único modo de clasificar la precipitación, por ello se plantearon siete tipos de clasificaciones en el caso de la precipitación diaria y uno para el caso de los eventos.

En la Tabla 5-1 se observa la asignación de las etiquetas de acuerdo con la variación de los rangos establecidos para los valores de precipitación diaria interpolada encontrados en los campos de precipitación. Se presenta un caso especial dentro de las clasificaciones, la clasificación tipo 7 surge debido a que al tener gran cantidad de valores de precipitaciones bajas o nulas es frecuente la concentración de registros hacia estos valores (Ver los histogramas en el Apéndice A), así que se optó por realizarle una transformación logarítmica a los datos de la precipitación interpolada con la hipótesis de que ello mejoraría la clasificación, en el sentido de poder tener más detalle en los valores de baja precipitación.

Tabla 5-1. Asignación de etiquetas de acuerdo a los rangos de precipitación diaria [mm].

Tipo de clasificación	Definición del rango y etiquetas
Tipo 1	RANGO=[0 0.2 0.5 0.7 1.0 5.0 10.0 100] CLASIFICADOR=[1 2 3 4 5 6 7]
Tipo 2	RANGO=[0 0.3 1.0 10 20 30 40 50 60 70 80 90 100] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10 11]
Tipo 3	RANGO=[0 4.8975 16.45179 28.0060 39.5602 51.1144 62.6686 74.2228 85.7771 97.3313] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10]
Tipo 4	RANGO=[0 9.8 20 30.2 40.9 54.1 76 100] CLASIFICADOR=[1 2 3 4 5 6 7]
Tipo 5	RANGO=[0 1 5 10 15 20 25 30 35 40 45 50 55 90] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10 11 12 13]
Tipo 6	RANGO=[0 0.1 0.3 1 1.8 2.83 4 5.2 7.7 11 17.16 76] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10 11]
Tipo 7*	RANGO=[1 1.10 1.20 1.30 1.40 1.50 1.60 1.70 1.80 1.90 2] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10]

*El rango de los valores de precipitación esta en unidades de logaritmo milímetros.

En el caso de la precipitación a nivel horario y subhoraria se estableció una sola clasificación para todos los eventos, pero con variaciones en los rangos según la agregación temporal (Ver Tabla 5-2).

Tabla 5-2. Asignación de etiquetas de acuerdo a los rangos de precipitación [mm] por agregación temporal en evento.

Resolución temporal	Definición del rango y etiquetas
2 horas	RANGO=[0 0.008 0.028 0.146 0.5 0.999 1.432 2.316 2.851 4.353 5.963 24.3] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10 11]
1 hora	RANGO=[0 0,004 0,017 0,119 0,370 0,600 1,133 1,813 2,33 3,65 24,3] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10]
30 minutos	RANGO=[0 0,00219 0,008 0,0355 0,149 0,3 0,517 0,8 1,250 1,908 24,3] CLASIFICADOR=[1 2 3 4 5 6 7 8 9 10]

5.2.1.3 Frontera

Con el fin de mantener de una manera más fiel los valores de las celdas en los bordes y así evitar aumentar la incertidumbre de los valores en los bordes, la frontera se establece como “conocida”. Esto se puede asumir si en el momento de realizar la tabla de patrones y la reproducción de modelo de autómatas, se inicia en la posición 2,2 en el caso de la vecindad Moore o en la posición 3,3 para la vecindad de Von Neuman, como se puede ver en la Figura 5-2.

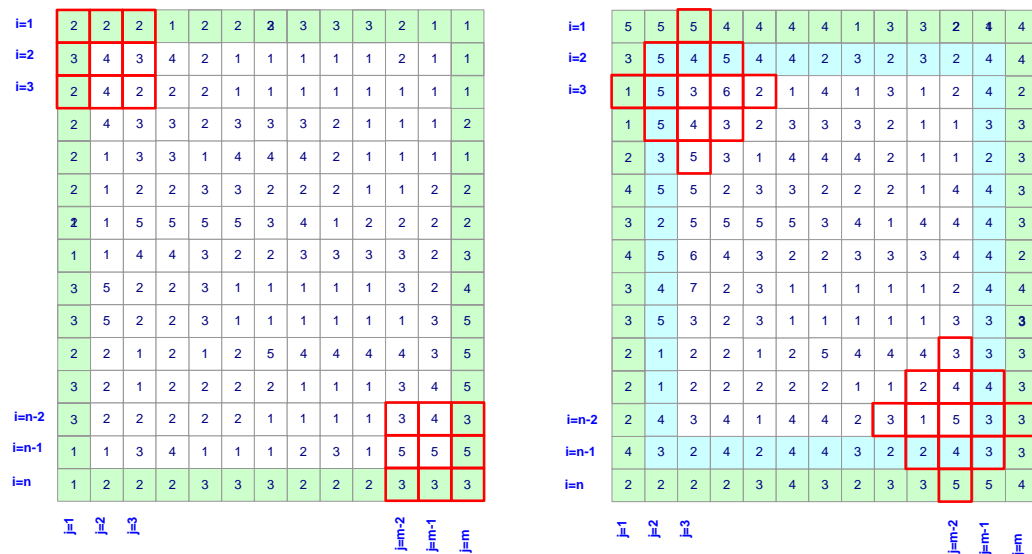


Figura 5-2. Establecimiento de la condición de frontera para la vecindad de Moore y Von Neuman.

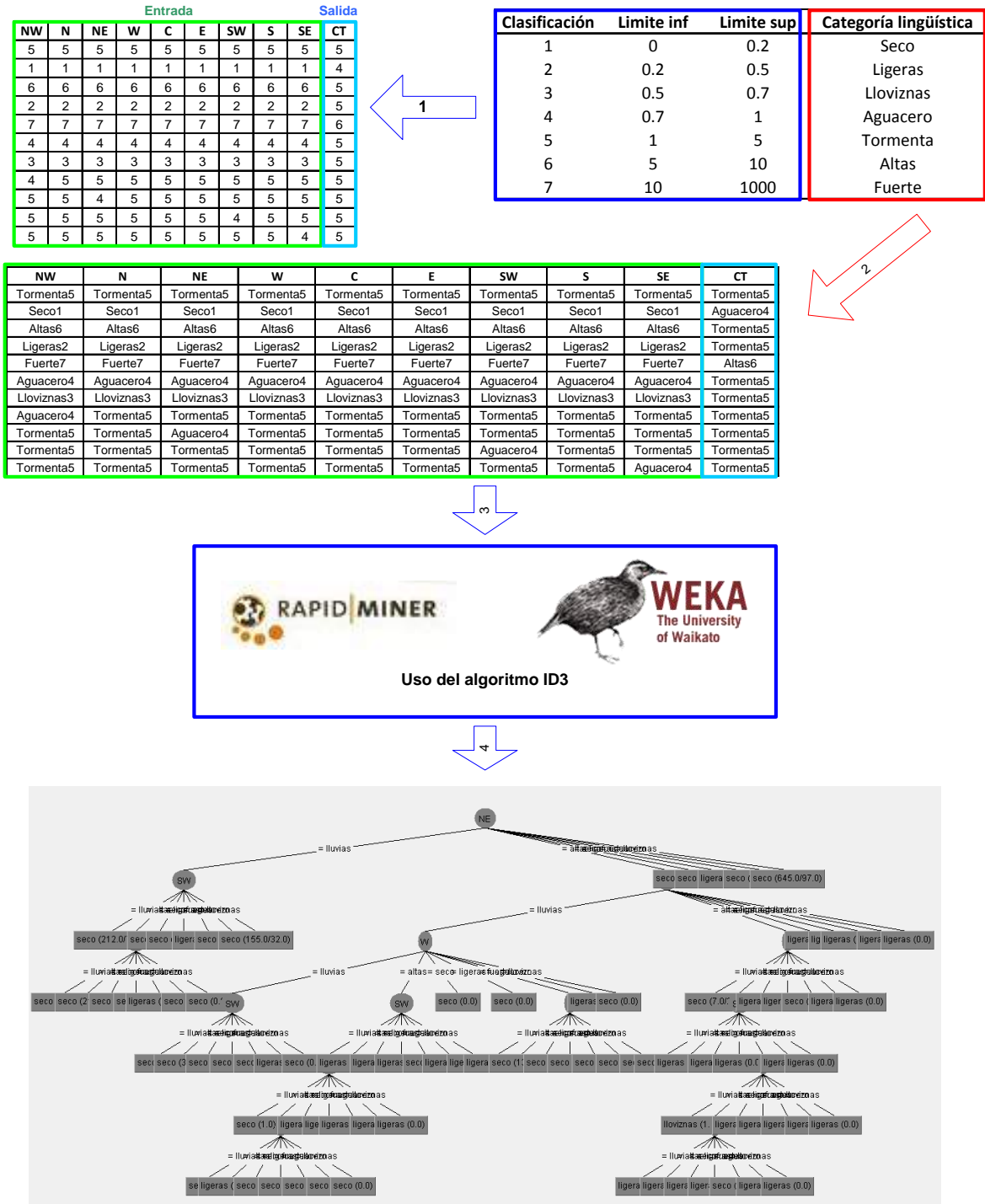
5.2.1.4 Reglas de Evolución

El reconocimiento de los patrones de distribución espacial de las precipitaciones es necesario, pues con base a éste se propone el método de análisis de los datos. Mediante la utilización de técnicas encaminadas a la extracción de conocimiento de los mapas clasificados o la tabla de patrones se busca tener un criterio de “conocimiento experto” para la identificación de patrones y reglas del comportamiento de la precipitación.

Usando una vecindad de Moore la tabla de patrones está representada por una matriz de diez columnas, donde los valores de la columna 1 hasta la columna 9 representan el valor de las nueve posiciones de la vecindad en una celda específica a este conjunto de celdas se les denominó “Entrada”. La columna diez indica el valor de la celda central de la vecindad en la misma posición pero en el tiempo inmediatamente después y se llamó “Salida”. En el caso de estudio para un tamaño de celda de 500 metros y una área de 50*70 kilómetros, el número de filas de la tabla de patrones usando una vecindad de Moore para un instante de tiempo es igual a 13524 (resultado de multiplicar 138 filas * 98 columnas). En el caso de usar la vecindad de Von Neuman el número de columnas es 14 y de filas 13056.

En el mercado existen varios aplicativos computacionales de uso libre especializados que llevan a cabo aplicaciones usando técnicas de minería de datos. En la presente investigación para la construcción del árbol de decisión se usó el software Weka version 3-6 y RapidMiner (<http://www.cs.waikato.ac.nz/ml/weka/> y <http://rapid-i.com/>, respectivamente). El software utilizado para la construcción de la tabla de patrones es MATLAB® versión 2009a.

El tipo de algoritmo de árbol de decisión seleccionado es el ID3. Para ello el software requiere que la tabla de patrones este en categorías lingüísticas y no con valores numéricos (Ver Figura 5-3).



El proceso de construcción se la tabla de patrones para la precipitación diaria se hizo para cada año por separado y después se conjugó en una sola tabla para introducirlo al software de minería de datos. En la Figura 5-4 se muestra la variación del número de patrones por año según la vecindad adoptada. Se observa que sin importar el tipo de vecindad utilizada el año con mayor variabilidad es el año 1995 y el de menor el 1997. Cabe anotar que la construcción de la tabla de patrones para la vecindad de Von Neuman se dificulto para algunas clasificaciones, por lo que se recomienda usar otros lenguajes de programación o sistemas operativos más eficientes. Al final del documento en el compendio de algoritmos, se encuentra el código desarrollado en MATLAB® para la construcción de la tabla de patrones.

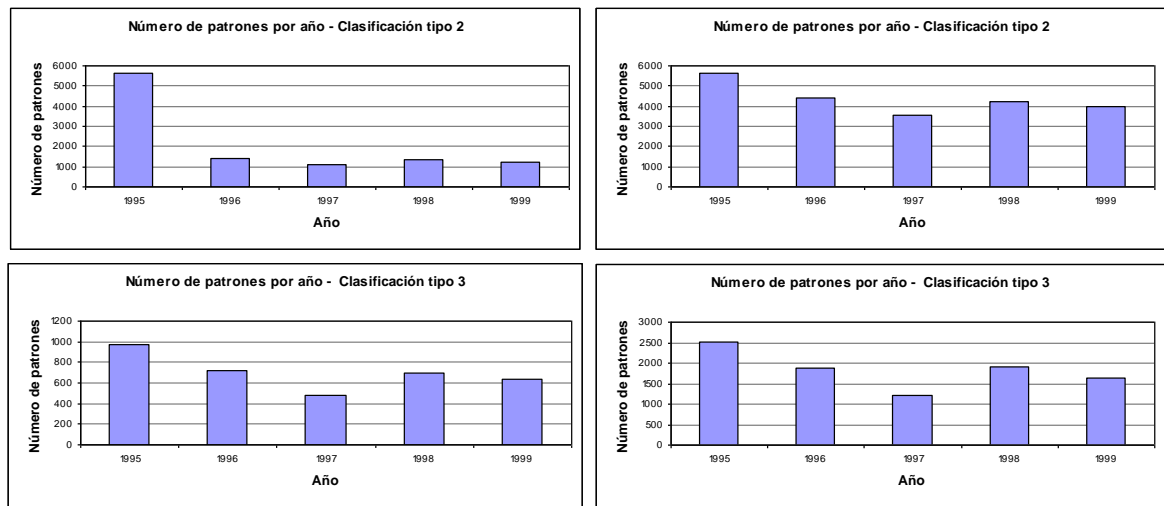


Figura 5-4. Variación de la cantidad de patrones por año según la vecindad.
Nota: Derecha usando vecindad Moore, izquierda vecindad Von Neuman.

Es frecuente encontrar en la tabla de patrones comportamientos idénticos en las celdas de “entrada” con diferente salida, por ejemplo la entrada con configuración 1,1,1,1,1,1,1,1 con salida 1 y la misma configuración de entrada con salida 3 (Ver Tabla 5-3). Esta situación genera conflicto en el momento de construir el árbol de decisión, para darle solución se plantearon dos alternativas:

Tabla 5-3. Ejemplo de patrones con igual valor en la entrada pero diferente salida.

Patrón	Entrada	Salida	Frecuencia
1111111111	111111111	1	1064314
1111111112	111111111	2	58364
1111111113	111111111	3	3359
1111111114	111111111	4	340
1111111115	111111111	5	44
1111111121	111111112	1	418
1111111122	111111112	2	110
1111111123	111111112	3	15
1111111124	111111112	4	1

- Alternativa 1: Hacer una variación en la lectura de la celda en la posición central en el mapa clasificado en el tiempo $t + 1$ de manera que no lea ese valor, sino el valor de la misma celda en el tiempo $t + 1$ pero en el mapa interpolado. Al final de recorrer todas las celdas en todo el periodo de interés se procede a estimar el valor de la celda de salida para una misma configuración de las celdas de la “Entrada” haciendo un promedio de valores encontrados. Luego se identifica en que rango se encuentra el promedio y se le asigna la etiqueta de acuerdo con la clasificación con la que se esté trabajando.
- Alternativa 2: Elegir el patrón con mayor frecuencia

Con estas dos alternativas la tabla de patrones y por ende los posibles modelos para el pronóstico a nivel diario son:

Tabla 5-4. Número de modelos propuestos para el pronóstico de precipitación diaria, en función de la estimación de los patrones.

ID modelo	Modelo	Tipo de clasificación	Alternativa resolución conflicto de patrones
1	Clase1	Tipo 1	1
2	Clase2	Tipo 2	1
3	Clase3	Tipo 3	1
4	Clase4	Tipo 4	1
5	Clase5	Tipo 5	1
6	Clase6	Tipo 6	2
7	Clase7	Tipo 7*	2
8	Moda1	Tipo 1	2
9	Moda2	Tipo 2	2
10	Moda3	Tipo 3	2
11	Moda4	Tipo 4	2
12	Moda5	Tipo 5	2

En el caso de los eventos se tomo en cuenta la alternativa número dos.

Una vez solucionado los conflictos de la tabla de patrones, el paso a seguir es asignarle a los valores numéricos de las etiquetas categorías lingüísticas. De esta manera se procede a introducirlos al aplicativo computacional de minería de datos. Se eligió el algoritmo Id3, tomando todos los valores para el entrenamiento. En la Tabla 5-5 se muestran el número de patrones encontrados para cada modelo propuesto para el pronóstico de precipitación diaria y el número de reglas obtenidas por el árbol de decisión. El pronóstico se realizó para los meses de marzo, abril y mayo de los años 1995 a 1999 excepto para el modelo “Clase6” donde la capacidad de computo no permitió obtener una tabla de patrones para los años 1996 a 1999, por lo que solo se presenta el pronóstico de los meses de marzo, abril y mayo de 1995.

Tabla 5-5. Número de patrones y reglas por clasificación para una plantilla Moore de radio 1. Reglas para el pronóstico de la precipitación diaria.

Modelo	Tipo de clasificación	Número de patrones	Número de reglas	Reducción de información
Clase1	Tipo 1	9176	6414	30%
Clase2	Tipo 2	9176	6500	29%
Clase3	Tipo 3	3501	1477	58%
Clase4	Tipo 4	1505	463	69%
Clase5	Tipo 5	6793	4904	28%
Clase6*	Tipo 6	13552	8775	35%
Moda1	Tipo 1	6743	5218	23%
Moda2	Tipo 2	1898	1095	42%
Moda3	Tipo 3	985	445	55%
Moda4	Tipo 4	989	269	73%
Moda5	Tipo 5	4129	3061	26%

5.2.2 Reproducción del Modelo de Autómata Celular

Teniendo como referencia las reglas del numeral anterior se realizó pronóstico de la precipitación tanto a nivel diario como tormenta. La programación se hizo en el software MATLAB® 2009b. El tiempo computacional de cada modelo por año se muestra en la Tabla 5-6, para un total de aproximadamente 150 días de computo. Es claro que los modelos cuyas clasificaciones no discretizan detalladamente los valores bajos se demoran menos en hacer los cálculos. Al final de documento en el compendio de algoritmos se encuentra el código elaborado para la ejecución del autómata.

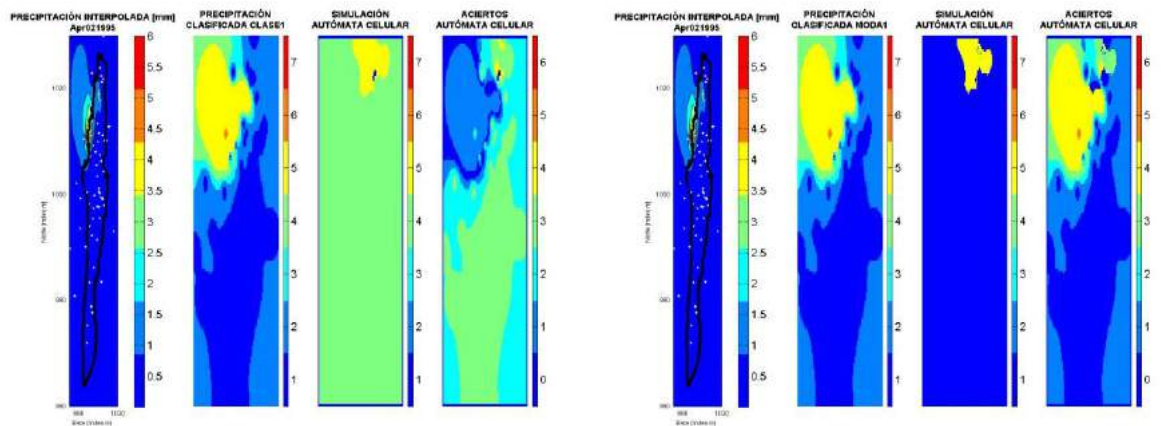
Tabla 5-6. Tiempo de cómputo del pronóstico de precipitación diaria para los meses de marzo, abril y mayo.

Modelo	1995	1996	1997	1998	1999
CLASE1	5.30	5.30	5.30	5.30	5.30
CLASE2	5.90	5.90	5.90	3.83	3.83
CLASE3	0.25	0.26	0.14	0.36	0.19
CLASE4	0.05	0.02	0.01	0.04	0.02
CLASE5	3.97	2.47	3.87	2.76	5.36
CLASE6	3.97	3.97	3.97	3.97	3.97
CLASE7	3.25	3.25	3.25	3.25	3.25
MODA1	3.26	3.26	2.63	2.63	5.57
MODA2	5.90	5.90	0.59	0.40	0.31
MODA3	0.25	0.26	0.14	0.36	0.19
MODA4	0.04	0.01	0.01	0.01	0.03
MODA5	2.17	1.59	2.00	3.14	1.63

Nota: Solo incluye la etapa de pronóstico.

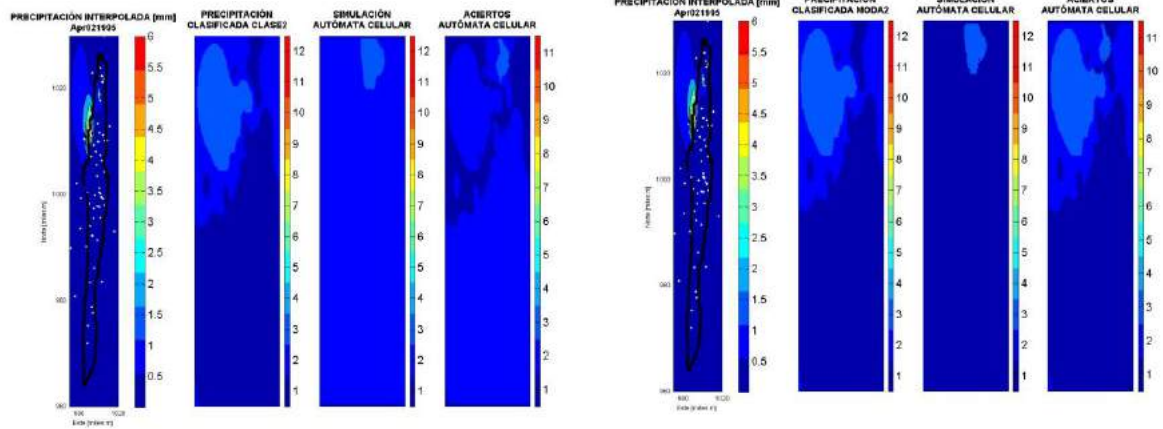
A manera de ejemplo la Figura 5-5 se muestra el pronóstico para el día 02 de marzo del año 1995 tomando como condición inicial el 01 de marzo del mismo año, el pronóstico esta realizado por varios modelos. En la primera columna de cada grafica esta la precipitación interpolada en milímetros, la segunda columna muestra la precipitación clasificada de acuerdo el tipo de clasificación, la tercera muestra el pronóstico.

En la cuarta columna se encuentra el mapa de “aciertos” que corresponde al error absoluto. Para la lectura de ese mapa los colores azules oscuros corresponden a errores de nulos a bajos, colores verdes errores medios y colores calidos errores altos.



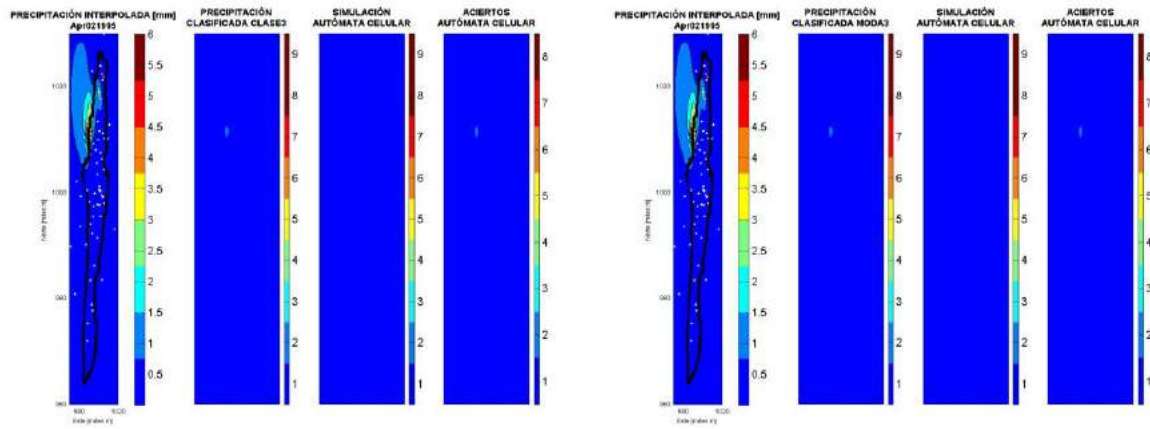
Clase 1

Modal



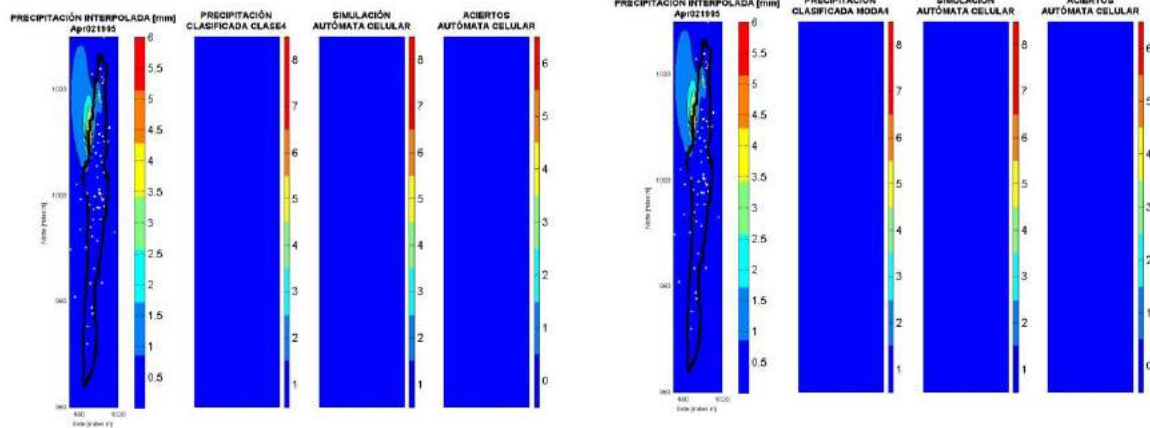
Clase2

Moda2



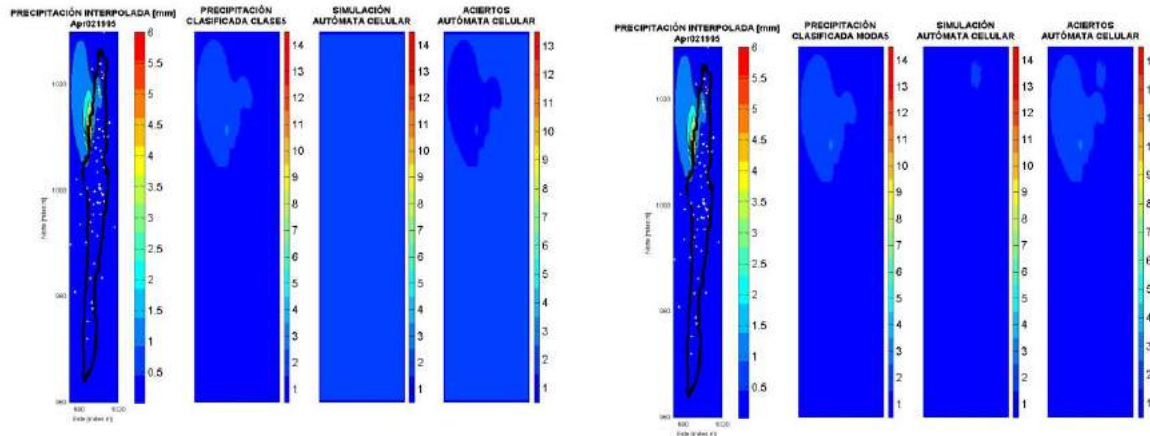
Clase3

Moda3



Clase4

Moda4



Clase5

Moda5

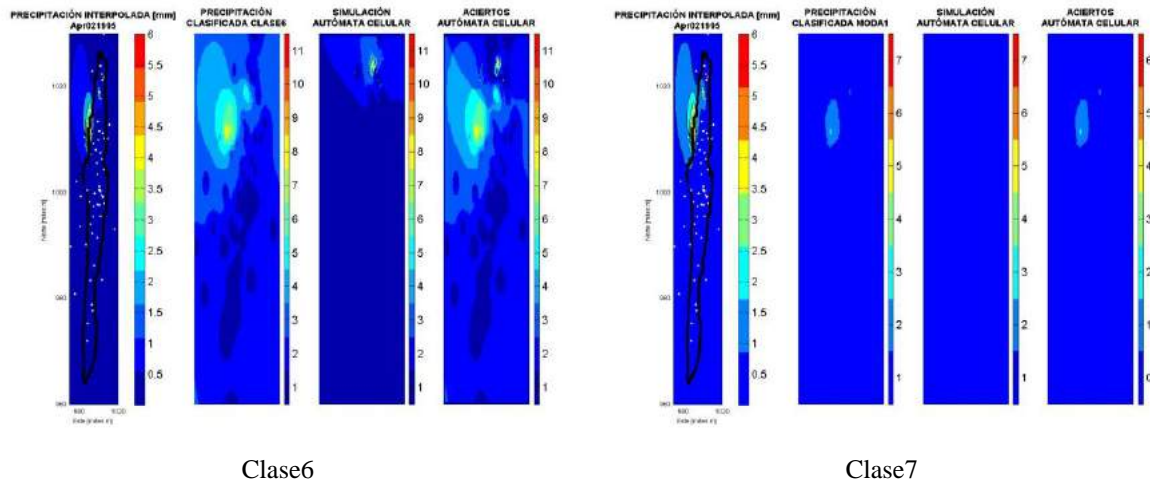
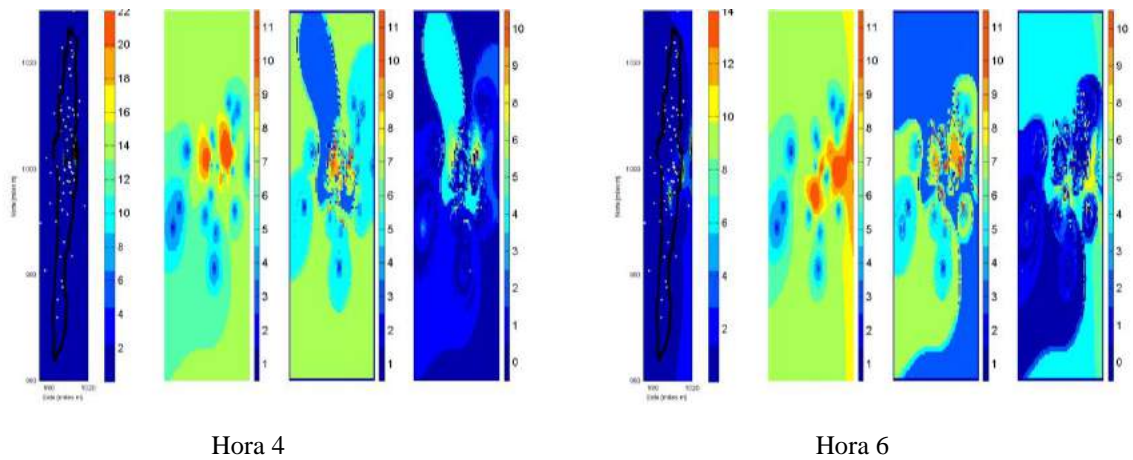


Figura 5-5. Ejemplo de pronóstico de precipitación diaria para el 2 de marzo de 1995 usando diferentes modelos.

En la

Figura 5-6 se muestra la evolución del pronóstico para un evento de tormenta que inicia el 12 de abril de 1995 a las 10 horas y termina el día siguiente a las 4 horas. Tanto el ejemplo de pronóstico de precipitación diaria como a nivel de evento se realizó actualizando las condiciones iniciales cada instante de tiempo (Figura 4-10) y vecindad de Moore.



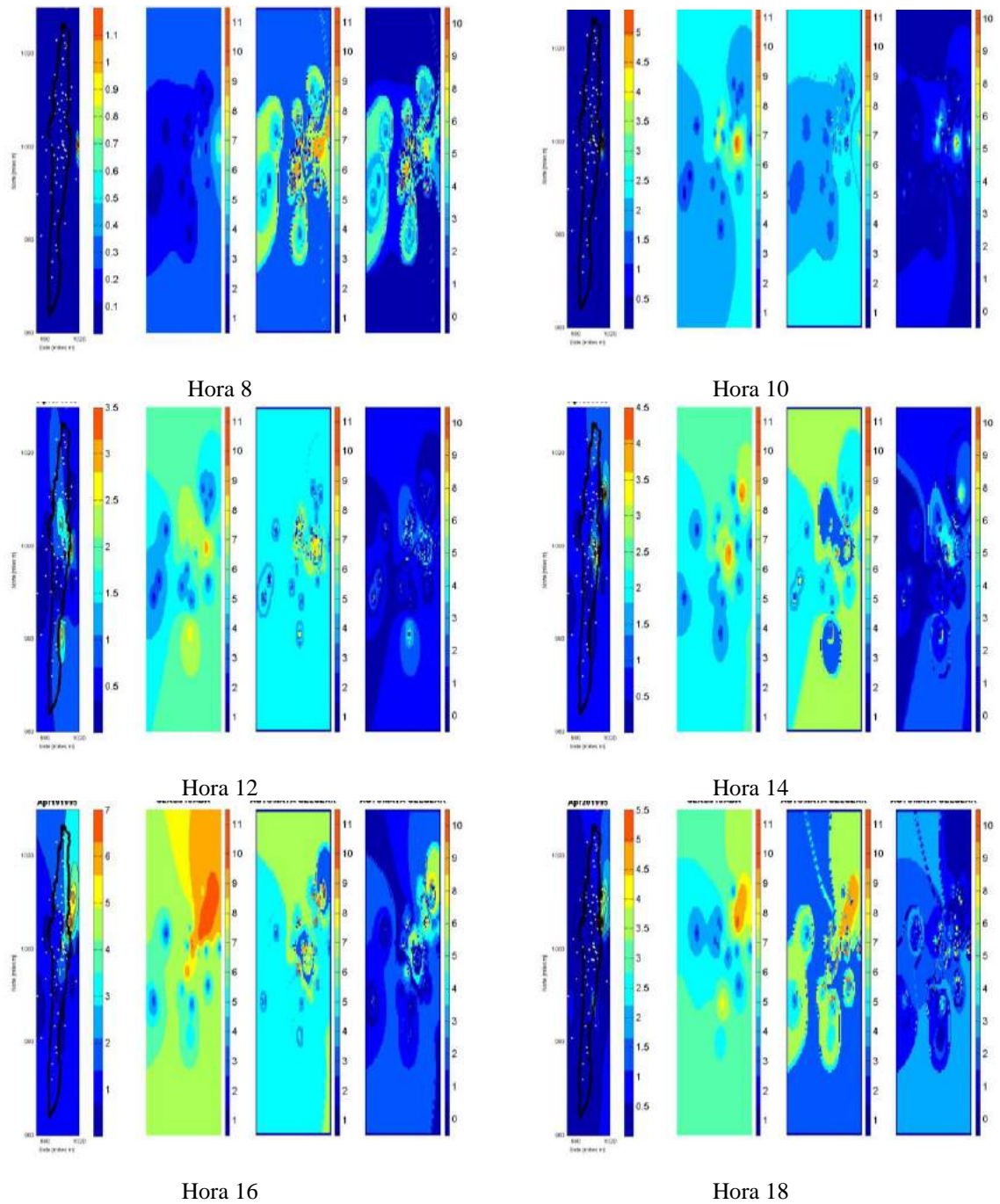


Figura 5-6. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 desde las 10 am y terminando el 13-abril-1995 a las 4 am. Resolución temporal 2 horas.

5.2.3 Desempeño del Modelo

En cada uno de los mapas mostrados en las anteriores figuras se puede observar una de las medidas de desempeño descritas anteriormente, el error absoluto. Es bueno aclarar que en términos generales un valor de error absoluto bajo es bueno, pero es importante recalcar que el cálculo del error está en función del mapa clasificado, por lo que se recomienda prestar atención a los tipos de clasificación a seleccionar.

Para las clasificaciones como la Clase 1, Moda 1 y Clase 6 es clara la variación espacial del error absoluto, mientras que para las otras clasificaciones es frecuente encontrar un error absoluto bajo, esto debido a que las clases de esos modelos agrupan gran parte de los valores bajos de precipitación y el mapa de clasificación es por lo general de un mismo valor de etiqueta. En el caso de la tormenta del 12 de abril los errores absolutos más altos se presentan en las celdas donde debería ocurrir un valor de precipitación clasificada alta, de lo que se infiere que al modelo se le dificulta reproducir los valores altos de la precipitación.

Una de las inquietudes que surgen ante la evaluación del desempeño es conocer que componente preserva el modelo. En ese sentido, se propone mostrar los valores de precipitación total tanto para las series diarias como para el caso de las tormentas y se comparan con los valores de precipitación total observada. La Figura 5-7 muestra la precipitación total en cada instante de tiempo de la tormenta del 12 de abril de 1995. En la gráfica superior están consignados los valores de precipitación total para el mapa clasificado y el pronosticado. Esta es una comparación de valores adimensionales pues se suman valores de etiquetas. Aun así se muestra la precipitación total interpolada con el fin de comparar la textura de las gráficas y así observar si se preservan las variaciones de agua precipitada.

En la gráfica localizada en la parte inferior se transforman los valores de las celdas del mapa clasificado y pronosticado respecto a los rangos de etiquetas. De esta manera al realizar la suma las dimensiones pueden ser comparadas con la precipitación total del mapa interpolado. Este tipo de análisis es conveniente en el sentido de que logra dar un orden de magnitud de la posible precipitación que se va a presentar en un instante de tiempo.

Los diagramas de dispersión son otra manera de evaluar el desempeño del modelo de pronóstico. En ellos se compara el valor de la celda en el mapa clasificado en la posición (i, j) versus el valor de la misma celda en el mapa de pronóstico. Adicionalmente se muestra la frecuencia con la que se presentó una condición dada.

La Figura 5-8 muestra la secuencia en el tiempo de los diagramas de dispersión para la tormenta de abril 12 de 1995 con una agregación temporal de 2 horas (Ver Figura 5-6).

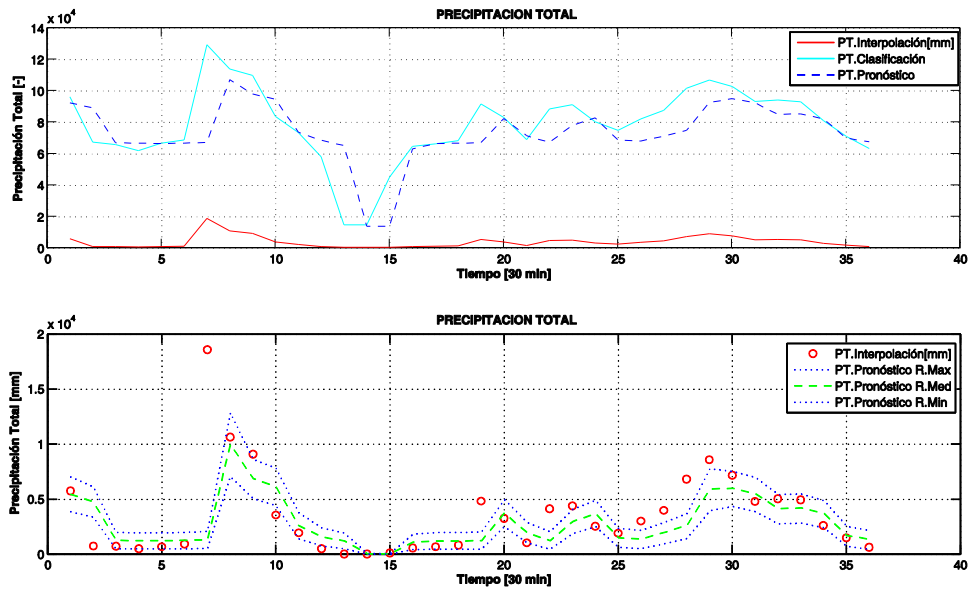
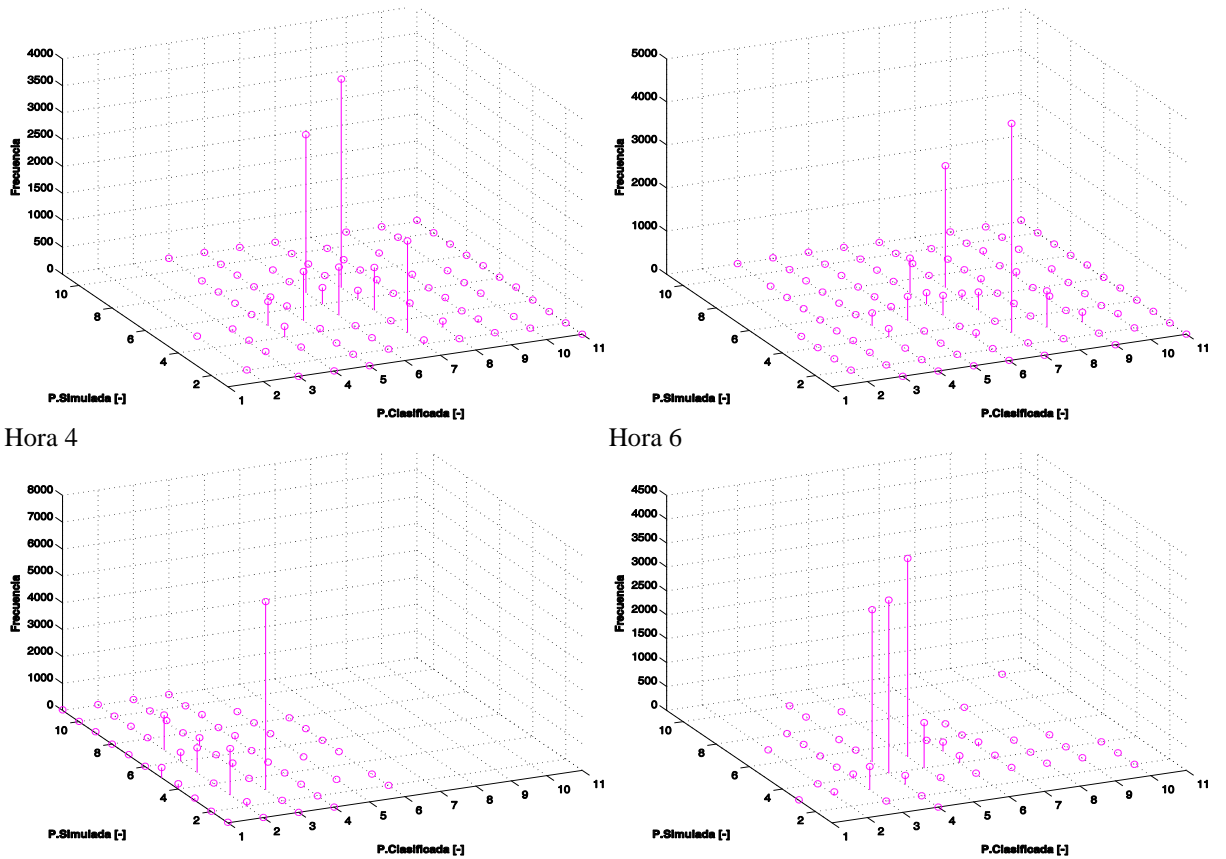
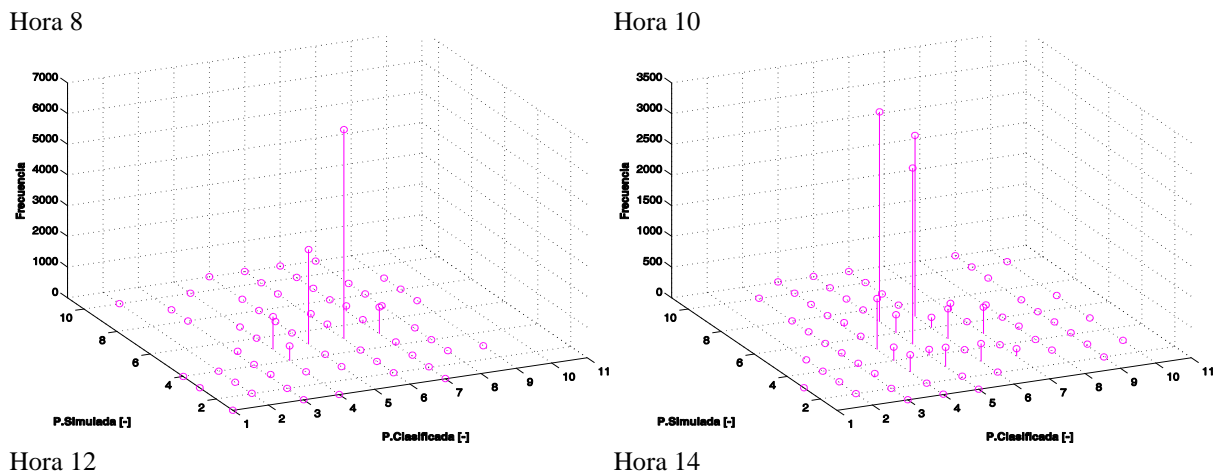


Figura 5-7. Variación de la precipitación total para la tormenta del 12 de abril. Resolución temporal 30 minutos.





Hora 8
Hora 10
Hora 12
Hora 14
Figura 5-8. Diagrama de dispersión para comparar el valor del celda clasificado y el valor pronosticado para las primeras 14 horas de la tormenta de abril 12 de 1995.

5.2.4 Análisis de Sensibilidad

El análisis de sensibilidad proporciona información detallada acerca de las limitaciones del AC y contribuye a la mejora futuras modelaciones con dicha teoría (Kocabas y Dragicevic, 2006). Inicialmente se propuso verificar si el uso de varias plantillas mejoraba el desempeño del modelo de pronóstico, así que se decidió usar una vecindad tipo Von Neuman, pero a la capacidad de memoria del aplicativo computacional no se logró ejecutar ese tipo de modelo. Entonces el desarrollo del análisis de sensibilidad se basó en los siguientes aspectos:

- Cambios en la clasificación
- Resolución de conflictos en patrones repetidos
- Asignación de condiciones iniciales
- Niveles de agregación temporal de las series en el caso de los eventos
- Reducción del número de reglas

A pesar que los tres primeros corresponden a la implementación de nuevos modelos, se consideró importante discutir acerca de las variaciones de los elementos del autómatas y las implicaciones en los resultados.

5.2.4.1 Cambios en la Clasificación

De acuerdo con los siete tipos de clasificación expuestos en la Tabla 5-1 y los mapas de muestra de la Figura 5-5, clasificaciones como la tipo 1, 2 realizan una clara identificación de los valores bajos de precipitación, mientras la clase 3 y 5 agrupa en la primera categoría las precipitaciones bajas, la clasificación tipo 3, 4 agrupa de manera “gruesa” las precipitaciones bajas. Una clasificación gruesa en la que se agrupen los valores bajos del campo de precipitación genera un mapa clasificado donde es frecuente que todos los

pixeles correspondan a un mismo valor de etiqueta que por lo general es 1 o 2, es decir condiciones de precipitación nula o tiempo seco. Ello implica que en ocasiones no se presenten condiciones propicias para la presencia de una condición húmeda y por ende errores en el pronóstico.

La clasificación de los valores de precipitación tiene una directa relación con la etapa de construcción de las reglas de evolución del autómatas. La técnica de minería de datos seleccionada se basa en el autoaprendizaje de los datos. Se infiere que si en la tabla de patrones se encuentra una cantidad considerable de datos con etiqueta que hace referencia a precipitaciones bajas, los resultados de la decisión del árbol serán precipitaciones bajas. Esta situación se presenta por ejemplo en el modelo "Clase3" donde, más de la mitad de las hojas o salidas del árbol de decisión son precipitación con etiqueta "Seco" (aproximadamente de 0-5mm). Caso contrario del modelo "Clase1", en el que la decisión más frecuente en la decisión es "Tormenta" (Ver Figura 5-9).

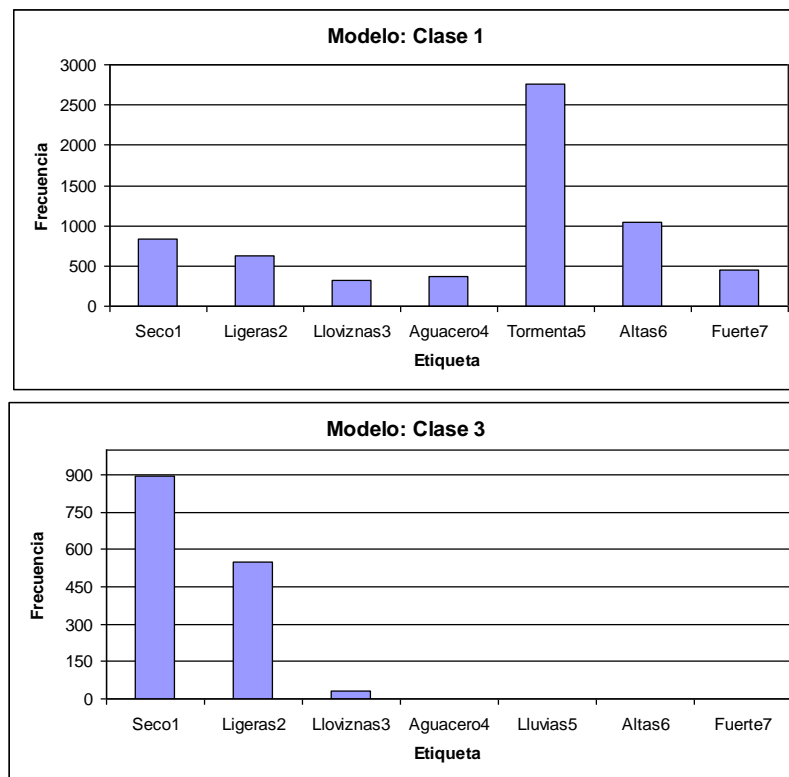


Figura 5-9. Frecuencia de posibles salidas en la decisión del árbol de decisión para el modelo de pronóstico de precipitación diaria
Nota: Superior modelo "Clase 1", inferior modelo "Clase 3"

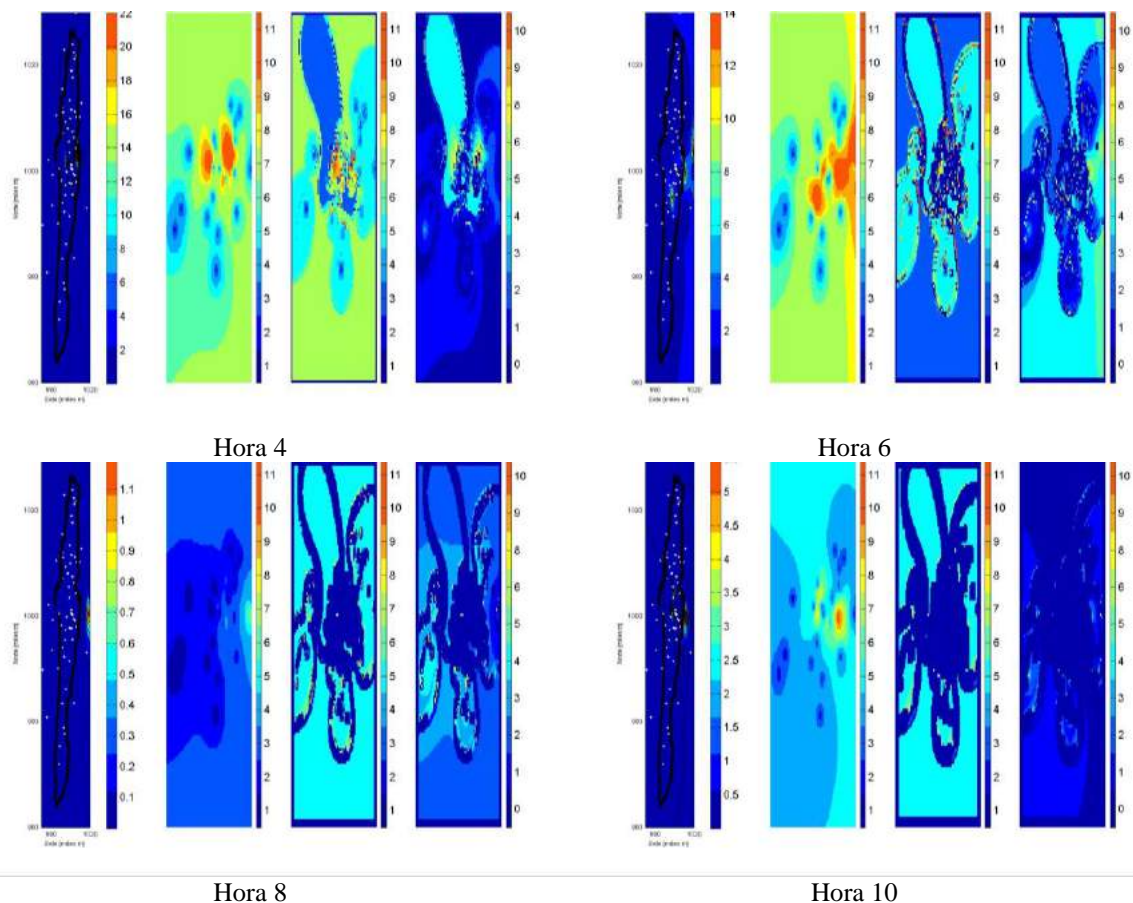
Con el fin de evaluar si una clasificación que discretizara de manera muy fina desde los valores bajos hasta los valores altos de precipitación mejoraba los resultados, se intentó probar proponer una nueva clasificación, pero los tiempos computacionales y la capacidad del software no lo permitieron.

5.2.4.2 Condiciones Iniciales

Los resultados de la modelación de la tormenta del abril 12 de 1995 mostrada en los resultados en la

Figura 5-6 fueron realizados bajo el supuesto de que la condición inicial para el pronóstico cada dos horas se actualizaba con el mapa clasificado correspondiente para el intervalo de tiempo en interés (Ver en la metodología Figura 4-10). Si por el contrario las condiciones iniciales para hacer el pronóstico en el tiempo t fueran el mapa pronosticado del tiempo inmediatamente anterior $t-1$ (Ver en la metodología Figura 4-11), existe una gran variación en los resultados.

Por un lado las formas presentes en el mapa de las condiciones iniciales en el primer instante se mantiene a través del tiempo, advirtiendo que no en la misma magnitud. Usando esta nueva propuesta de condición inicial, es frecuente que dentro de las reglas del árbol de decisión no existan varios patrones de los que se están creando con la nueva condición, por lo que en el momento de tomar la decisión el algoritmo opta por la decisión “Vacío” (en MATLAB® es NaN) en los mapas aparece con color azul oscuro, por lo que suele confundirse con el valor de etiqueta 1. Ocurre lo mismo en el mapa de errores absolutos, donde es frecuente encontrar colores azules oscuros que no corresponden a cero sino a un “vacío” debido que no puede restar un valor numérico del mapa clasificado con un “vacío” del mapa de pronóstico (Ver Figura 5-10).



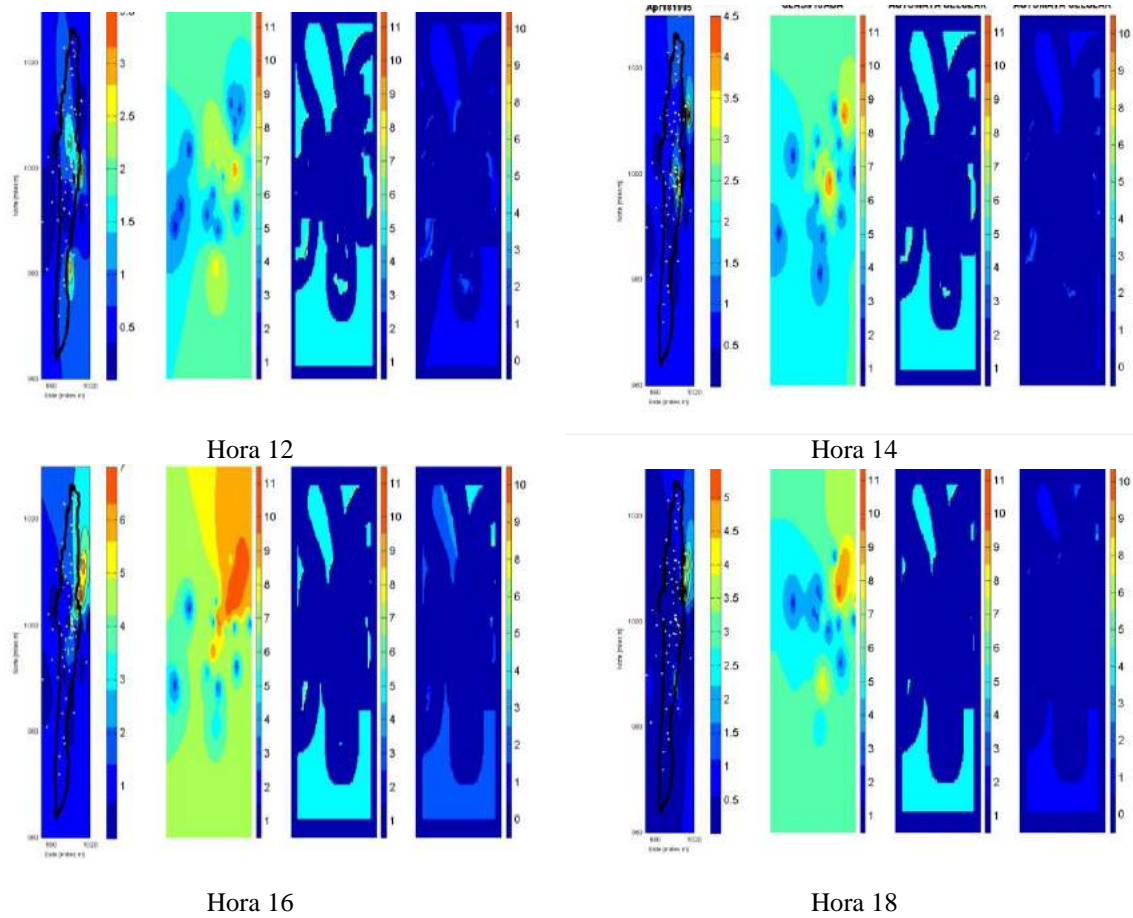
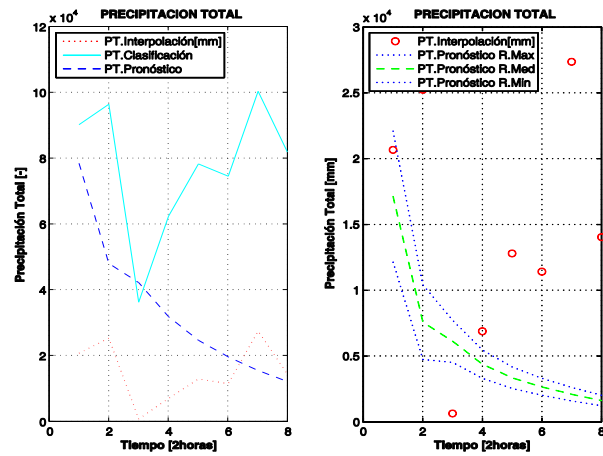
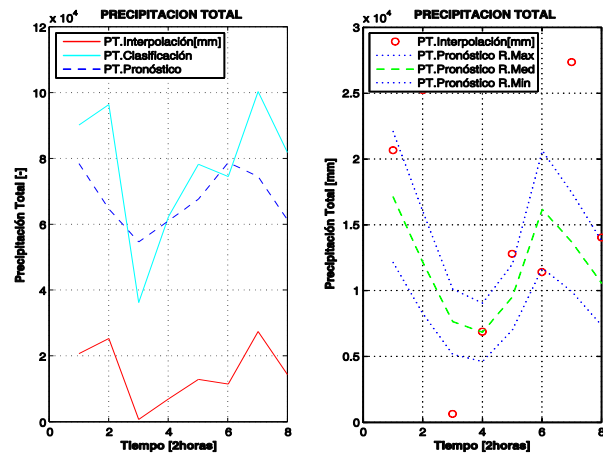


Figura 5-10. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 cambiando la condición inicial para cada instante de tiempo.

En la Figura 5-11 se comparan los valores de precipitación total para las dos posibilidades de condición inicial. Es evidente que la condición inicial para cuando se actualiza en cada instante no representa de manera adecuada la precipitación total. En la Figura 5-12 está la comparación de los gráficos de dispersión. En la condición que no actualiza existen mayores frecuencias en las celdas correctamente pronosticadas.

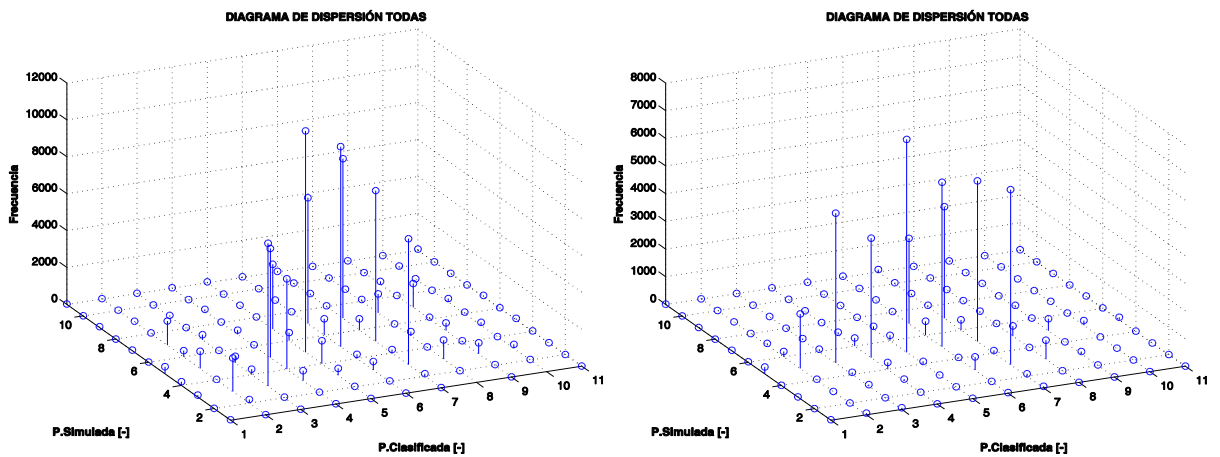


Sin actualizar



Actualizando

Figura 5-11. Comparación de los diagrama de precipitación total variando las condiciones iniciales para cada instante de tiempo



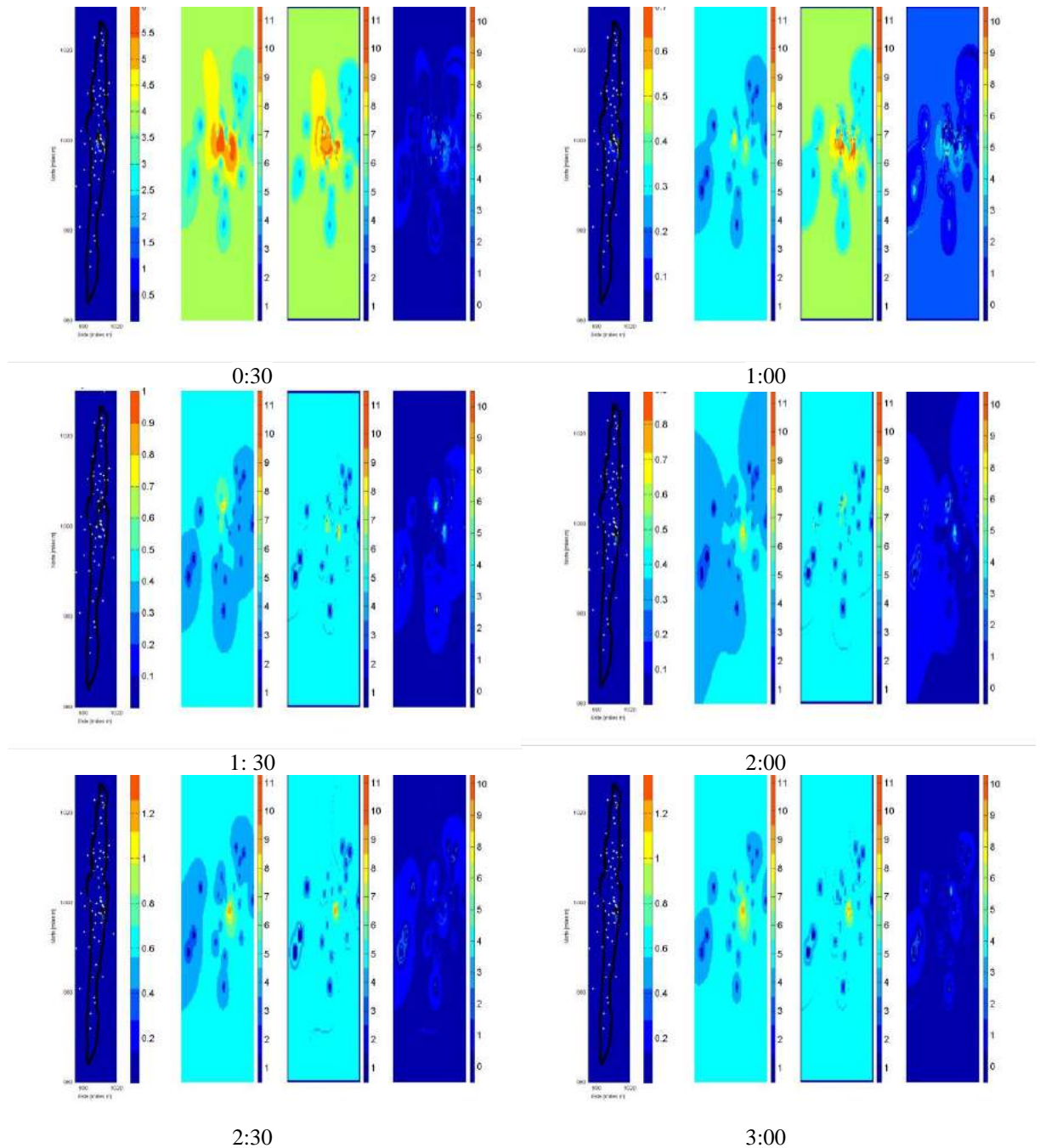
Sin actualizar

Actualizando

Figura 5-12. Comparación de los diagrama de dispersión variando las condiciones iniciales para cada instante de tiempo

5.2.4.3 Niveles de Agregación Temporal

Como ya se había sugerido, la selección de la resolución temporal del pronóstico está en función del requerimiento para lo que se desea. Con el fin de comparar los resultados obtenidos al implementar el modelo con diferentes escalas temporales, se implementó la metodología del AC a la tormenta del abril 12 a resolución de 30 minutos. En la Figura 5-13 se muestra la secuencia de las primeras 6 horas de la tormenta.



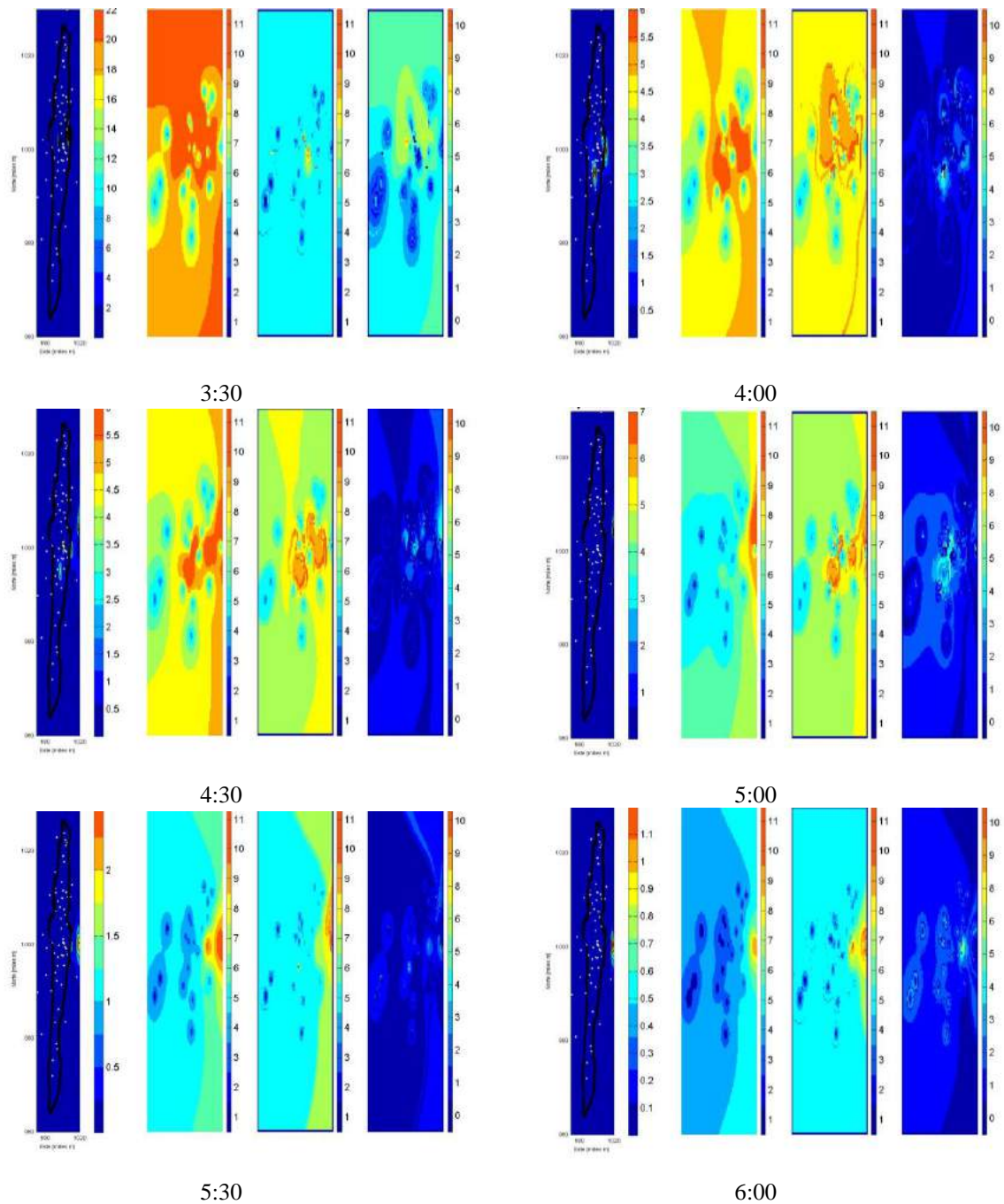


Figura 5-13. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 con resolución temporal de 30 minutos.

Nota: Solo se muestran las primeras 6 horas del evento

Los mapas de errores absolutos o “aciertos” muestran un buen desempeño del modelo cuando la resolución temporal es 30 minutos. Existe mayor frecuencia de errores absolutos bajos que en el caso de la resolución temporal de 2 horas (

Figura 5-6).

En la Figura 5-14 se comparan los diagramas de dispersión para la misma tormenta pero a diferentes agregaciones temporales para la tormenta de abril 12 de 1995. Es claro que con la resolución temporal mas fina es mas frecuente que las celdas con valores de clasificación intermedios sea mejor pronosticados.

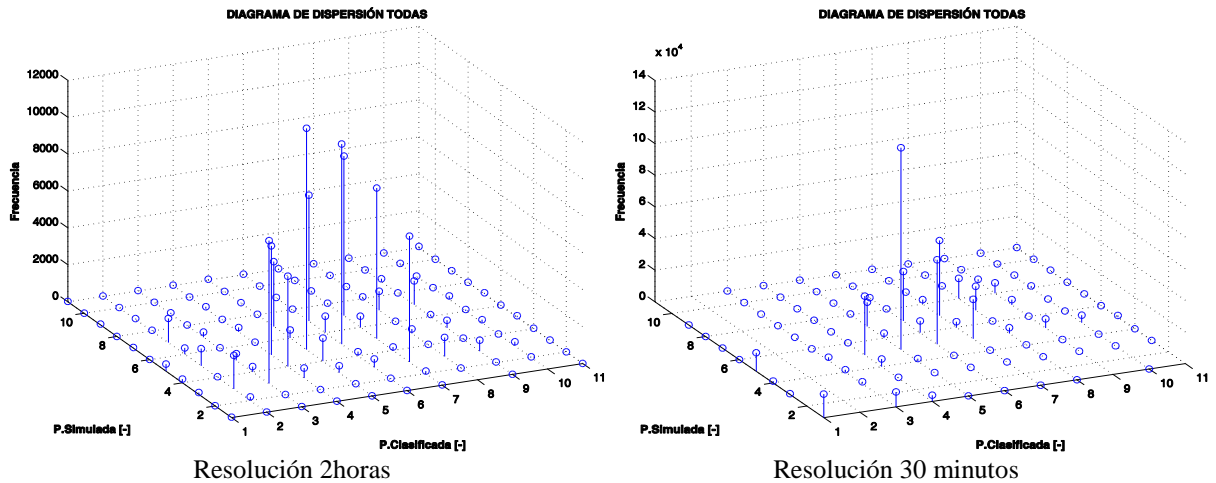


Figura 5-14. Diagrama de dispersión para comparar el valor del celda clasificado y el valor pronosticado

5.2.4.4 Reducción del Número de Reglas

Ya se vio que los modelos tienen una gran cantidad de reglas para reproducir el fenómeno de la precipitación, pero surge la inquietud de qué tanto afecta a los resultados la utilización de solo un porcentaje de las reglas. En el caso de la tormenta de abril 12 se presentan a continuación dos situaciones: 1) utilización del 20% de las reglas más frecuentes del modelo original, y 2) uso de las 3 reglas más frecuentes por salida del modelo original. Entiéndase por salida la decisión del árbol. La Tabla 5-7 muestra las etiquetas asignadas para cada rango.

Tabla 5-7. Asignación de etiquetas de acuerdo a los rangos de precipitación por agregación temporal para el evento de abril 12 de 1995.

Resolución temporal	Definición del rango [mm] y etiquetas
2 horas	RANGO=[0 0.008 0.028 0.146 0.5 0.999 1.432 2.316 2.851 4.353 5.963 24.3] CLASIFICADOR=[11 12 13 14 15 16 17 18 19 20 21]

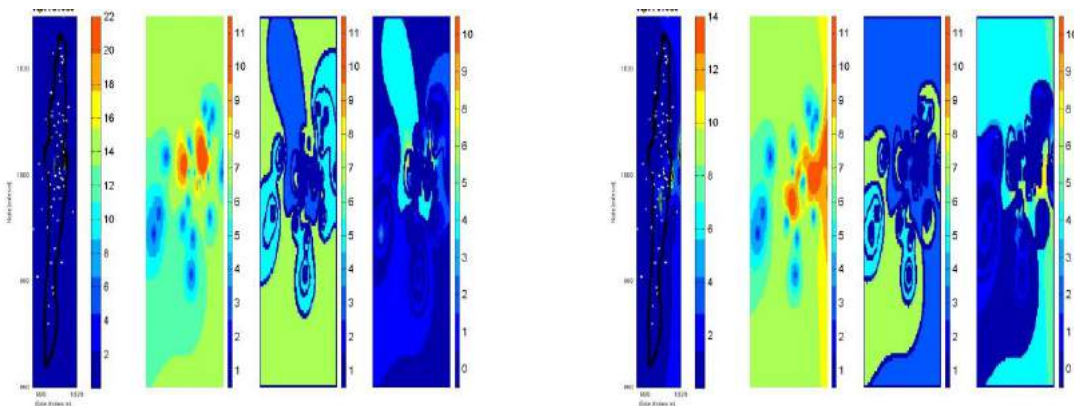
En el caso de la situación uno, la selección del nuevo conjunto de reglas se llevó a cabo al organizar de manera descendente las reglas más utilizadas en el modelo original y elegir solo el 20% de estas. El árbol de decisión para la tormenta de abril 12, arrojó un conjunto de 1704 reglas, de las cuales solo 13 representan el 20% más frecuente. En la Tabla 5-8 se muestran las 25 reglas más usadas en el modelo original de pronóstico. Se identifican las 9 posiciones con la letra P (Plantilla Moore) con su respectivo valor de acuerdo con las etiquetas asignadas con la letra V. La columna “OUT” indica las “salida” del patrón representado en las primeras 18 columnas.

Gráficamente los resultados de los mapas de pronóstico se encuentran en la Figura 5-15. Debido a la reducción del número de reglas es posible encontrar patrones que no están dentro del nuevo conjunto de reglas, por lo que la respuesta del modelo es “vacío” o “NaN”. Este es un ejemplo recurrente en los bordes de las zonas que contienen valores de precipitación similar.

Tabla 5-8. Frecuencias de las reglas más utilizadas en el modelo de AC para la tormenta de abril 12 de 1995 con resolución temporal 2h

Rank	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8	V8	P9	V9	OUT	FRECUENCIA
1	5	17	9	17	7	17	1	17	3	17	2	17	6	17	4	17	8	17	13	23543
2	5	16	7	16	9	16	3	16	1	16	2	16	4	16	8	16	0	0	17	21856
3	5	15	3	15	9	15	1	15	7	15	6	15	2	15	4	15	8	15	15	15195
4	5	14	3	14	9	14	7	14	2	14	4	14	1	14	8	14	0	0	15	9249
5	5	13	9	13	1	13	7	13	3	13	2	13	6	13	4	13	8	13	15	7409
6	5	12	3	12	4	12	6	12	8	12	2	12	0	0	0	0	0	0	14	3974
7	5	19	9	19	7	19	1	19	6	19	3	19	2	19	4	19	0	0	17	1753
8	5	18	7	18	1	18	6	18	9	18	2	18	4	18	0	0	0	0	13	1503
9	5	11	3	11	1	11	0	0	0	0	0	0	0	0	0	0	0	0	13	597
10	5	15	3	15	9	15	1	14	8	15	0	0	0	0	0	0	0	0	15	547
11	5	16	7	16	9	16	3	16	1	15	0	0	0	0	0	0	0	0	16	435
12	5	15	3	15	9	14	1	15	7	15	0	0	0	0	0	0	0	0	15	354
13	5	15	3	15	9	15	1	15	7	14	0	0	0	0	0	0	0	0	15	339
14	5	14	3	14	9	13	1	14	0	0	0	0	0	0	0	0	0	0	14	339
15	5	15	3	15	9	16	7	15	1	15	2	15	0	0	0	0	0	0	16	327
16	5	14	3	13	7	14	9	14	0	0	0	0	0	0	0	0	0	0	14	308
17	5	20	3	20	7	20	1	20	9	20	2	20	4	20	6	20	8	20	19	307
18	5	17	9	17	7	16	1	16	3	17	0	0	0	0	0	0	0	0	17	302
19	5	14	3	14	9	15	1	14	4	14	2	14	7	14	0	0	0	0	15	277
20	5	14	3	15	9	14	7	14	1	14	4	14	8	14	0	0	0	0	15	271
21	5	14	3	14	9	14	7	15	1	15	6	14	0	0	0	0	0	0	15	247
22	5	15	3	14	7	15	9	15	2	14	4	15	0	0	0	0	0	0	15	240
23	5	16	7	16	9	16	3	15	4	16	2	15	0	0	0	0	0	0	15	238
24	5	14	3	14	9	14	7	13	8	14	0	0	0	0	0	0	0	0	14	226
25	5	17	9	17	7	16	1	17	3	17	4	17	0	0	0	0	0	0	17	224

Nota: solo se muestran las 25 primeras reglas más frecuentes de las 1704 reglas del árbol de decisión.



Hora 4

Hora 6

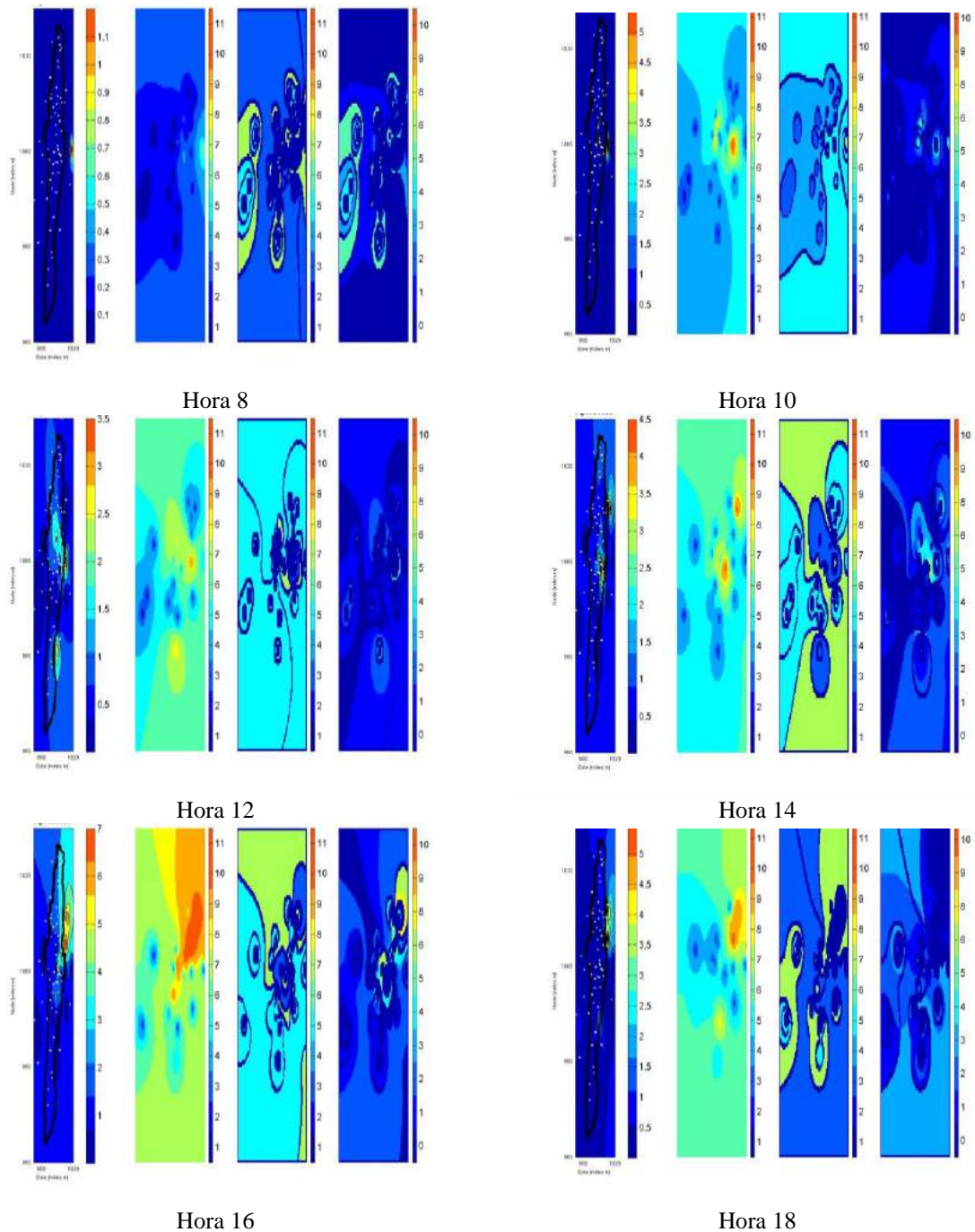


Figura 5-15. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 con resolución temporal de 2h. Simulación solo con las 13 reglas mas frecuentes del modelo inicial.

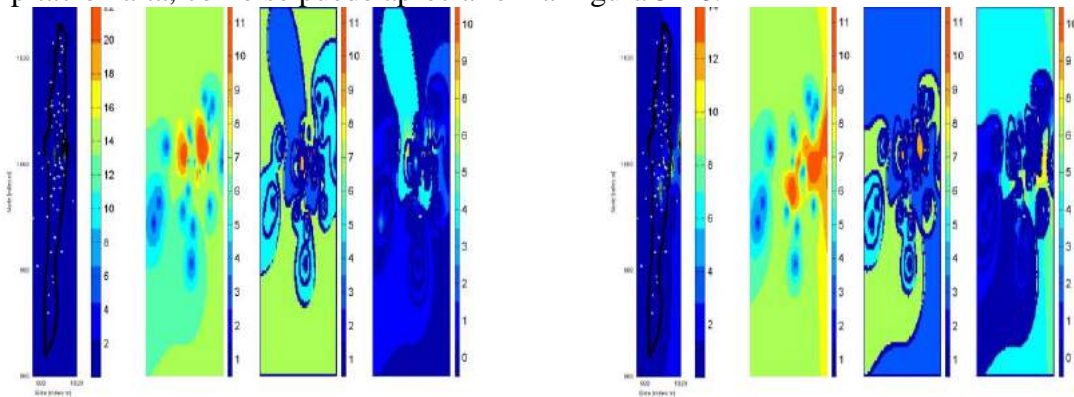
Adicional a la situación donde se presentan las celdas con valor de NaN se observa que dentro de las reglas más frecuentes no se encuentran reglas con salida con categoría alta (18, 19, 20 y 21) lo que conlleva que en los mapas pronóstico no existan valores de precipitación alta. Es decir, que no reproduce las condiciones de precipitaciones máximas. Con el fin de minimizar esa situación se planteó la alternativa de seleccionar las reglas de

manera que todas las posibles “salidas” se encontraran en el nuevo conjunto de reglas. Para ello se eligieron las tres reglas más frecuentes por salida. El nuevo conjunto de 33 reglas se muestra en la Tabla 5-9.

Tabla 5-9. Organización de reglas de acuerdo a la frecuencia por salida de las reglas más utilizadas en el modelo de AC para la tormenta de abril 12 de 1995 con resolución temporal 2h

id	P1	V1	P2	V2	P3	V3	P4	V4	P5	V5	P6	V6	P7	V7	P8	V8	P9	V9	OUT
1	5	15	3	14	7	16	1	15	9	16	4	15	0	0	0	0	0	0	11
2	5	15	3	14	7	16	1	14	9	16	0	0	0	0	0	0	0	0	11
3	5	14	3	13	7	15	2	13	9	14	6	14	0	0	0	0	0	0	11
4	5	16	7	16	9	17	1	16	3	17	4	16	2	16	8	16	0	0	12
5	5	17	9	17	7	17	1	17	3	16	2	17	6	16	0	0	0	0	12
6	5	17	9	16	7	17	1	17	3	17	6	16	8	17	0	0	0	0	12
7	5	17	9	17	7	17	1	17	3	17	2	17	6	17	4	17	8	17	13
8	5	18	7	18	1	18	6	18	9	18	2	18	4	18	0	0	0	0	13
9	5	11	3	11	1	11	0	0	0	0	0	0	0	0	0	0	0	0	13
10	5	12	3	12	4	12	6	12	8	12	2	12	0	0	0	0	0	0	14
11	5	14	3	14	9	13	1	14	0	0	0	0	0	0	0	0	0	0	14
12	5	14	3	13	7	14	9	14	0	0	0	0	0	0	0	0	0	0	14
13	5	15	3	15	9	15	1	15	7	15	6	15	2	15	4	15	8	15	15
14	5	14	3	14	9	14	7	14	2	14	4	14	1	14	8	14	0	0	15
15	5	13	9	13	1	13	7	13	3	13	2	13	6	13	4	13	8	13	15
16	5	16	7	16	9	16	3	16	1	15	0	0	0	0	0	0	0	0	16
17	5	15	3	15	9	16	7	15	1	15	2	15	0	0	0	0	0	0	16
18	5	16	7	16	9	15	1	16	6	16	3	16	0	0	0	0	0	0	16
19	5	16	7	16	9	16	3	16	1	16	2	16	4	16	8	16	0	0	17
20	5	19	9	19	7	19	1	19	6	19	3	19	2	19	4	19	0	0	17
21	5	17	9	17	7	16	1	16	3	17	0	0	0	0	0	0	0	0	17
22	5	19	9	19	7	19	1	20	8	19	6	19	4	20	0	0	0	0	18
23	5	19	9	19	7	20	3	19	4	20	0	0	0	0	0	0	0	0	18
24	5	19	9	19	7	19	1	20	8	19	6	19	4	19	2	19	0	0	18
25	5	20	3	20	7	20	1	20	9	20	2	20	4	20	6	20	8	20	19
26	5	21	3	21	1	21	7	21	8	21	6	21	0	0	0	0	0	0	19
27	5	20	3	20	7	21	2	20	0	0	0	0	0	0	0	0	0	0	19
28	5	18	7	18	1	18	6	17	2	18	9	17	3	17	0	0	0	0	20
29	5	17	9	17	7	18	2	18	0	0	0	0	0	0	0	0	0	0	20
30	5	15	3	14	7	16	1	16	9	15	6	14	8	16	0	0	0	0	20
31	5	15	3	16	8	14	4	15	1	16	6	15	7	14	0	0	0	0	21
32	5	16	7	17	9	16	3	14	0	0	0	0	0	0	0	0	0	0	21
33	5	15	3	16	8	15	1	14	7	15	6	16	0	0	0	0	0	0	21

Con este nuevo conjunto de reglas sí es posible encontrar celdas con valores de precipitación alta, como se puede apreciar en la Figura 5-16.



Hora 4

Hora 6

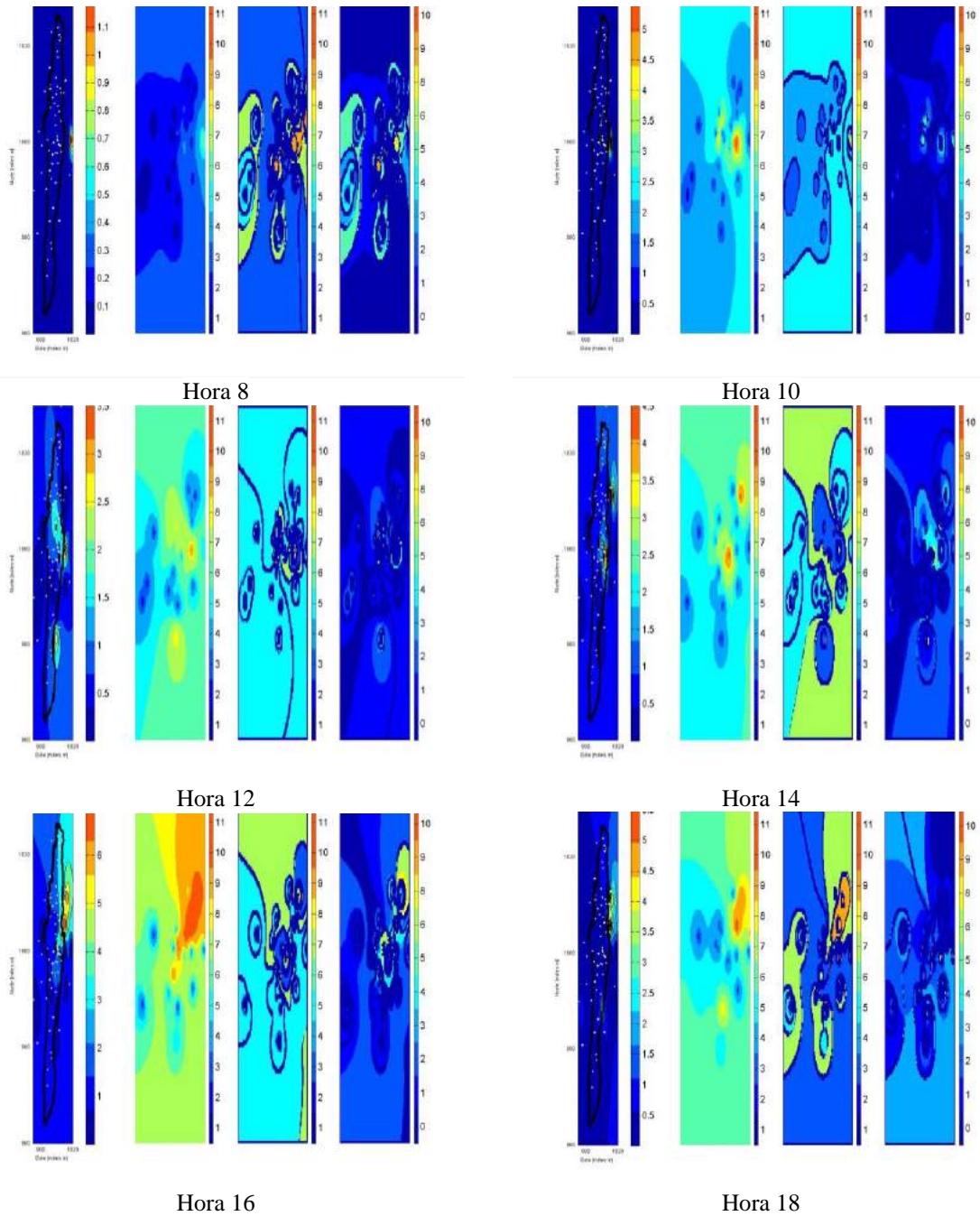


Figura 5-16. Ejemplo de pronóstico de precipitación horaria el evento ocurrido el 12-abril-1995 con resolución temporal de 2h. Simulación con 33 reglas.

Al comparar las gráficas de dispersión de las dos alternativas es claro que en el caso de las 13 reglas no existen celdas con valores pronosticados mayores a la etiqueta 8, mientras que en el caso de las 33 reglas si existen todas las posibilidades.

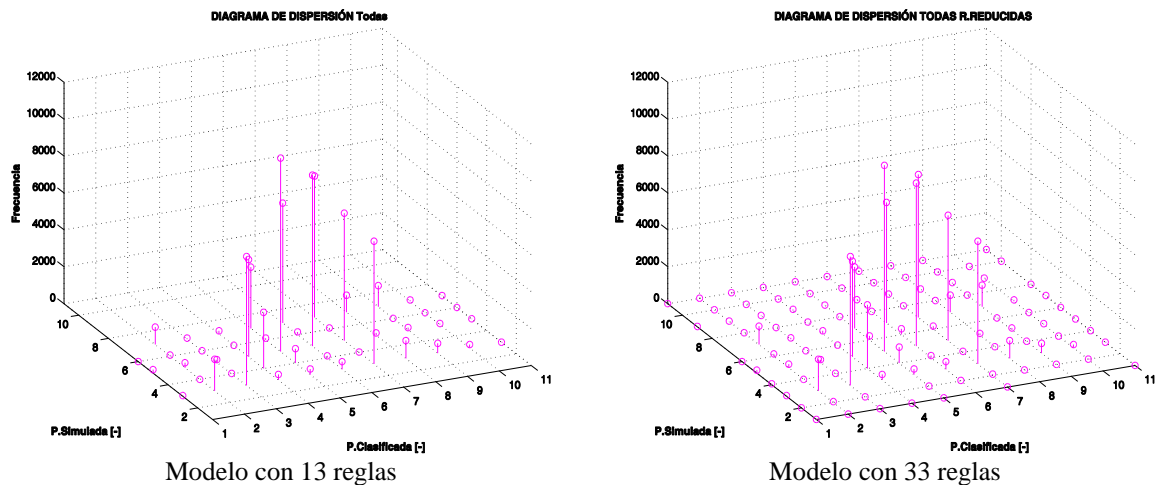


Figura 5-17. Comparación de los diagramas de dispersión para el evento ocurrido el 12-abril-1995 con resolución temporal de 2h, variando la cantidad de reglas de evolución.

Para darle solución a los valores de “Vacio” o “NaN” en las celdas donde no se encontró una regla de uso, se propone el uso de técnicas de procesamiento de imágenes. Actualmente existen una gran cantidad de algoritmos para la restauración y mejoramiento de imágenes. Con el fin de darle solución a las inconsistencias de los valores de las celdas de los mapas pronosticados solamente se hará mención de la operación de filtraje. El filtraje es una técnica de transformación de la imagen píxel a píxel que no depende de solamente el valor de un determinado píxel, sino también del valor de los valores de los pixeles vecinos en la imagen original. La imagen resultante después de la operación es una nueva imagen. Comúnmente los filtros sirven para suavizar o realzar detalles de la imagen, o inclusive minimizar efectos de ruido.

Una característica común a todos los tipos de datos ráster es la llamada "frecuencia espacial", que define la magnitud de cambios de los datos por unidad de distancia en una determinada zona de la imagen. Áreas de la imagen con pequeños cambios o con transiciones graduales en los valores de los datos se denominan áreas de bajas frecuencias (ej. superficie de una masa de agua en reposo). Áreas de grandes cambios o rápidas transiciones se conocen como áreas de altas frecuencias (ej. suelo urbano con densas redes de carreteras). El procedimiento matemático del filtraje consiste en relacionar cada píxel con sus vecinos, de acuerdo con unos coeficientes determinados por el usuario. En el caso de una matriz de 3*3, la formula para obtener el nuevo valor del píxel es (Chuvieco, 2002):

$$ND'_{i,j} = \frac{\sum_{p=-1,1} \sum_{q=-1,1} ND_{i+p,j+q} CF_{f+p,c+q}}{\sum_{p=-1,1} \sum_{q=-1,1} CF_{f+p,c+q}} \quad \text{Ecuación 5-1}$$

donde $ND_{i,j}$ indica el valor del Nivel Digital original del píxel i, j ; $ND'_{i,j}$ el Nivel Digital de salida para ese mismo píxel; CF el coeficiente de filtraje correspondiente, f y

c la fila y columna central de la matriz de filtraje. El valor de la fórmula se aproxima al entero más cercano. Para el uso de la Ecuación 5-1 en el caso del pronóstico de la precipitación, los Niveles Digitales de asocian al valor de precipitación clasificada.

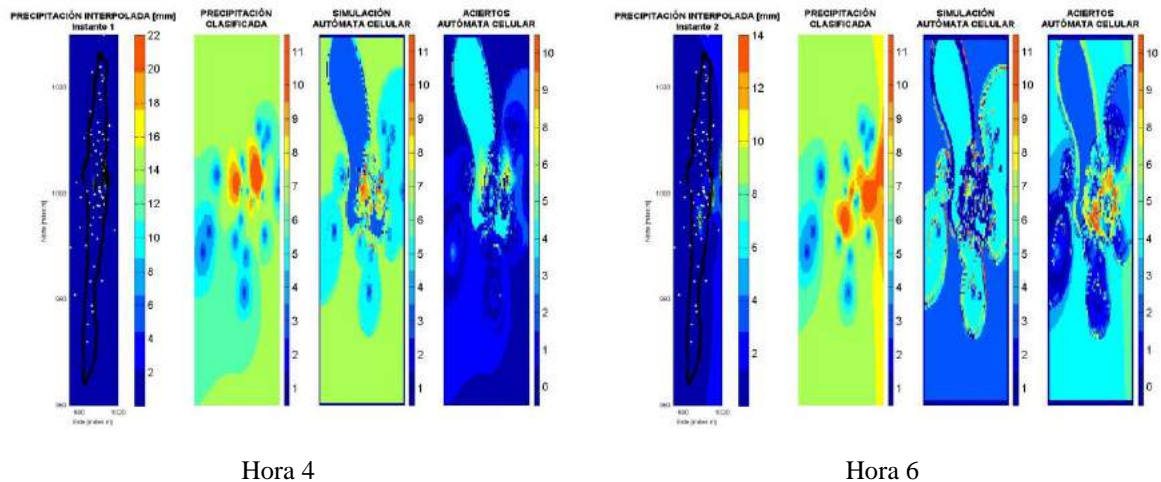
En la presente investigación se usó un filtro de paso bajo, que se caracteriza por enfatizar las bajas frecuencias, suavizando las imágenes y suprimiendo ruidos. Se busca entonces trata de asemejar el valor de cada celda con el valor de NaN al valor de la celda de los píxeles vecinos, reduciendo la variabilidad espacial de la imagen. Existen varios tipos de plantilla para los filtros de paso bajo, para fines del trabajo solo de uso la siguiente plantilla o kernel:

0.25	0.5	0.25
0.5	1	0.5
0.25	0.5	0.25

Figura 5-18. Matriz de coeficientes de filtraje. Kernel para el uso de un filtro de paso bajo.

Fuente: Chuvieco, 2002.

En las imágenes de la Figura 5-19 muestran el resultado de aplicar un filtro de paso bajo solo en las celdas con valores iguales a “Nan”.



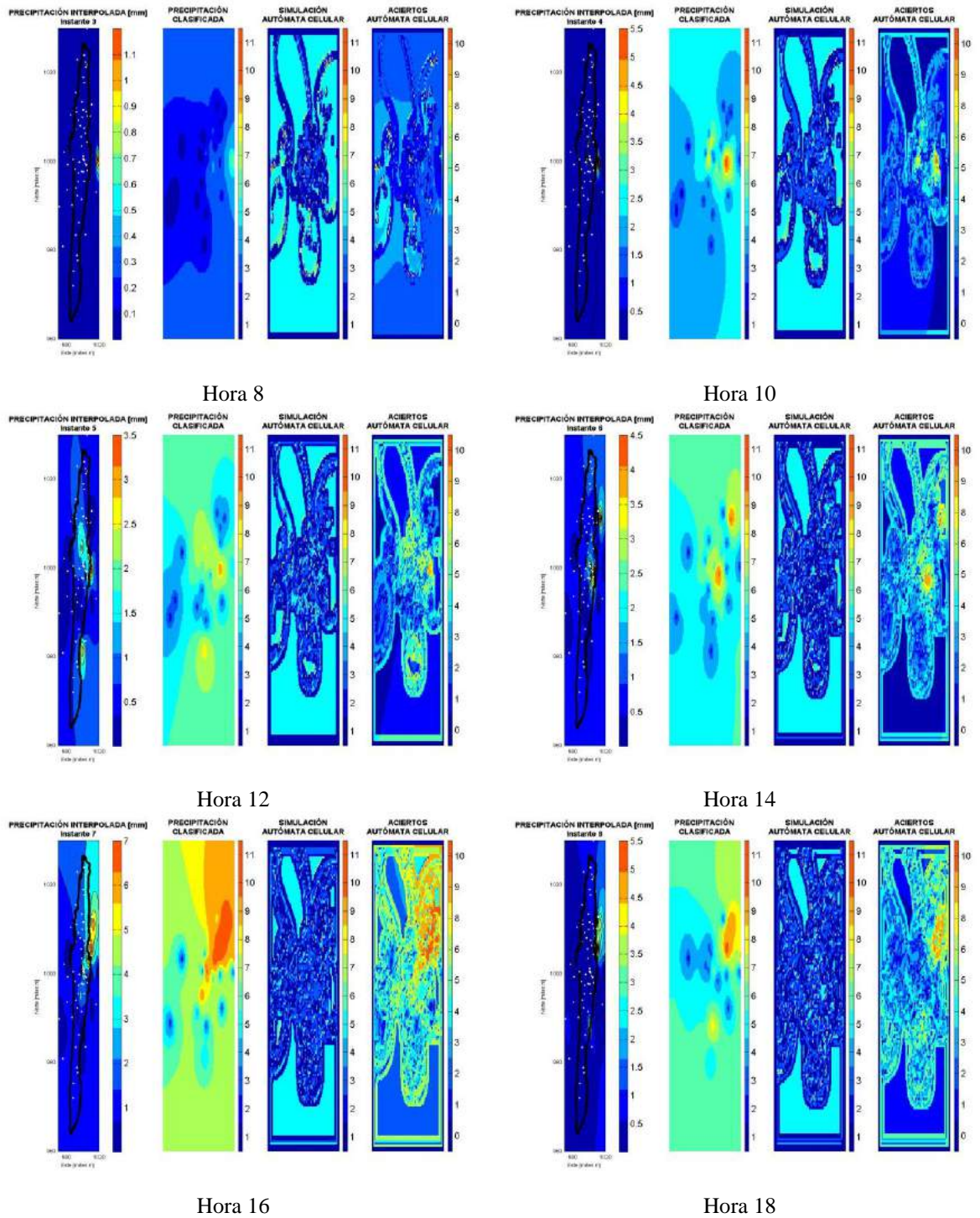


Figura 5-19. Aplicación de un filtro de paso bajo a las imágenes del modelo de AC para el evento ocurrido el 12-abril-1995 con resolución temporal de 2h

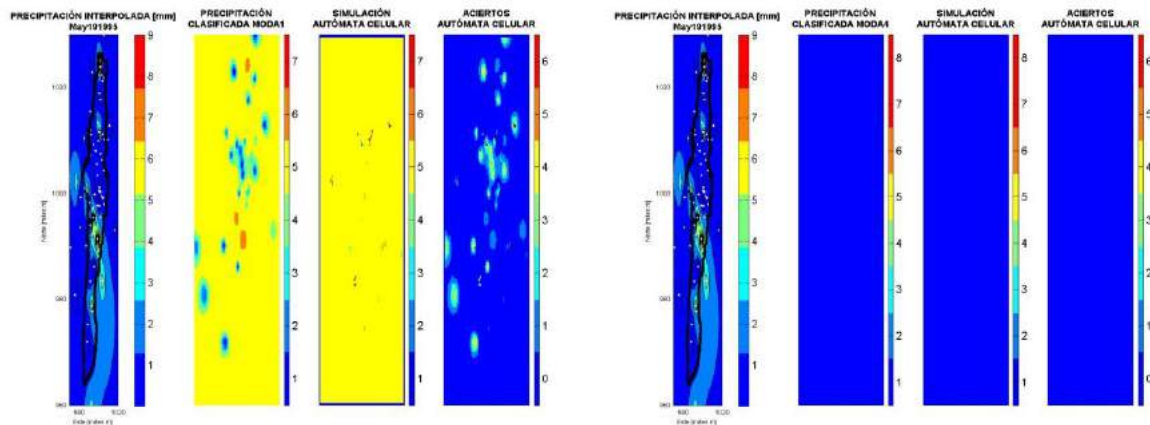
Nota: con condición de 33 reglas y actualizando la condición inicial cada 2 horas.

6 ESTADÍSTICA DE REGLAS

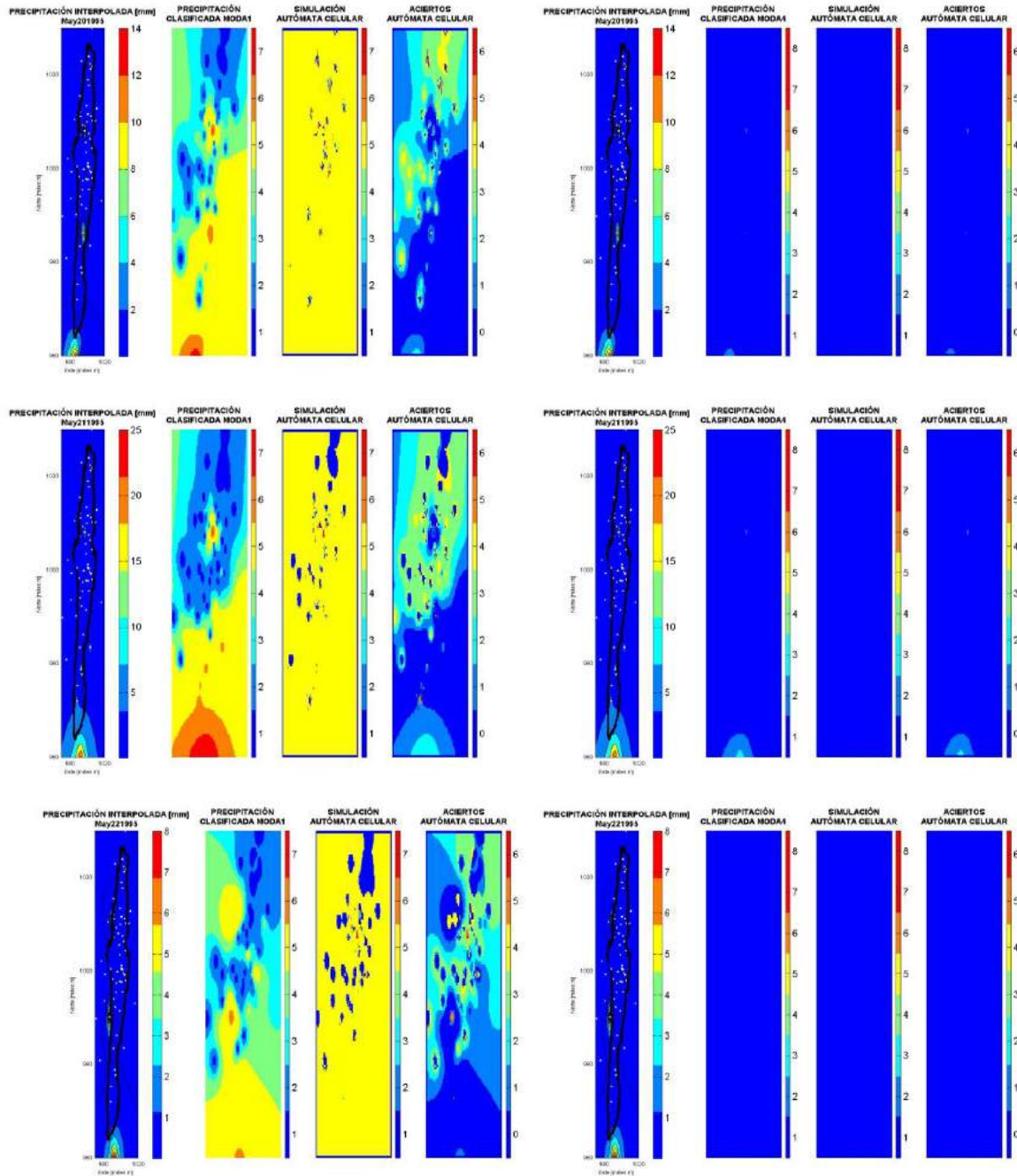
El desarrollo de este capítulo se basa en el análisis de las reglas obtenidas en cada uno de los modelos propuestos para el pronóstico de la precipitación diaria y a nivel evento. El objetivo de este capítulo es sugerir las variables predominantes que determinan el comportamiento de las reglas de evolución. En el Apéndice C, se pueden ver en detalle las reglas correspondientes a cada uno de los modelos propuestos.

6.1 REDUCCIÓN DE LA INFORMACIÓN

De acuerdo con lo visto en Tabla 5-5 el modelo que reduce más la cantidad de patrones al trasformarlo en reglas es el “Moda4” con una reducción del 73% respecto a los patrones iniciales. El modelo que reduce menos la información es el “Moda1” con un 23% y el intermedio es el “Moda2” con una reducción del 42%. En términos de eficiencia se pensaría que el modelo “Moda4” es el más apropiado, con el fin de verificar esa hipótesis se observaran a continuación una secuencia del pronóstico de precipitación de mayo 19 a mayo 23 de 1995, con el fin de evaluar visualmente el desempeño de estos modelos.



Implementación de un modelo de autómatas celulares para el pronóstico de la precipitación espacial. Caso de estudio ciudad de Bogotá (Colombia).



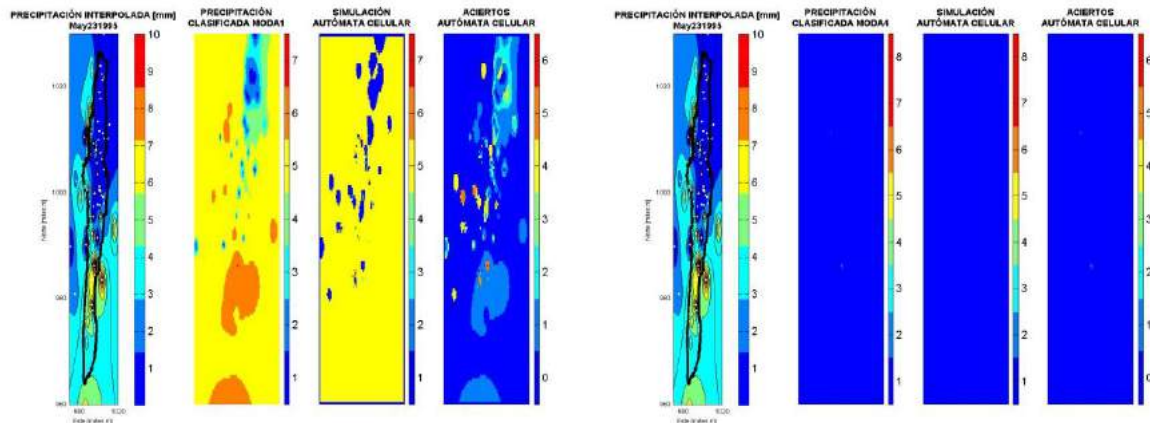


Figura 6-1. Secuencia de la serie de pronóstico para el periodo comprendido desde mayo 19 a mayo 23 de 1995 para los modelos "Moda1" y "Moda4".

Nota: Moda1 derecha, Moda4 izquierda.

Mediante la inspección de las anteriores gráficas, se evidencia que una alta reducción de la información de patrones a reglas no es un indicativo de ser la mejor opción si se comparan varios modelos.

La mayor reducción de la información ocurrió en un modelo donde la clasificación es muy gruesa “Moda4”. Ello implica que desde el mapa clasificado se visualicen pocos valores de lluvia lo que se ve representado en mapas de pronóstico con escasos rastros de la ocurrencia del fenómeno. El conjunto de reglas del modelo “Moda1” muestra dinámica espacial de la precipitación. Es frecuente encontrar que el pronóstico de un día determinado se parezca al mapa clasificado del día anterior. Esto es un indicio de la fuerte inercia que trae el modelo, y al mismo tiempo sugiere que se deberían probar con otros rezagos en el momento de la construcción. En de la tabla de patrones.

6.2 USO DE REGLAS EN EL PRONÓSTICO DE LA PRECIPITACIÓN

En el caso de la construcción de la tabla de patrones para los datos de precipitación total diaria, se tomaron en cuenta todos los patrones de los cinco años en una sola tabla. Es frecuente que en el momento de realizar el pronóstico para un año específico no se utilicen todas las reglas del árbol, pues no todos los patrones son comunes en los 5 años.

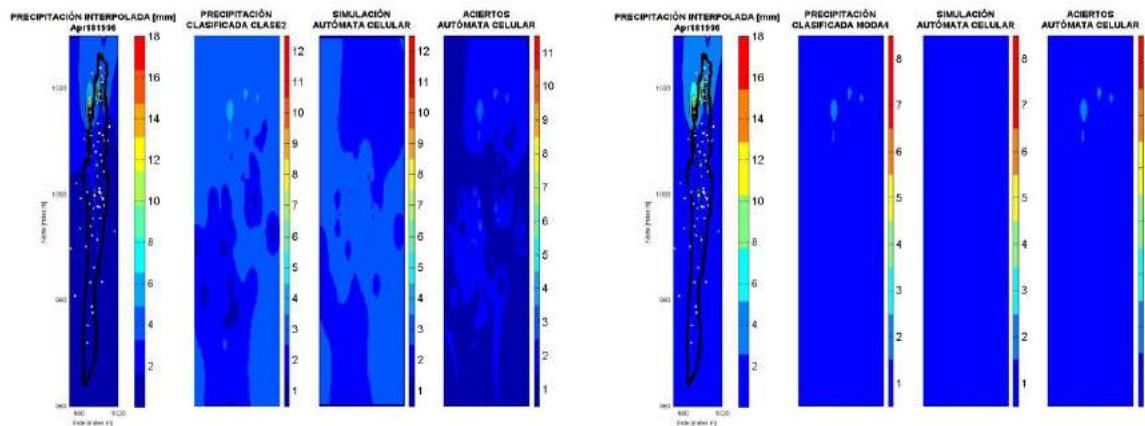
Al comparar el total de reglas por modelo (Ver Tabla 5-5) establecidas mediante el uso del algoritmo ID3, con el total de reglas realmente utilizadas en la implementación del autómatas celulares por año, se observa que en algunos casos solo se utilizaron el 11%, caso específico del modelo “Clase2” mientras que otros modelos usan más del 70% de las reglas originales, como es el caso del modelo Calse4, Moda1 y Moda4 (Ver

Tabla 6-1).

Tabla 6-1. Numero de reglas utilizadas y no utilizadas en cada modelo de pronóstico de precipitación.

Indicador	Modelo	Año					Indicador	Modelo	Año				
		1995	1996	1997	1998	1999			1995	1996	1997	1998	1999
Número de reglas usadas	clase1	3748	3465	3105	3217	3265	Número de reglas sin usar	clase1	2666	2949	3309	3197	3149
	clase2	1187	908	717	837	764		clase2	5313	5592	5783	5663	5736
	clase3	306	236	163	223	199		clase3	1171	1241	1314	1254	1278
	clase4	330	223	165	227	208		clase4	133	240	298	236	255
	clase5	2937	2053	1284	2140	1757		clase5	1967	2851	3620	2764	3147
	moda1	4026	3531	3234	3289	3856		moda1	1192	1687	1984	1929	1362
	moda2	861	791	589	660	710		moda2	234	304	506	435	385
	moda3	347	335	230	287	326		moda3	98	110	215	158	119
	moda4	216	205	139	172	192		moda4	53	64	130	97	77
	moda5	2203	1962	1195	1771	1847		moda5	858	1099	1866	1290	1214
Porcentaje de reglas usadas	clase1	58	54	48	50	51	Porcentaje de reglas sin usar	clase1	42	46	52	50	49
	clase2	18	14	11	13	12		clase2	82	86	89	87	88
	clase3	21	16	11	15	13		clase3	79	84	89	85	87
	clase4	71	48	36	49	45		clase4	29	52	64	51	55
	clase5	60	42	26	44	36		clase5	40	58	74	56	64
	moda1	77	68	62	63	74		moda1	23	32	38	37	26
	moda2	79	72	54	60	65		moda2	21	28	46	40	35
	moda3	78	75	52	64	73		moda3	22	25	48	36	27
	moda4	80	76	52	64	71		moda4	20	24	48	36	29
	moda5	72	64	39	58	60		moda5	28	36	61	42	40

En la Figura 6-2 se muestra la secuencia de pronóstico de abril 18 a abril 20 de 1996 con el fin de comparar el número de reglas usadas por el modelo. En la columna izquierda están los pronósticos del modelo “Clase2”, el que para el año 1996, uso el 14% de las reglas suministradas por el árbol de decisión. El modelo “Moda4” uso para el mismo año el 76% de las reglas. No es fácil comparar los resultados de los dos modelos, pues el pronóstico está altamente influenciado por las clasificaciones. Es mayor el número de reglas usadas para el modelo “Clase2”, lo que era de esperar pues al realizar una discretización más fina de los valores de precipitación, la cantidad y variedad de patrones es superior.



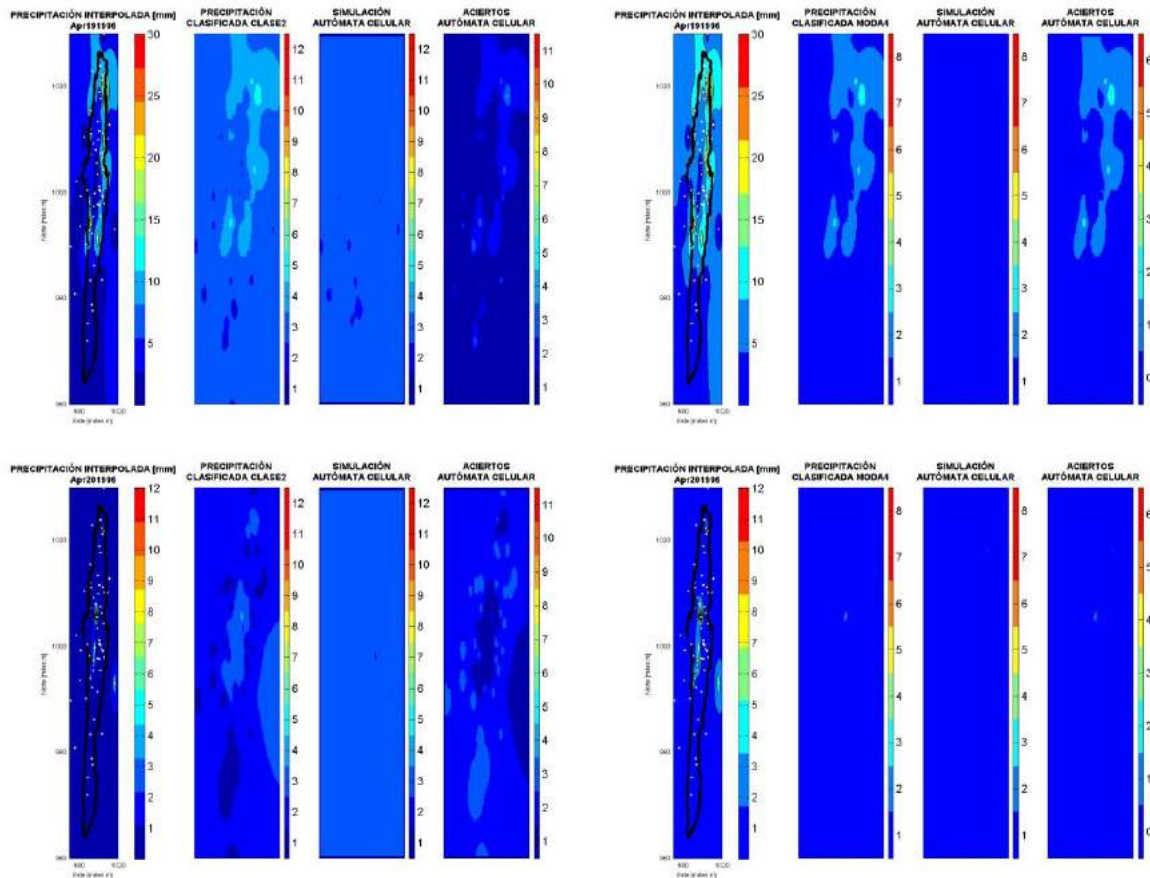


Figura 6-2. Secuencia de la serie de pronóstico para el periodo comprendido desde abril 18 a abril 20 de 1996 para los modelos "Clase2" y "Moda4".
 Nota: Clase2 derecha, Moda4 izquierda.

6.3 ANÁLISIS DE FRECUENCIAS POR POSICIÓN Y ETIQUETA

De acuerdo con los análisis anteriores el modelo que puede representar mejor el pronóstico de la precipitación diaria es el “Moda1”, uno de los modelos con bajo rendimiento es el “Moda4” y se eligió “Moda5” para un modelo con rendimiento medio. A continuación se analizaran el análisis de frecuencias de estos tres modelos, (en el Apéndice B pueden verse en detalle los resúmenes de todas las clasificaciones).

Para efectos del análisis de este capítulo se llamarán posiciones a cada uno de los niveles del árbol de decisión en el momento de hacer la pregunta del valor de una vecindad. Para una plantilla de Moore, como existen 9 tipos de vecindades pueden existir máximo 9 niveles de árbol. Los nodos pueden existir en n cantidad pero sus etiquetas corresponden a los nombres de cada una de las vecindades de la plantilla tipo Moore. Los conectores de los nodos son los posibles valores de las etiquetas y la salida es la decisión después de recorrer

toda una la rama del árbol que cumpla los patrones encontrados en el mapa de condiciones iniciales. Ver el detalle en la Figura 6-3.

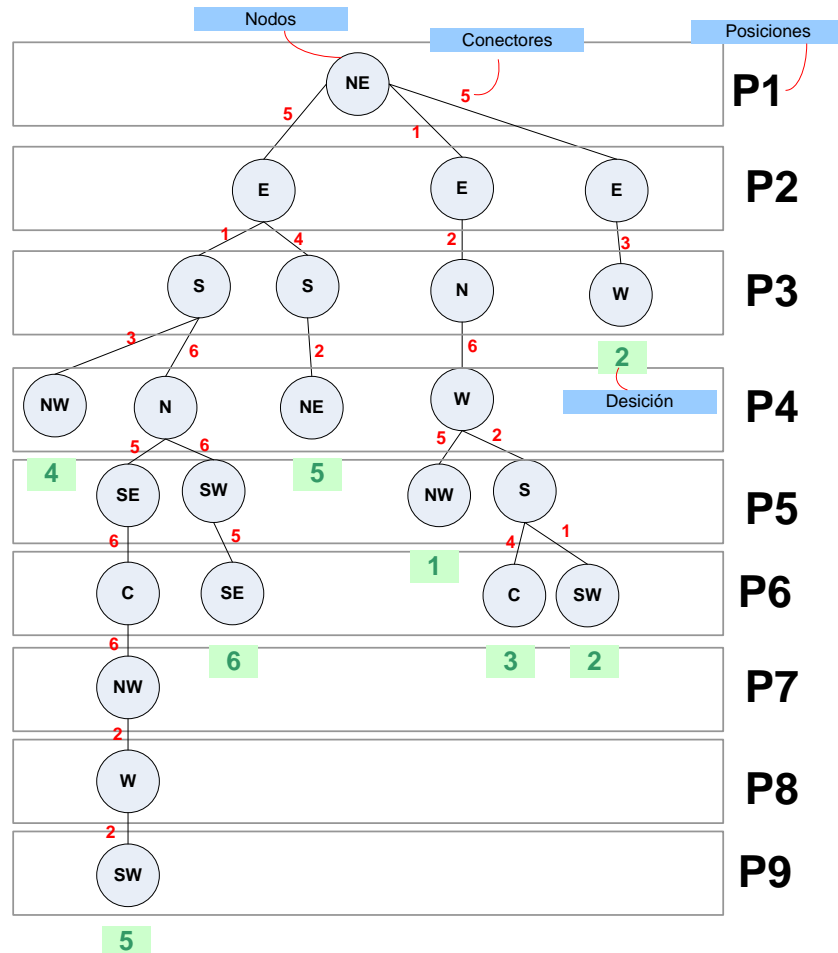


Figura 6-3. Identificación de los niveles, nodos y conectores del árbol de decisión
Fuente: La autora

En la Tabla 6-2 se observan los vecinos a los que con mayor frecuencia se les hace una pregunta para tomar una decisión en cada uno de los niveles del árbol. En el modelo Moda es poco frecuente preguntarle al vecino SW, C y W en el Moda5 y en la Moda4 el árbol le pregunta en igual frecuencia a todos los vecinos. De este análisis se podría inferir que las celdas N, NE, NW y E tienen más peso que las C, W, SE, S y SW en el momento de tomar una decisión.

Tabla 6-2. Vecino más frecuente al que se le hace para tomar una decisión usando una plantilla MOORE.

Moda1					
Posición	Frecuencia Vecino ¹				
	1995	1996	1997	1998	1999
P1	NW	NW	NW	NW	NW
P2	N	N	N	N	N
P3	NE	NE	NE	NE	NE
P4	W	W	W	W	W
P5	SW	SW	SW	SW	C
P6	SW	SW	SW	SW	SW
P7	C	C	C	C	C
P8	E	E	E	E	E
P9	S	S	S	S	S

SE

Moda5					
Posición	Frecuencia Vecino ¹				
	1995	1996	1997	1998	1999
P1	NW	NW	NW	NW	NW
P2	N	N	N	N	N
P3	NE	NE	NE	NE	NE
P4	SE	SE	SE	SE	SE
P5	W	W	W	W	W
P6	W	W	W	W	W
P7	W	W	W	W	W
P8	E	E	E	W	W
P9	S	S	S	S	S

C

SW

Moda4					
Posición	Frecuencia Vecino ¹				
	1995	1996	1997	1998	1999
P1	NW	NW	NW	NW	NW
P2	N	N	N	N	N
P3	NE	NE	NE	NE	NE
P4	W	W	W	W	W
P5	E	E	E	E	E
P6	S	C	C	S	NE
P7	E	E	E	E	E
P8	SE	S	SE	SE	S
P9	S	S	SW	S	S

-

De los modelos Moda4, Moda5 y Moda1 la respuesta mas frecuente del árbol es 11, 12 y 15, correspondiente a valores en los rangos de 0-9, 0-1 y 0-20 milímetros respectivamente. Los valores menos frecuentes corresponden a las etiquetas que contienen los valores altos de precipitación. Ver Tabla 6-3, Tabla 6-4y Tabla 6-5

Tabla 6-3. Resumen de estadística de reglas para el modelo Moda4 en el año 1995.

Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³	Rango		Etiqueta
	Max	Mín					
NW	12	16	11	1	0	9,80	11
N	12	16	12	2	9,80	20,00	12
NE	12	15	13	1	20,00	30,20	13
W	13	15	14	1	30,20	40,90	14
C	12	11	15	1	40,90	54,10	15
E	13	15	16	1	54,10	76,00	16
SW	13	16	17	#N/A	76,00	200,00	17
S	13	15	18				
SE	13	16	19				
Ct	11	14	20				
			21				
			22				

Tabla 6-4. Resumen de estadística de reglas para el modelo Moda5 en el año 1995.

Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³	Rango		Etiqueta
	Max	Mín					
NW	14	23	11	2	0	1,00	11
N	14	23	12	1	1,00	5,00	12
NE	14	23	13	1	5,00	10,00	13
W	14	23	14	1	10,00	15,00	14
C	13	22	15	2	15,00	20,00	15
E	14	22	16	2	20,00	25,00	16
SW	14	23	17	1	25,00	30,00	17
S	14	23	18	1	30,00	35,00	18
SE	14	23	19	1	35,00	40,00	19
Ct	12	23	20	1	40,00	45,00	20
			21	1	45,00	50,00	21
			22	1	50,00	55	22

Tabla 6-5. Resumen de estadística de reglas para el modelo Moda1 en el año 1995.

Vecino	Frec. Valor ²		Etiqueta	Frecuencia Vecino ³	Rango		Etiqueta
	Max	Mín					
NW	15	17	11	1	0	0,20	11
N	12	17	12	2	0,20	0,50	12
NE	15	17	13	2	0,50	0,70	13
W	14	17	14	2	0,70	1,00	14
C	13	17	15	1	1,00	5,00	15
E	12	17	16	2	5,00	10,00	16
SW	14	17	17	1	10,00	1000,00	17
S	13	17	18				
SE	14	17	19				
Ct	15	14	20				
			21				
			22				

7 RESULTADOS Y DISCUSIÓN

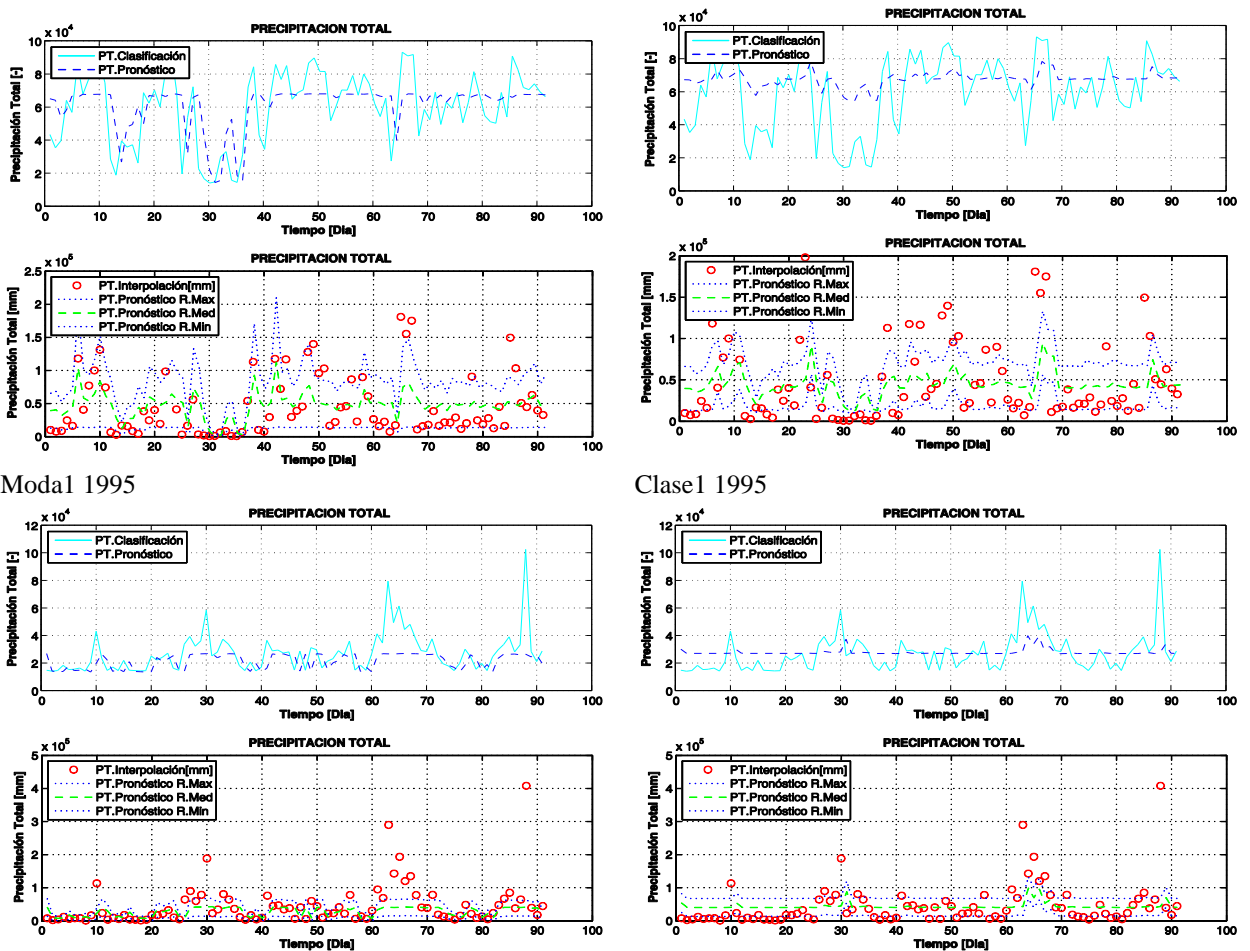
Respecto a la creación de los campos de precipitación, los resultados de la validación cruzada recomiendan que de los tres métodos utilizados el *IDW* es el que mejor representa el comportamiento espacial de la precipitación, debido a que presenta los menores errores absolutos en la comparación de los valores interpolados y los observados. El tamaño de la celda fue asignado mediante el análisis del comportamiento de los centros de tormenta de 19 eventos con resolución de 30 minutos. Se buscó que el tamaño de celda no superara la distancia mínima más frecuente del desplazamiento del centro de la tormenta. Además, también se asumió el criterio de comparar los resultados al interpolar con diferentes tamaños de celda (100, 200, 250 y 500 metros), concluyendo que con menores resoluciones la eficacia no mejora de manera sustancial. Se tomó entonces una resolución de celda igual a 500 metros, alineadas en dirección norte-sur.

Con el fin de verificar los resultados de la implementación de la metodología de pronóstico de precipitación con autómatas celulares se verificaron las medidas de desempeño propuestas de acuerdo con los resultados expuestos en los capítulos anteriores. A nivel diario se implementó el modelo para una red de 77 estaciones, con el fin de obtener el pronóstico de precipitación total diaria en la época húmeda comprendida entre los meses de marzo, abril y mayo para cada uno de los cinco años disponibles (año 1995 al 1999). Se plantearon 12 modelos de clasificación y obtención de patrones. Los modelos con una mejor discretización en los valores de baja precipitación representan de mejor manera la dinámica del fenómeno que es el caso de los modelos: Clase1, Moda1, Clase5, Moda5 y Clase6. Cabe destacar que un elemento importante en la obtención de reglas es la resolución de conflicto de patrones repetidos. Se observa que para una misma clasificación, pero con diferente técnica de resolución de conflicto los resultados son diferentes (Clase1-Moda1 y Clase5 - Moda5). Los modelos donde la técnica de resolución de conflictos fue el uso de la salida más frecuente (Ver numeral 5.2.1.4) usan mayor número de reglas suministradas por el árbol de decisión y produce menores errores absolutos en el mapa de pronóstico, además al comparar las gráficas de precipitación total observada con la pronosticada dan mejores resultados.

El modelo “Moda1” es uno de los que mejor representa la dinámica de la precipitación. Como ya se observó en la Figura 6-1, los errores relativos en general alcanzan a tener valores bajos en casi el 60% de la cobertura espacial del mapa de “aciertos”. Además, en la Figura 7-2 (Líneas azul y cian) se observa que la variación de la precipitación total para cada día en el periodo de interés se asemeja a la estimada en los mapas de clasificación. Se identifican varios aspectos en la gráfica: 1) Existe una similitud en la textura de las líneas. Se infiere entonces que se mantiene la tendencia en los “altibajos”, a pesar que es poco frecuente que en los días con precipitaciones altas los valores de precipitación total se parezcan. Ello concluye unas de las debilidades del modelo que es la “Dificultad en reproducir condiciones extremas de precipitación”. 2) Otro aspecto es la presencia de un rezago en las líneas de precipitación total. Este hecho se le atribuye a la inercia de la que se

había hablado en capítulos anteriores, en la que un mapa de pronóstico del día t se parece al mapa clasificado del día $t-1$. Es así como los valores totales de precipitación total pronosticada se ven rezagados un día respecto a la precipitación total clasificada. En la figura de precipitación total en unidades de milímetros, se observa que la precipitación total observada está cerca del límite inferior del corredor delimitado por los rangos de precipitación interpolada, tendiendo frecuentemente al límite inferior. En términos generales se considera una buena estimación, bajo la salvedad de que los valores altos no son fáciles de reproducir. Esta medida de desempeño se considera de gran utilidad pues dicta órdenes de magnitud de la precipitación total en un área específica.

Comparando estos análisis con las gráficas del modelo Clase1 es evidente que el mejor rendimiento lo tiene Moda1. Existen épocas donde el modelo Clase1 ni siquiera es capaz de reproducir precipitaciones bajas como en los días 15 y 35. También se ve que la precipitación total interpolada esta por fuera del límite superior del rango de precipitación clasificada con más recurrencia que en el modelo Moda1.



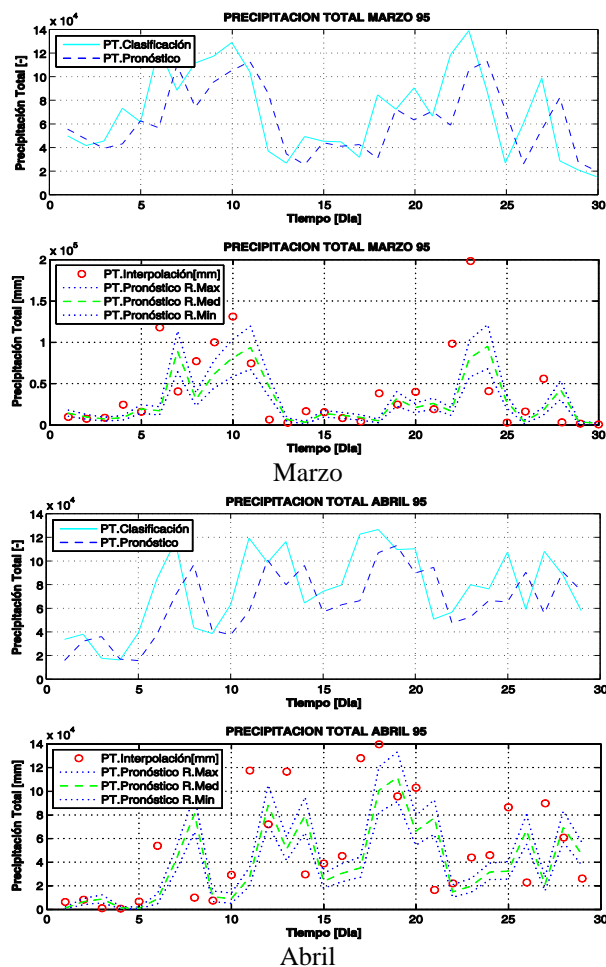
Moda5 año 1998

Clase5 año 1998

Figura 7-1. Variación de la precipitación total diaria para el periodo marzo, abril y mayo en los modelos Moda1, Clase1, Moda5 y Clase5.

También en el caso de la precipitación total para Moda5, el modelo no logra reproducir las precipitaciones altas. Se observan variaciones temporales en la línea de la precipitación total, mientras que en el modelo Clase5 parecería tener una condición “constante” hasta el instante 65, donde se observa un indicio de simular el pico de la línea de precipitación total observada en ese instante del tiempo. Este fenómeno no se observa en el gráfico de Moda5. Por lo general los valores de precipitación total interpolada están dentro del rango asignado por la clasificación.

El modelo Clase6 tiene un gran número de patrones y reglas, por lo que computacionalmente sólo se logró simular los tres meses lluviosos del año 1995. La Figura 7-2 ratifica la presencia de inercia en el modelo, mostrando el rezago de un día, las texturas de las líneas de precipitación total son similares y se encuentran mejores acercamientos de los valores de la precipitación total interpolada a la franja de rango de precipitación clasificada para el mes de mayo.



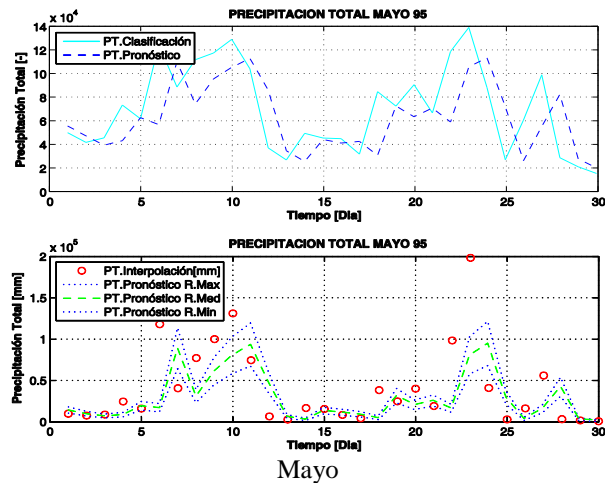
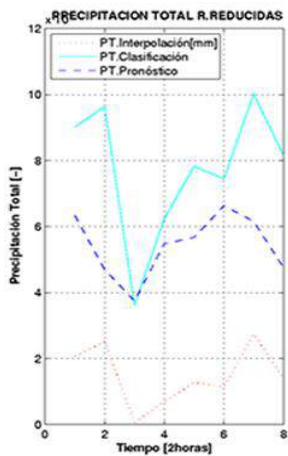


Figura 7-2. Variación de la precipitación total para el modelo "Clase6" en el año 1995.

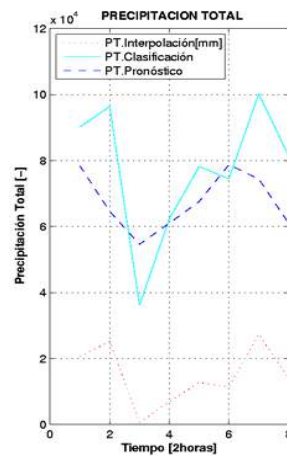
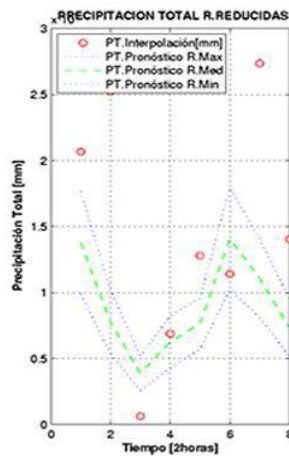
El análisis de los resultados del modelo de pronóstico con autómatas celulares a nivel de los eventos de tormenta a diferentes agregaciones temporales se concentró en varias líneas de estudio: 1) análisis de los resultados con el modelo convencional y 2) análisis de sensibilidad variando elementos del autómatas. El estudio se concentró en la tormenta ocurrida en abril 12 de 1995. El pronóstico de la precipitación haciendo uso del autómatas para la resolución de 2 horas (

Figura 5-6) y 30 minutos (Figura 5-13) muestra lo importante que es el detalle en el momento de querer observar comportamientos característicos de la precipitación como la emergencia y la complejidad. A niveles de agregación temporal más finos la dinámica en el espacio y el tiempo es sustancial, los errores relativos son bajos y muestran un buen desempeño alrededor del 70% del mapa de aciertos. En cuanto al análisis de sensibilidad, se recomienda actualizar el mapa de condiciones iniciales del modelo cada instante de tiempo, con el fin de evitar propagación de errores y de los contornos de las concentraciones de precipitación a través del tiempo. Este aspecto podría cambiar si se incorporaran nuevas variables de estudio y si en la construcción de la tabla de patrones se añadiera la opción de 2, 3 o más tiempos o rezagos para la toma de decisión del estado de una celda en el tiempo futuro. En cuanto a la reducción del número de reglas se observa un buen desempeño de la tormenta cada 2 horas, si se usan solo el 20% de las reglas del árbol de decisión. Este es un gran aporte de la investigación pues de alguna manera deja abierta la inquietud de si se podrían conformar un conjunto de reglas que definieran el comportamiento de la lluvia. El inconveniente que se puede presentar ante esta situación es la inexistencia condiciones en el árbol después de haber eliminado algunas reglas. Incluso esta situación ocurre cuando algún patrón está mal clasificado en el árbol de decisión. Se planteó la utilización de un filtro de paso bajo para estimar los valores de las celdas que no lograron tomar la decisión con las reglas reducidas o inexistentes.

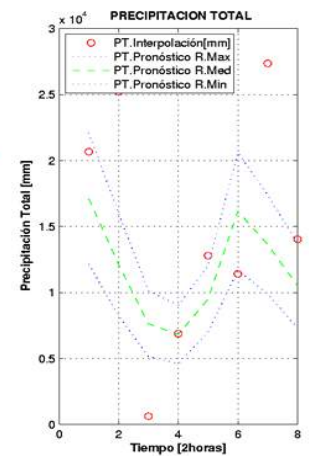
Al realizar el filtraje los resultados cambian sutilmente en algunos instantes de tiempo. Por ejemplo, al utilizar solo 33 reglas de las 1704 deducidas por el árbol de decisión para el caso de la tormenta de abril 12 se observa que por lo general la precipitación total estimada por el pronóstico es inferior a la precipitación total de la observada. Aplicando el filtro se muestra que los valores de precipitación total aumentan un poco, acercándose a los valores medios de la precipitación observada.



Reglas reducidas por frecuencia de salida (33 reglas)



Después del filtraje



8 CONCLUSIONES Y RECOMENDACIONES

La presente investigación reporta los resultados de un modelo guiado por datos para hacer el pronóstico de precipitación. Se reconocen las limitantes como la falta de información detallada en el espacio y las dificultades en los software utilizados en términos de tiempo computacional o capacidad de cómputo. Los resultados se consideran un avance en el estudio de la teoría del autómata celular en el pronóstico de variables meteorológicas por lo que se considera una investigación base para profundizar en el desarrollo de modelos de autómatas más sofisticados. Se considera que se deben incluir más variables involucradas en el proceso de formación de lluvia, mejorar los estimativos de los campos de precipitación como sería el caso de obtenerlos con un radar, y buscar la integración con otros modelos hidrológicos. De otro lado, también se deben buscar otras alternativas de lenguaje de programación, equipos o sistemas operativos, pues un volumen de información superior al utilizado en esta investigación la capacidad de cómputo se reduciría considerablemente.

La falta de instrumentación en la red de monitoreo genera una incertidumbre espacial al tener que generar los campos de precipitación con técnicas de interpolación, más aun cuando la densidad no es la deseada. Aun así que realizaron algunas consideraciones como el tamaño de celda, alineación de la malla y validación cruzada para intentar establecer lo mejor posible los campos de precipitación. El análisis del desplazamiento de los centros de tormenta dio evidencia de lo complejo que es el fenómeno de la precipitación y sirvió como base de referencia para la elección del tamaño de celda de los campos de precipitación. Se sugirió que el tamaño apropiado del tamaño de las celdas no excediera la distancia mínima del movimiento del centro de tormenta. Bajo este supuesto se analizaron 19 tormentas y se estableció que el tamaño de celda fuera 500 metros. Distancias inferiores implican más tiempo computacional y no mejora de manera contundente los resultados de la validación cruzada. La aplicación de tres diferentes métodos de interpolación (inverso de la distancia, V4 y Krigging) con variaciones en el tamaño de celda ha permitido concluir que para el periodo de estudio el mejor desempeño lo tiene el método IDW. Esta selección se llevó a cabo mediante el proceso de validación cruzada en la que se extrajeron del cálculo de interpolación 5 estaciones de precipitación diaria localizadas estratégicamente en 5 zonas de la ciudad de Bogotá de las 77 estaciones de precipitación diaria para llevar a cabo ese cálculo.

La metodología propuesta para el pronóstico de precipitación con autómatas celulares es sencilla de aplicar en otras ciudades o regiones, tanto en pequeñas como en grandes extensiones de territorio. Quizás el componente que más condiciona el uso de la metodología en zonas donde la adquisición de la información de lluvia es a partir de estaciones puntuales es la disponibilidad y densidad de estaciones en la zona de estudio para la generación de los campos de precipitación.

El diseño experimental comprende la generación de varios campos simulados a diferentes escalas temporales y la variación de las reglas de evolución. Los resultados

obtenidos permiten evidenciar la fortaleza y robustez del modelo de autómeta celular para el problema de pronóstico. Se demostró que para el período comprendido entre el 01-ene-1995 al 31dic-1999 la metodología propuesta para pronóstico de precipitación diaria con autómetas celulares planteando varios modelos de clasificación funciona de manera adecuada para algunos de ellos. Los resultados muestran variaciones en la dinámica espacial y temporal. De igual forma se observaron situaciones en las que el la precipitación total del pronóstico se acerca a el valor de precipitación total del área de estudio.

Al analizar la distribución espacial de la precipitación diaria se observa que las mayores concentraciones se localizan en el corredor de los cerros orientales y la zona centro de la ciudad. De igual forma ocurre con los eventos analizados. Se recomienda que el estudio de estas características podría dar indicios de una zonificación dentro del área de estudio. Esto posiblemente contribuiría en la reducción del número de patrones y reglas que se ve representado en menor tiempo de cómputo y de esta manera se podría involucrar otras variables y hacer análisis bajo otras condiciones de vecindad y tipo de clasificación.

El modelo es capaz de reproducir la variabilidad a nivel temporal y espacial. Temporalmente, las gráficas de precipitación total son un indicativo de cómo varia la cantidad de lluvia precipitada, los valores en el tiempo aumentan o disminuyen durante el periodo en consideración. Espacialmente se observa la variación de la ubicación de las zonas con concentración de lluvia a pesar que en ocasiones se observen valores similares en localizaciones iguales a las que tenía una celda en el tiempo anterior, debido a que el modelo está concebido con una fuerte dependencia a la precipitación del instante de tiempo inmediatamente anterior.

Con el fin de realizar un estudio más detallado de las potencialidades de los modelo de autómeta celular en el campo de la hidrología, se recomienda desarrollar modelos de pronóstico con AC acoplados con técnicas como “*Data assimilation*” que permita dar una visión más detallada del comportamiento de la formación de la precipitación, su movimiento en el espacio, la verificación de la influencia de componentes geográficos como el relieve y los usos de suelo.

9 TRABAJO FUTURO

De acuerdo con los resultados obtenidos y con el fin de darle continuidad al presente trabajo se proponen las siguientes investigaciones:

- Verificar si un modelo de pronóstico con autómeta celular logra sugerir la presencia o no de una anomalía climática como el ENSO.
- Desarrolla un modelo de AC que considere otras variable hidrometeorológicas y así identificar cuáles son los factores que juega un papel relevante en la formación de la precipitación. En ese sentido sería productivo platear una red de medición más eficiente o la incorporación de nuevas tecnologías como los sensores remotos.
- Comparar los resultados de simulación de un AC con los desarrollados por otros modelos como el MM5 o *Weather Research and Forecasting* -WRF.
- Realizar un análisis temporal de los patrones de precipitación con el fin de identificar si las actividades antrópicas en especial las industriales han contribuido a los cambios en la precipitación.
- Diseñar una metodología para el acople de los resultados del pronóstico con AC y modelos distribuidos existentes en el mercado como SWAT.
- Incorporar otras técnicas de minería de datos (ej. redes neuronales o algoritmos genéticos) para obtención de las reglas de evolución del autómeta celular.
- Segmentar el área de estudio por regiones implementando la metodología en cada una de ellas por separado y luego comparar los resultados con los obtenidos sin haber regionalizado el espacio.
- Evaluar el desempeño de un modelo de AC para el pronóstico de precipitación en distintas ciudades localidades, de manera que se puedan determinar los límites espacio-temporales de aplicación de las reglas obtenidas.
- Diseñar un modelo que permita implementar la metodología desarrollada a datos meteorológicos obtenidos mediante sensores remotos, con el fin de realizar el estudio de variables meteorológicas donde la adquisición de información se realiza mediante técnicas modernas

BIBLIOGRAFÍA

- Anhert, P., Krajewski, W., Johnson, E., 1986. Kalman filter estimation of radar rainfall field bias, *Memorias XXIII Conferencia de radar meteorológico*, Amer. Meteor. Soc., pp JP 33-37.
- Ashraf, M., Loftis, J.C., Hubbard, K.G., 1997. Application of geostatistics to evaluate partial weather station networks. *Agricultural and Forest Meteorology* . Número 84, pp 255–271.
- Austin, P., Houze, R. 1972. Analysis of the structure of precipitation patterns in New England. *Journal of applied meteorology*. Volumen 11. Número 6. pp 926–935.
- Barnston, A.G., Livezey, R., 1987. Classification, seasonality and persistence of low frequency atmospheric circulation patterns. *Monthly Weather Review* . Volumen 115. pp 1083–1126.
- Bartier, P.M., Keller, C.P., 1996. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). *Computers & Geosciences* . Volumen 22. Número 7. pp 795–799.
- Bernal, F.A., 2007. Desarrollo de modelos conceptuales y computacionales para simular estocásticamente la dinámica espacio - temporal de la precipitación en Bogotá. Tesis de Maestría en Recursos Hidráulicos, Universidad Nacional de Colombia.
- Brown, D., *et al.*, 2002. Stochastic simulation of landcover change using geostatistics and generalized additive models. *Photogrammetric Engineering & Remote Sensing*. Volumen 10. Número 68. pp 1051-1061.
- Bodri, L., Čermák, V., 2000. Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia. *Advances in Engineering Software*. Volumen 31. pp 311–321.
- Burlando, P., *et al.*, 1993, Forecasting of short-term rainfall using ARMA models *Journal of Hydrology*. Volumen 144. Número 1-4. pp 193-211.
- Burrough, P., McDonnell. R., 1998. Principles of geographical information systems. New York, Oxford University Press. 333 p.

- Bras, R., Rodríguez-Iturbe, I., 1994. Random Functions and Hydrology. Dover Books on Advanced Mathematics. New York.
- Carr, J. 1995. Numerical Analysis for the Geological sciences. Prentice Hall, New Jersey.
- Chambers, J. M., et al., 1983. Graphical Methods for Data Analysis. Duxbury Press. Boston.
- Cheng, C.T., Ou, C.P., Chau, K.W., 2002. Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. *Journal of Hydrology* . Número 268. pp 72–86.
- Chen, W. 2004. Cellular Automata and Artificial Intelligence. En Ecohydraulics Modelling. Ed Taylor & Francis Group. Londres
- Chuvieco. E. 2002. Teledetección ambiental. Editorial Ariel. España.
- Clark, I. (1979). Practical Geostatistics. Blackwell Science.
- Coppola, E., et al., 2007. Cellular automata algorithms for drainage network extraction and rainfall data assimilation. *Hydrological Sciences–Journal–des Sciences Hydrologiques*. Volumen 3. Número 52. pp 579-592.
- Cooper, N.G., 1987. Los Alamos Science Special Issue. Los Alamos National Laboratory, New México.
- Crisci, G., Rongo, R., Di Gregorio, S., Spataro, W., 2004. The simulation model SCIARA: the 1991 and 2001 lava flows at Mount Etna. *Journal of Volcanology and Geothermal Research*. Volumen 3010. pp1-15.
- Departamento Técnico Administrativo del Medio Ambiente (DAMA). 2006. Informe anual de calidad del aire de Bogotá año 2005. Bogotá.
- Dalonte D., et al., 2007. A validated fast algorithm for simulation of flooding events in plains. *Hydrological Processes*. Volumen 21. pp 1115-1124.
- D'Odorico P., Rodríguez-Iturbe, I., 2000. Space-time self-organization of mesoscale rainfall and soil moisture. *Advances in Water Resources*. Volumen 23. pp 349-357
- Gardner, M., 1970. Mathematical Games - The fantastic combinations of John Conway's new solitaire game Life. *Scientific American*. Volumen 223. pp 120-123.

- Gamal, A., Smith, D., 2002. A neural network model to predict the wastewater inflow incorporating rainfall events. *Water Research*. Volumen 36. pp 1115–1126.
- Gar-On Yeh, A., Li X. 2004. Data mining of cellular automata's transition rules. *Int. J. Geographical information science*. Volumen 18. Número 8. pp 723–744.
- Georgakakos, K. P., Krajewski, W. F., 1994. Stochastic-dynamical generation of tropical rainfall fields. *Annales Geophysicae*, 12(II): C406.
- Giraldo, R., 2002. Introducción a la Geoestadística, teoría y aplicación. Universidad Nacional de Colombia. Bogotá D.C. 94 páginas.
- Hernandez, J., et. al. 2008. Minería de datos. Pearson Prentice Hall. España.
- Hsu, K-L., Gupta, H.V., Gao, X., 2000. Rainfall estimator from satellite imagery. En Artificial Neural Networks in Hydrology, Water Science and Technology Library. USA. pp 209-234.
- Huang, X., Jensen, J. R., 1997, A machine-learning approach to automated knowledge-base building for remote sensing image analysis with GIS data. *Photogrammetric Engineering & Remote Sensing*. Volumen 63. pp 1185–1194.
- Hipel, K., McLeod, I., 1994. Time Series Modelling of Water Resources and Environmental Systems. ELSEVIER
- Ibañez, J.J., Machado, C., 1995. Análisis de la variabilidad espacio-temporal y procesos caóticos en ciencias medioambientales. Geoforma Ediciones. pp 9-15.
- Ilachinski, A., 2001. Cellular automata: a discrete universe. World Scientific. 808 pp.
- Isaaks, E. y Srivistava, R., 1989. Applied Geostatistics. Oxford University Press, New York.
- Jenkins, G. M., Watts, D. G., 1968. Spectral Analysis and its Applications. Holden-Day, San Francisco.
- Jiménez, A., Posadas, A., Marfill, J.M., 2000. A probabilistic seismic hazard model based on cellular automata and information theory. *Nonlinear Processes in Geophysics*. Volumen 12. pp 1–16.
- Jiménez, A., et al., 2008b. Variabilidad Espacio-Temporal de la Precipitación y su Impacto Hidrológico. XVIII Seminario Nacional de Hidráulica e Hidrología- Bogotá D.C.

- Karafyllidis, I., Thanailakis, A., 1997. A model for predicting forest fire spreading using cellular automata. *Ecological Modelling*. Volumen 99. pp 87- 97.
- Kocabas, V., Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban System*. Volumen 30. Número 6. pp 921-953.
- Krajewski, W.F., 1987. Cokriging radarrainfall and rain gage data. *Journal of Geophysical Research*. Volumen 92. Número D8. pp 9571-9580.
- Krzysztofowicz, R., Maranzano, C. J., 2004. Bayesian system for probabilistic stage transition forecasting. *Journal of Hydrology*. Volumen 299. pp 15-44
- Lam, N., 1983. Spatial interpolation methods: a review. *The American Cartographer*. Volumen 10(2). pp 129-149
- Mejía, J. F., Poveda, G, 2005. Ambientes Atmosféricos de Sistemas Convectivos de Mesoescala sobre Colombia durante 1998 según la misión TRMM y el re-análisis NCEP/NCAR. *Rev. Acad. Colomb. Cienc*. Volumen 29. Numero 113. pp 495-514. ISSN 0370-3908.
- Mitchell T. 1997. Machine Learning. McGraw Hill, New York (1997).
- Mora, L., *et al.* 2005. Modeling time series of climatic parameters with probabilistic finite automata . *Environmental Modelling & Software*. Número 20. pp 753-760
- Muñoz, J.D., 1996. Autómatas Celulares Y Física Digital, en: *Memorias del Primer Congreso Colombiano de NeuroComputación*. Santafé de Bogotá, D. C.: Academia Colombiana de Ciencias Exactas, Físicas y Naturales, 28 p. ISBN 958-9205-17-8.
- Nauffal, M. S, Obregón, N. N, 2008. Dinámica espacio - temporal de tormenta registrada en Bogotá en el año 1993. En XVIII Seminario Nacional de Hidráulica e Hidrología. Bogotá.
- Ortiz D. P., Penagos J.C., Lizcano A., 1995. Determinación de lluvias críticas que activan deslizamientos empleando redes neuronales. Universidad de los Andes, Bogotá D.C., Colombia.
- Packard, N. H., Wolfram, S., 1985. Two-dimensional cellular automata. *J. Statist. Phys*. Volumen 38. pp 901-946.

- Preis A., Ostfeld A., 2008. A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds. *Journal of Hydrology*. Volumen 349, pp 364– 375
- QI, Y., WU, J., 1996. Effects of changing spatial resolution on the results of landscape pattern analysis using autocorrelation indices. *Landscape Ecology*. Volumen 1. Número 11. pp 39-49.
- Rinaldi P., *et al.*, 2007. Cellular automata algorithm for simulation of surface flows in large plains. *Simulation Modeling Practice and Theory*. Volumen 15. pp 315-327.
- Rodríguez-Iturbe, I., Febres De Power, B., Sharifi, M.B., Georgakakos, K.P., 1989. Chaos in rainfall. *Water Resources Research*. Volumen 25. Número 7. pp 667-1675.
- Rodríguez, E., *et al.*, 2008a. Análisis hidrometeorológico granizada sobre el campus de la Universidad Nacional – Bogotá. *XVIII Seminario Nacional de Hidráulica e Hidrología- Bogotá D.C.*
- Rodríguez, E., *et al.*, 2008b. Variabilidad espacio-temporal de la precipitación en una microcuenca urbana. *XVIII Seminario Nacional de Hidráulica e Hidrología- Bogotá D.C.*
- Schärer M., Page T., Beven K., 2006. A fuzzy decision tree to predict phosphorus export at the catchment scale. *Journal of Hydrology*. Volumen 331, pp 484– 494.
- Sivakumar, B., Liong, S.-Y., Liaw, C.-Y., 1998. Evidence of chaotic behavior in Singapore rainfall. *Journal of the American Water Resources Association*. Volumen 34. Número 2. pp 301–310.
- Sivakumar. B., 2000. Chaos theory in hydrology: important issues and interpretations. *Journal of Hydrology*. Volumen 227. pp 1–20.
- Sui, D. Z., Zeng, H. 2001. Modeling the dynamics of landscape structure in Asia’s emerging desakota regions: a case study in Shenzhen. *Landscape and Urban Planning*. Volumen 1-4. Número 53. pp 37-52.
- Tukey, J. W., 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.
- Vargas, A. *et. Al.* 2010. Consideraciones en la Estimación de los Campos de Precipitación en la Ciudad de Bogotá. *XXIV Congreso Latinoamericano de Hidráulica*. Punta del Este, Uruguay. En evaluación.

- Von Neumann, J. 1996. Theory of Self-Reproducing Automata. University of Illinois Press, Illinois. Edited and completed by A.W. Burks.
- Véneré, M. J., *et al.*, 2004. Simulación de Inundaciones en Llanuras. Aplicación a la Cuenca del Arroyo Santa Catalina-Azul. *Mecánica Computacional*. Volumen 23. pp 1135-1150.
- Ware, C., Knight, W., Wells, D., 1991. Memory intensive algorithms for multibeam bathymetric data. *Computers & Geosciences*. Volumen 17. Número 7. pp 985–993.
- Weimar, J. 1997. Simulation with cellular automata. Logos Verlag Berlin. 199 pp.
- Wolfram, S., 1986. Cellular automata fluids. *Journal of Statistical Physics* . Volumen 45. pp 3-4.
- Wolfram, S., 1986a. Theory and Applications of Cellular Automata. World Scientific Press, Singapore.
- Wolfram, S., 2002. A New Kind of Science. USA: Wolfram Media Inc, Champaign, IL. 1197 pp.
- Wei. C, Hsu. N, 2009. Optimal tree-based release rules for real-time flood control operations on a multipurpose multireservoir system. *Journal of Hydrology*. Volumen 365. pp 213–224.
- Zawadzki, I.I., 1973. Statistical Properties of Precipitation Patterns. *Journal of Applied Meteorology*. Volumen 12. Número 3. pp 459-472.
- Zhou, G., Liebhold, A., 1995. Forecasting the Spatial Dynamics of Gypsy-Moth Outbreaks Using Cellular Transition Models. *Landscape Ecology*. Volumen 3. Número 10. pp 177-189.

ALGORITMOS

Algoritmo 1. Interpolación especial con el método del inverso cuadrático de la distancia

```
% 1.1 Lectura de las matrices de los mapas de salida CT

% DELTA DE 2H
PRECIPITACION=xlsread('Evento7.xls','2h','I3:A112');
DIA_INICIAL=datenum('12-abr-1995');
DIA_FINAL=datenum('20-abr-1995');
% 1.2 Lectura de Archivo de entrada-Coordenadas
COORDENADAS=xlsread('Evento7.xls','coordenadas','B3:C28');
% % Crea la matriz de las fechas
MATRIZ_NUMERICA_FECHAS=(DIA_INICIAL:DIA_FINAL);
MATRIZ_CARACTER_FECHAS=datestr(MATRIZ_NUMERICA_FECHAS,'mmmD
DYyyy');

%% 2. DEFINICION DE PARAMETROS DE LA INTERPOLACION %%
% 2.1 Valores de la grilla
delta=500;
Emin=970000; Emax=1020000; Nmin=960000; Nmax=1030000;
[MXMalla,MYMalla] = meshgrid(Emin:delta:Emax, Nmax:-delta:Nmin);
% 2.2 Numero de dias y de estaciones activas
TAMANO_DATOS=size(PRECIPITACION)-1;
NUMERO_DE_DIAS=TAMANO_DATOS(1);
NUMERO_DE_ESTACIONES=TAMANO_DATOS(2);
%% 2.3 METODO 1: Metodo del inverso de la distancia cuadrada (IDW)%%
METODO=1;
k(1:NUMERO_DE_DIAS,1)=0;

for i=1:NUMERO_DE_DIAS
% 2.3.1 Vector de precipitacion y coordenadas
PRECIPITACION_AUXILIAR=NaN;
COORDENADAS_AUXILIARES_X=NaN;
COORDENADAS_AUXILIARES_Y=NaN;
for j=1:NUMERO_DE_ESTACIONES
if PRECIPITACION(i+1,j+1) >= 0
k(i,1)=k(i,1)+1;
PRECIPITACION_AUXILIAR(k(i,1),1)=PRECIPITACION(i+1,j+1);
COORDENADAS_AUXILIARES_X(k(i,1),1)=COORDENADAS(j,1);
COORDENADAS_AUXILIARES_Y(k(i,1),1)=COORDENADAS(j,2);
end
end
end
```

```
%% 2.3.2 Interpolacion IDW
if METODO == 1
    NUMERADOR_PRINCIPAL=0;
    DENOMINADOR_PRINCIPAL=0;
    for m=1:k(i,1)
        DIFERENCIA_X=(COORDENADAS_AUXILIARES_X(m,1)-MXMalla);
        DIFERENCIA_Y=(COORDENADAS_AUXILIARES_Y(m,1)-MYMalla);

        DISTANCIA=sqrt(DIFERENCIA_X.*DIFERENCIA_X+DIFERENCIA_Y.*DIFERENCIA_Y);
        DISTANCIA_2=DISTANCIA.*DISTANCIA;
        NUMERADOR=PRECIPITACION_AUXILIAR(m)./DISTANCIA_2;
        DENOMINADOR=1./DISTANCIA_2;
        NUMERADOR_PRINCIPAL=NUMERADOR_PRINCIPAL+NUMERADOR;

    DENOMINADOR_PRINCIPAL=DENOMINADOR_PRINCIPAL+DENOMINADOR;

    % 2.3.2.1 Guardar la matriz interpolada en una estructura

    PRECIPITACION_INTERPOLADA.(MATRIZ_CARACTER_FECHAS(i,1:9))=NUMERADOR_PRINCIPAL./DENOMINADOR_PRINCIPAL;
    end
end
end
%% 3. PRECIPITACION POR PIXEL
PRECIPITACION_PIXEL=NaN;
for i=1:NUMERO_DE_DIAS

    PRECIPITACION_PARA_PIXELES=PRECIPITACION_INTERPOLADA.(MATRIZ_CARACTER_FECHAS(i,1:9));

    TAMANO_MATRIZ_PRECIPITACION_PARA_PIXELES=size(PRECIPITACION_PARA_PIXELES);
    for j=1:TAMANO_MATRIZ_PRECIPITACION_PARA_PIXELES(1)-1
        for k=1:TAMANO_MATRIZ_PRECIPITACION_PARA_PIXELES(2)-1

            PRECIPITACION_PIXEL(j,k)=(PRECIPITACION_PARA_PIXELES(j,k)+PRECIPITACION_PARA_PIXELES(j,k+1)+PRECIPITACION_PARA_PIXELES(j+1,k)+PRECIPITACION_PARA_PIXELES(j+1,k+1))/4;
            end
        end
    PRECIPITACION_POR_PIXEL.(MATRIZ_CARACTER_FECHAS(i,1:9))=PRECIPITACION_PIXEL;
end
End
```

Algoritmo 2. Construcción de la tabla de patrones

```
% 4.1.1 Clasificación tipo 1
RANGO=[0 0.2 0.5 0.7 1.0 5.0 10.0 1000];
CLASIFICADOR=[1 2 3 4 5 6 7];

for i=1:NUMERO_DE_DIAS

TAMANO_MATRIZ_INTERPOLACION=size(PRECIPITACION_POR_PIXEL.(MATRIZ_CARACTER_FECHAS(i,1:9)));
    for j=1:TAMANO_MATRIZ_INTERPOLACION(1)
        for k=1:TAMANO_MATRIZ_INTERPOLACION(2)

l=find(RANGO>PRECIPITACION_POR_PIXEL.(MATRIZ_CARACTER_FECHAS(i,1:9))(j,k));

PRECIPITACION_CLASIFICADA.(MATRIZ_CARACTER_FECHAS(i,1:9))(j,k)=CLASIFICADOR(l(1)-1);
            end
        end
    end

% 4.2 Lectura de la plantilla por los píxeles de la matriz precipitación
% para estimar la tabla de patrones. Plantilla de MOORE radio 1, matriz 9*9
for i=1:NUMERO_DE_DIAS

MATRIZ_AUXILIAR=PRECIPITACION_CLASIFICADA.(MATRIZ_CARACTER_FECHAS(i,1:9));
    TAMANO_MATRIZ_AUXILIAR=size(MATRIZ_AUXILIAR);
    NUMERO_DE_FILAS=(TAMANO_MATRIZ_AUXILIAR(1)-2)*(TAMANO_MATRIZ_AUXILIAR(2)-2);
    MATRIZ_MOORE=NaN;
    j=1;
    for k=2:TAMANO_MATRIZ_AUXILIAR(1)-1
        for l=2:TAMANO_MATRIZ_AUXILIAR(2)-1
            MATRIZ_MOORE(j,1)=MATRIZ_AUXILIAR(k-1,l-1);
            MATRIZ_MOORE(j,2)=MATRIZ_AUXILIAR(k-1,l);
            MATRIZ_MOORE(j,3)=MATRIZ_AUXILIAR(k-1,l+1);
            MATRIZ_MOORE(j,4)=MATRIZ_AUXILIAR(k,l-1);
            MATRIZ_MOORE(j,5)=MATRIZ_AUXILIAR(k,l);
            MATRIZ_MOORE(j,6)=MATRIZ_AUXILIAR(k,l+1);
            MATRIZ_MOORE(j,7)=MATRIZ_AUXILIAR(k+1,l-1);
            MATRIZ_MOORE(j,8)=MATRIZ_AUXILIAR(k+1,l);
            MATRIZ_MOORE(j,9)=MATRIZ_AUXILIAR(k+1,l+1);
            j=j+1;
        end
    end
end
```

```
    end
  end
  MATRIZ_MOORE_2.(MATRIZ_CARACTER_FECHAS(i,1:9))=MATRIZ_MOORE;
end
%%
for i=1:NUMERO_DE_DIAS

MATRIZ_AUXILIAR=PRECIPITACION_POR_PIXEL.(MATRIZ_CARACTER_FECHAS(i,1:9));
  TAMANO_MATRIZ_AUXILIAR=size(MATRIZ_AUXILIAR);
  NUMERO_DE_FILAS=(TAMANO_MATRIZ_AUXILIAR(1)-2)*(TAMANO_MATRIZ_AUXILIAR(2)-2);
  MATRIZ_MOORE_3=NaN;
  j=1;
  for k=2:TAMANO_MATRIZ_AUXILIAR(1)-1
    for l=2:TAMANO_MATRIZ_AUXILIAR(2)-1
      MATRIZ_MOORE_3(j,l)=MATRIZ_AUXILIAR(k,l);
      j=j+1;
    end
  end

MATRIZ_MOORE_4.(MATRIZ_CARACTER_FECHAS(i,1:9))=MATRIZ_MOORE_3;
end
%%

for i=1:NUMERO_DE_DIAS-1
  PATRONES=NaN;
  PATRONES=MATRIZ_MOORE_2.(MATRIZ_CARACTER_FECHAS(i,1:9));

PATRONES(:,10)=MATRIZ_MOORE_4.(MATRIZ_CARACTER_FECHAS(i+1,1:9))(:,1);
  PATRONES_2.(MATRIZ_CARACTER_FECHAS(i,1:9))=PATRONES;
end

%% Determinacion de los patrones : Deja un solo patron sin repetirlo
l=1;
MATRIZ_INTELIGENTE=NaN;
SUMATORIA=0;
MATRIZ_PATRON=PATRONES_2.(MATRIZ_CARACTER_FECHAS(1,1:9));
MATRIZ_INTELIGENTE(1,1:10)=MATRIZ_PATRON(1,1:10);
for i=1:NUMERO_DE_DIAS-1
  MATRIZ_PATRON=PATRONES_2.(MATRIZ_CARACTER_FECHAS(i,1:9));
  for j=1:NUMERO_DE_FILAS
    m=NaN;
    for k=1:l
```



```
if MATRIZ_PATRON(j,1:9)==MATRIZ_INTELIGENTE(k,1:9);
    m(k,1)=1;
else
    m(k,1)=0;
end
end
if sum(m) > 0
    SUMATORIA(1:l,1)=SUMATORIA+m;
    n=find(m);

MATRIZ_INTELIGENTE(n,10)=MATRIZ_INTELIGENTE(n,10)+MATRIZ_PATRON(j,
10);
else
    MATRIZ_INTELIGENTE(l+1,1:10)=MATRIZ_PATRON(j,1:10);
    l=l+1;
    SUMATORIA(l,1)=1;
end
end
end

MATRIZ_INTELIGENTE(:,11)=SUMATORIA;
MATRIZ_INTELIGENTE(:,12)=MATRIZ_INTELIGENTE(:,10)./MATRIZ_INTELIG
ENTE(:,11);
```

Algoritmo 3. Reproducción del autómata celular

```
REGLAS=xlsread('REGLA_COPIA2H');
ESTADISTICA_DE_REGLAS(1:1:length(REGLAS),1)=0;
for i=1:NUMERO_DE_DIAS-1
    regla_dia=i
    CT.(MATRIZ_CARACTER_FECHAS(i+1,1:9))(1:140,1:100)=NaN;
    for j=2:139
        for k=2:99
            m=0;
            for l=length(REGLAS):-1:1

MATRIZ_REAL=PRECIPITACION_CLASIFICADA.(MATRIZ_CARACTER_FECHAS
(i,1:9))(j-1:1:j+1,k-1:1:k+1);
            MATRIZ_AUXILIAR=MATRIZ_REAL;
            REGLA_ANALISIS=REGLAS(l,1:19);
            if find(REGLA_ANALISIS==1) > 0
                DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==1)+1));
                MATRIZ_AUXILIAR(1,1)=DATO_AUXILIAR-10;
            end
            if find(REGLA_ANALISIS==2) > 0
                DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==2)+1));
                MATRIZ_AUXILIAR(1,2)=DATO_AUXILIAR-10;
            end
            if find(REGLA_ANALISIS==3) > 0
                DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==3)+1));
                MATRIZ_AUXILIAR(1,3)=DATO_AUXILIAR-10;
            end
            if find(REGLA_ANALISIS==4) > 0
                DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==4)+1));
                MATRIZ_AUXILIAR(2,1)=DATO_AUXILIAR-10;
            end
            if find(REGLA_ANALISIS==5) > 0
                DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==5)+1));
                MATRIZ_AUXILIAR(2,2)=DATO_AUXILIAR-10;
            end
            if find(REGLA_ANALISIS==6) > 0
                DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==6)+1));
                MATRIZ_AUXILIAR(2,3)=DATO_AUXILIAR-10;
            end
            if find(REGLA_ANALISIS==7) > 0
                DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==7)+1));
                MATRIZ_AUXILIAR(3,1)=DATO_AUXILIAR-10;
            end
            if find(REGLA_ANALISIS==8) > 0
```

```
DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==8)+1));
MATRIZ_AUXILIAR(3,2)=DATO_AUXILIAR-10;
end
if find(REGLA_ANALISIS==9) > 0
    DATO_AUXILIAR=REGLA_ANALISIS(1,(find(REGLA_ANALISIS==9)+1));
    MATRIZ_AUXILIAR(3,3)=DATO_AUXILIAR-10;
end
if MATRIZ_REAL==MATRIZ_AUXILIAR
    CT.(MATRIZ_CARACTER_FECHAS(i+1,1:9))(j,k)=REGLAS(1,19)-10;
    m=1;
    ESTADISTICA_DE_REGLAS(1,1)=ESTADISTICA_DE_REGLAS(1,1)+1;
    break
end
end
if m == 0;
    CT.(MATRIZ_CARACTER_FECHAS(i+1,1:9))(j,k)=NaN;
end
end
end
end
%% 6. Estima los aciertos
for i=2:NUMERO_DE_DIAS

ACIERTOS.(MATRIZ_CARACTER_FECHAS(i,1:9))=CT.(MATRIZ_CARACTER_FECHAS(i,1:9))-
PRECIPITACION_CLASIFICADA.(MATRIZ_CARACTER_FECHAS(i,1:9));
end
%% 7. Estima los ERRORES RELATIVOS

for i=2:NUMERO_DE_DIAS

ERRORES_RELATIVOS_NUM.(MATRIZ_CARACTER_FECHAS(i,1:9))=abs(ACIERTOS.(MATRIZ_CARACTER_FECHAS(i,1:9)))/PRECIPITACION_CLASIFICADA.(MATRIZ_CARACTER_FECHAS(i,1:9));

ERRORES_RELATIVOS_NUM.(MATRIZ_CARACTER_FECHAS(i,1:9))=ERRORES_RELATIVOS_NUM.(MATRIZ_CARACTER_FECHAS(i,1:9))*100;
end
```

Algoritmo 4. Estadística de reglas

```
SUMA_DE_MATRIZ_PRONOSTICO=NaN;
SUMA_DE_MATRIZ_CLASIFICADA=NaN;
SUMA_DE_MATRIZ_INTERPOLADA=NaN;
for i=2:NUMERO_DE_DIAS
    MATRIZ_AUXILIAR=CT.(MATRIZ_CARACTER_FECHAS(i,1:9))(2:139,2:99);
    MATRIZ_AUXILIAR=MATRIZ_AUXILIAR(~isnan(MATRIZ_AUXILIAR));
    SUMA_DE_MATRIZ_PRONOSTICO(i-1,1)=sum(sum(MATRIZ_AUXILIAR));

    SUMA_DE_MATRIZ_PRONOSTICO(i-
1,1)=sum(sum(MATRIZ_AUXILIAR(~isnan(MATRIZ_AUXILIAR))));

    SUMA_DE_MATRIZ_CLASIFICADA(i-
1,1)=sum(sum(PRECIPITACION_CLASIFICADA.(MATRIZ_CARACTER_FECHAS(i,1
:9))));
    SUMA_DE_MATRIZ_INTERPOLADA(i-
1,1)=sum(sum(PRECIPITACION_POR_PIXEL.(MATRIZ_CARACTER_FECHAS(i,1:9)
)));
end

SUMA_DE_MATRIZ_PRONOSTICO_MAXIMO=NaN;
SUMA_DE_MATRIZ_PRONOSTICO_MEDIO=NaN;
SUMA_DE_MATRIZ_PRONOSTICO_MINIMO=NaN;

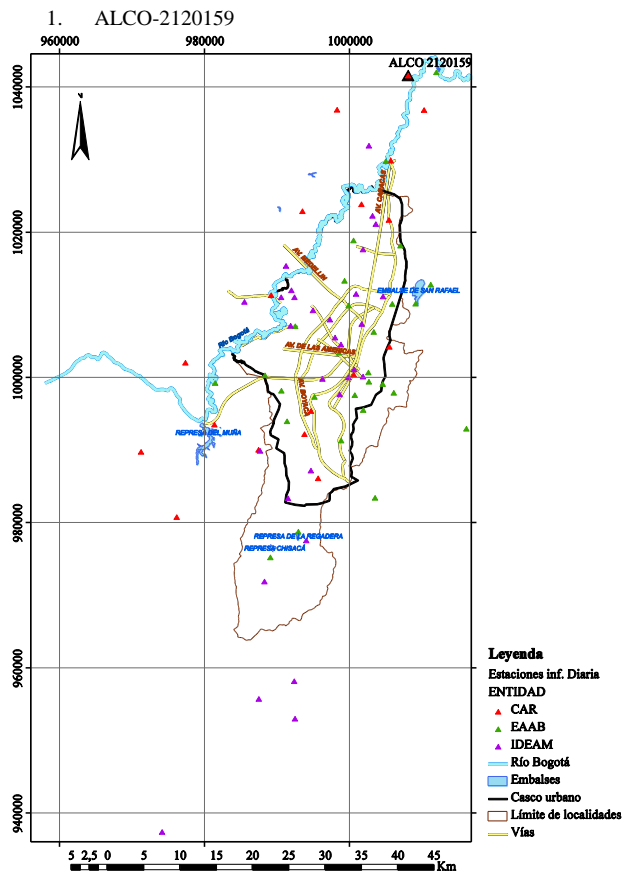
for i=2:NUMERO_DE_DIAS
    MATRIZ_AUXILIAR=CT.(MATRIZ_CARACTER_FECHAS(i,1:9))(2:139,2:99);
    MAXIMO=NaN;
    MEDIO=NaN;
    MINIMO=NaN;
    for j=1:138
        for k=1:98
            if MATRIZ_AUXILIAR(j,k) >= 0;
                MAXIMO(j,k)=RANGO(1,MATRIZ_AUXILIAR(j,k)+1);

MEDIO(j,k)=(RANGO(1,MATRIZ_AUXILIAR(j,k)+1)+RANGO(1,MATRIZ_AUXILIA
R(j,k)))/2;
                MINIMO(j,k)=RANGO(1,MATRIZ_AUXILIAR(j,k));
            end
        end
    end
    SUMA_DE_MATRIZ_PRONOSTICO_MAXIMO(i-
1,1)=sum(sum(MAXIMO(~isnan(MAXIMO))));
    SUMA_DE_MATRIZ_PRONOSTICO_MEDIO(i-
1,1)=sum(sum(MEDIO(~isnan(MEDIO))));
```

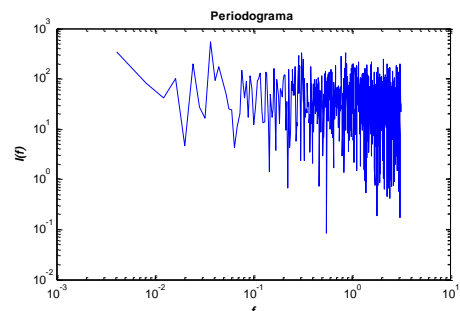
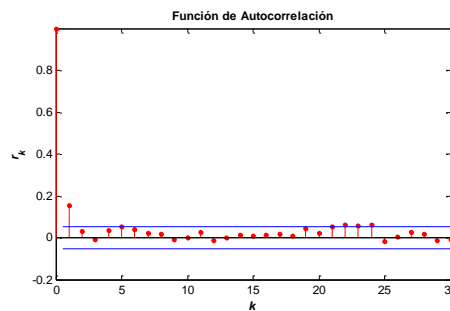
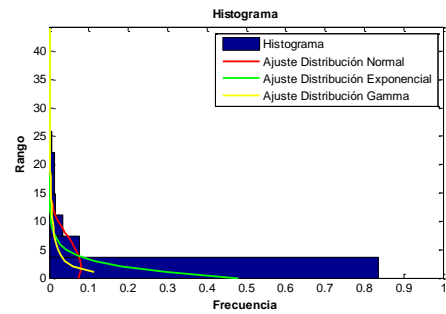
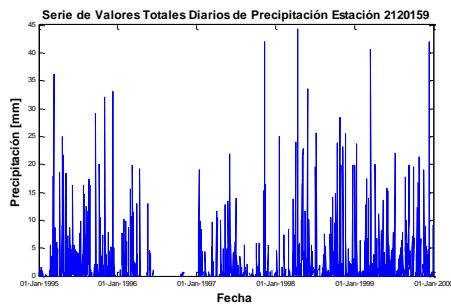
```
SUMA_DE_MATRIZ_PRONOSTICO_MINIMO(i-  
1,1)=sum(sum(MINIMO(~isnan(MINIMO))));  
end
```

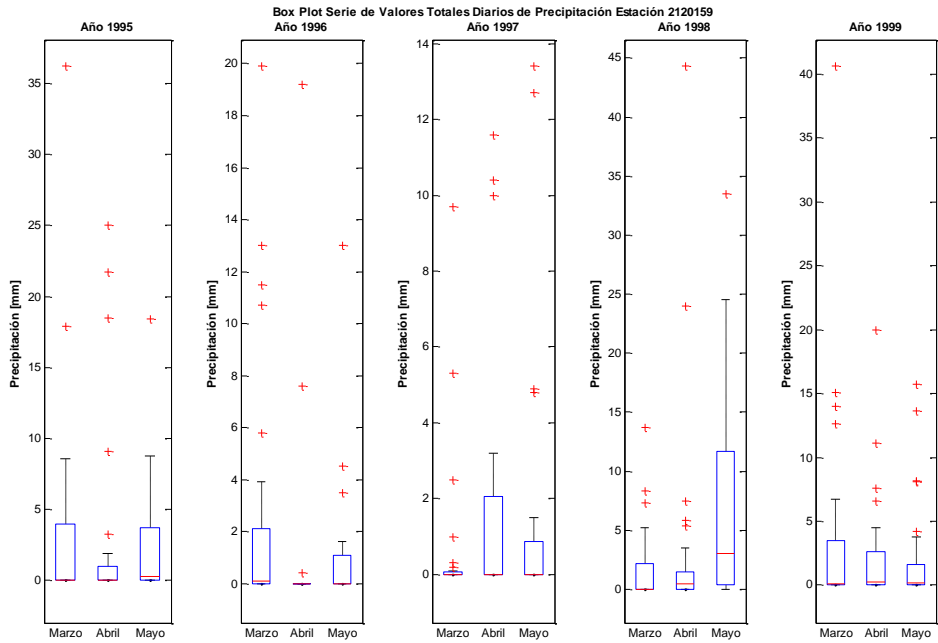
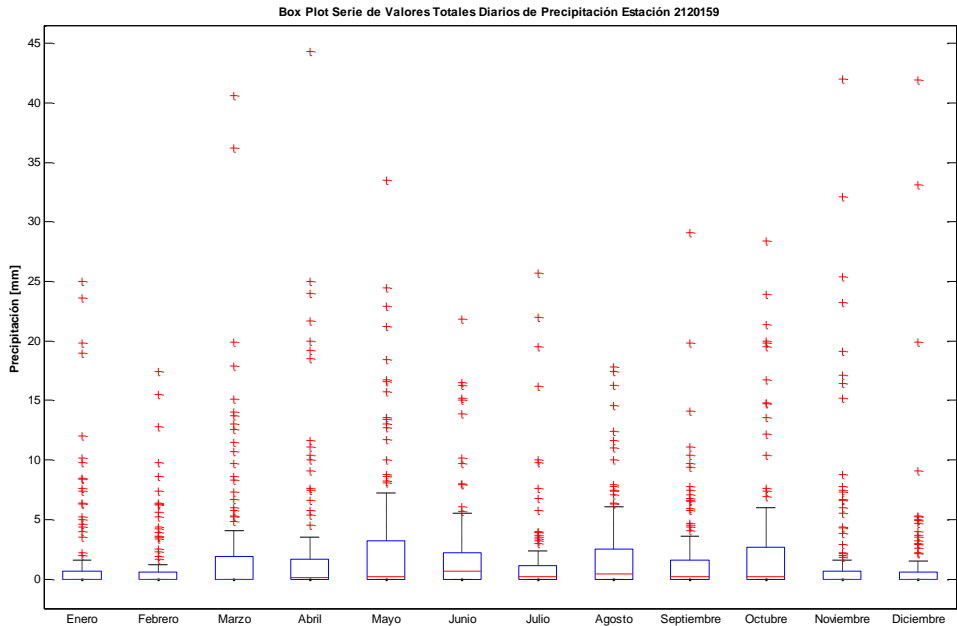
APENDICES

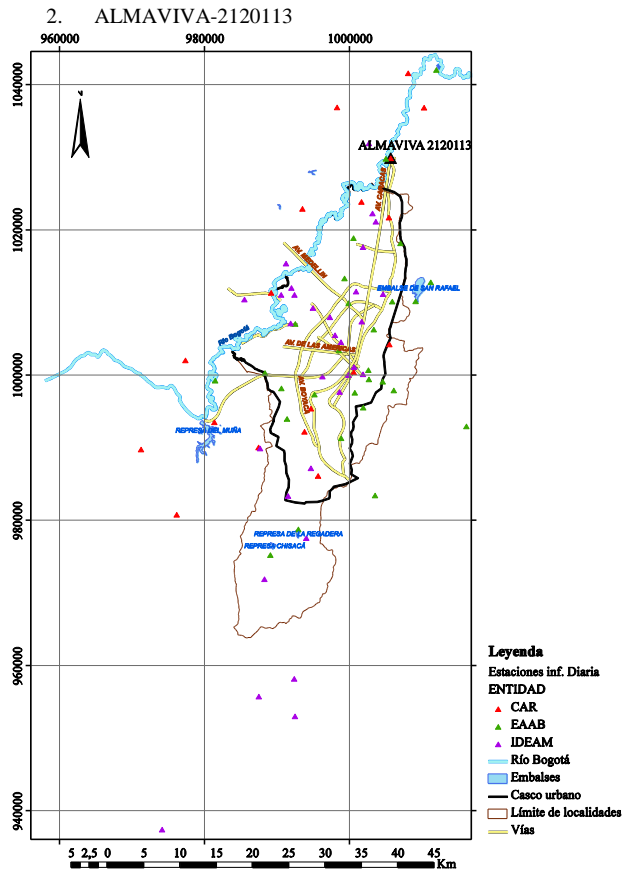
***APÉNDICE A. CARACTERIZACIÓN DE LA PRECIPITACIÓN
DIARIA POR ESTACIÓN***



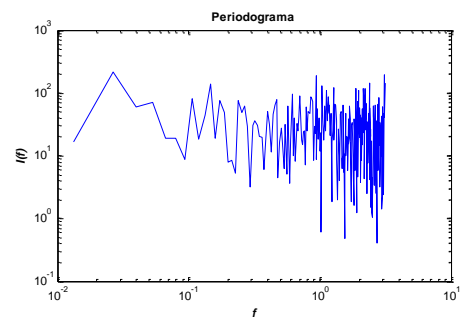
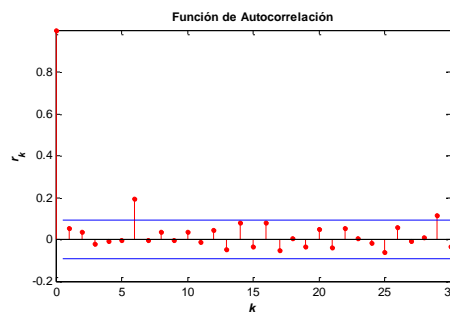
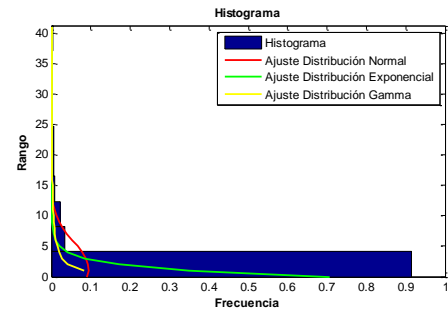
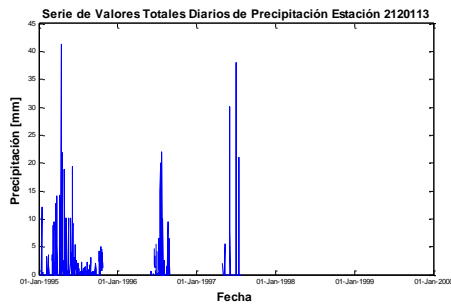
Este	1008000
Norte	1041700
Cota	2590
Número de Datos	1551
Número de Datos Sin Registro	275
Media	2.1
Mediana	0.1
Máximo	44.3
Desviación Estándar	5.0
Varianza	24.6
Coefficiente de Variación	8.1E-04
Coefficiente de Asimetría	4.0
Coefficiente de Curtosis	22.8

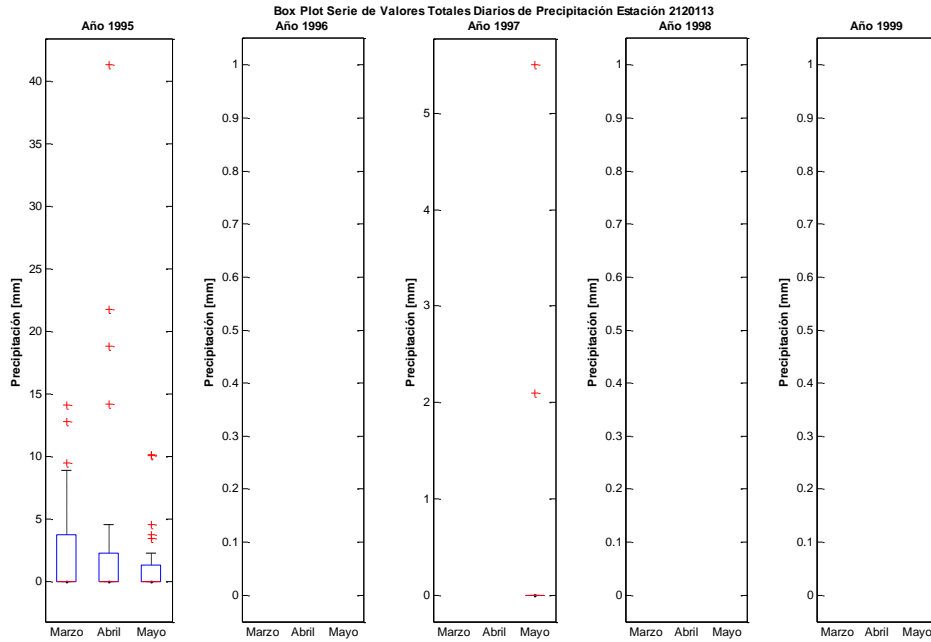
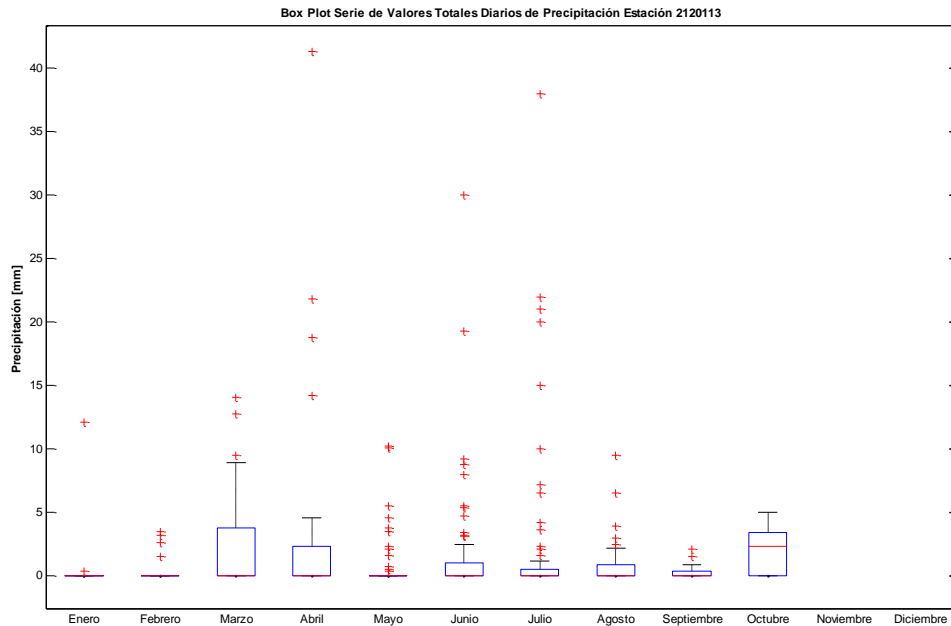


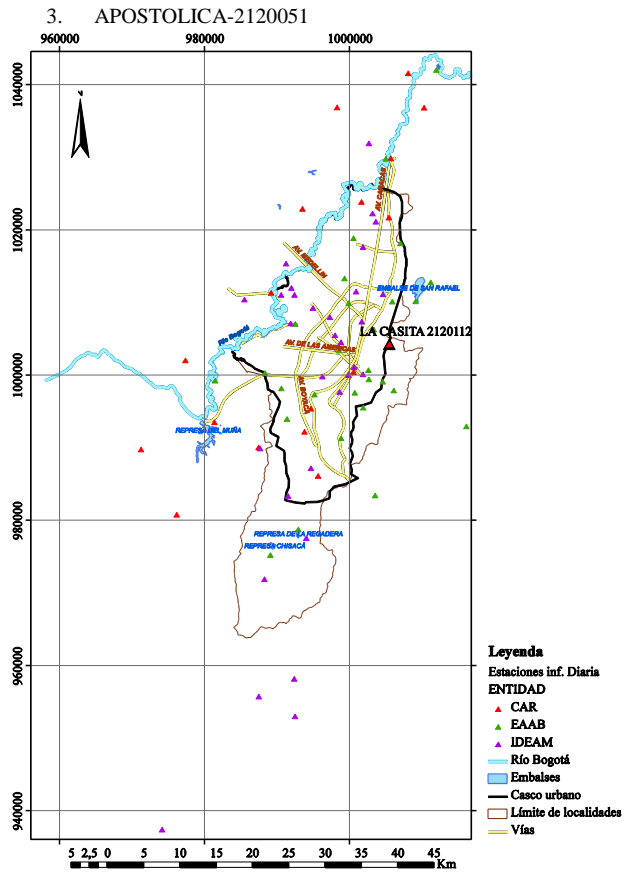




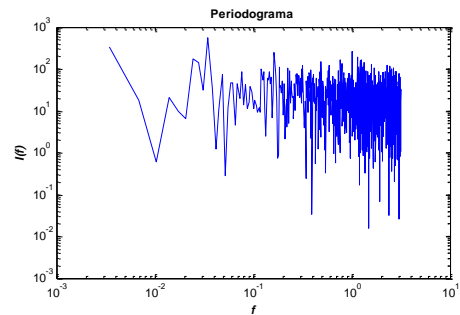
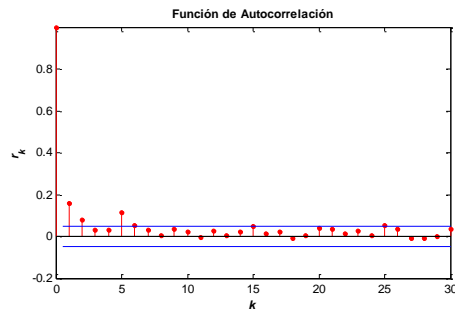
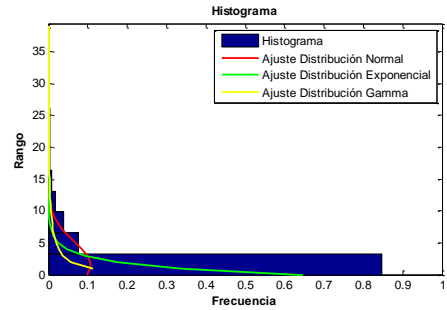
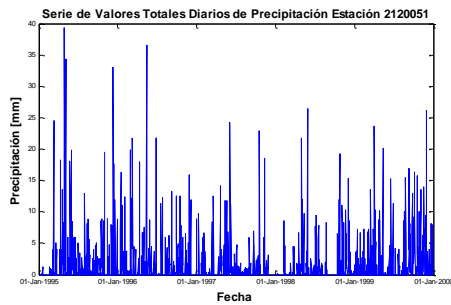
Este	1005600
Norte	1029965
Cota	2595
Número de Datos	471
Número de Datos Sin Registro	1355
Media	1.4
Mediana	0.0
Máximo	41.3
Desviación Estándar	4.2
Varianza	17.9
Coefficiente de Variación	5.5E-04
Coefficiente de Asimetría	5.4
Coefficiente de Curtosis	39.8

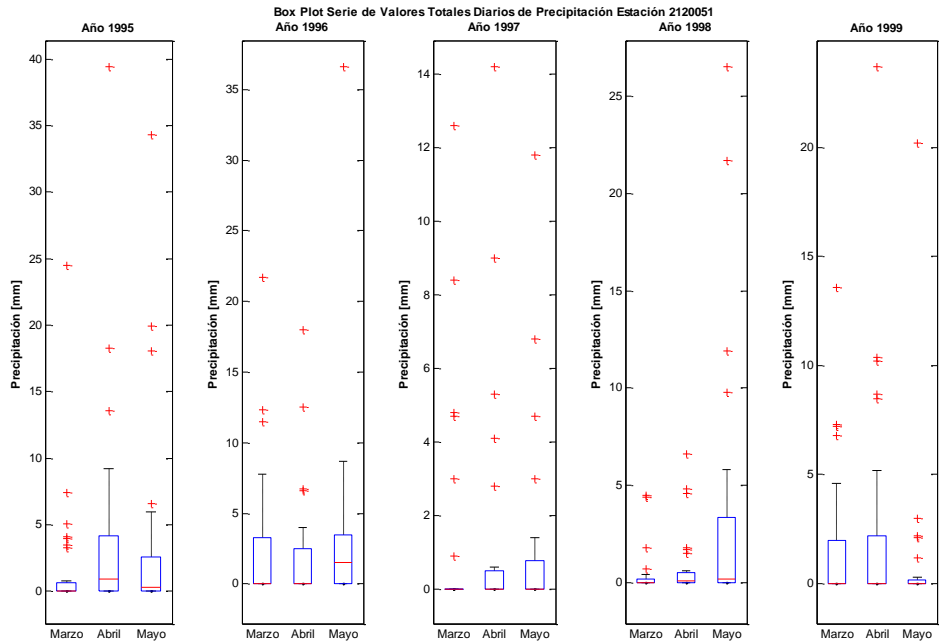
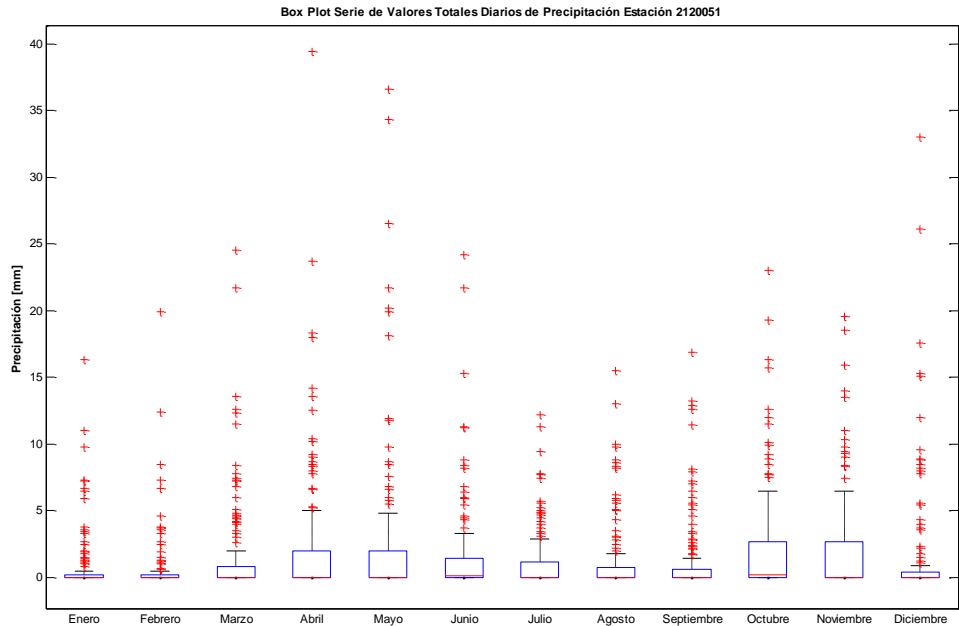




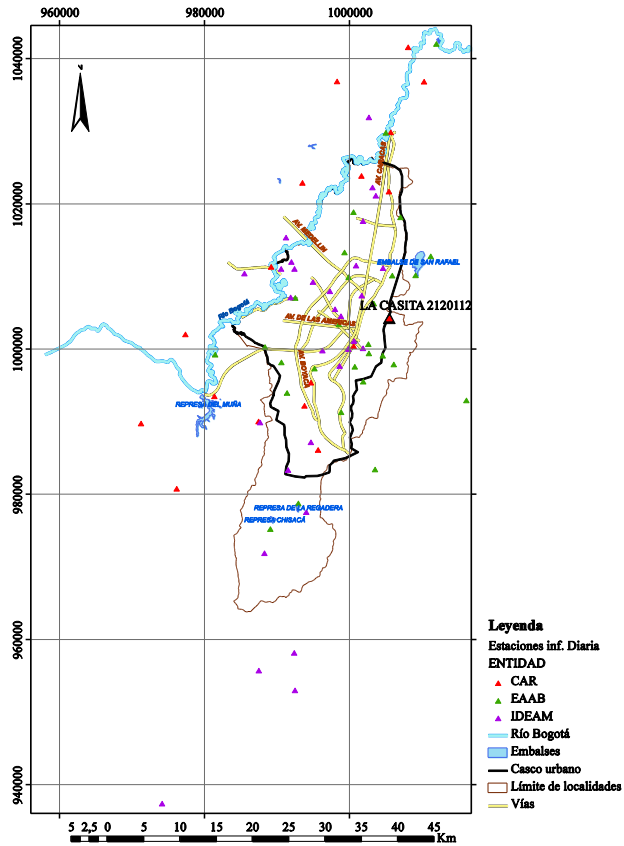


Este	971174
Norte	989870
Cota	2618
Número de Datos	1818
Número de Datos Sin Registro	8
Media	1.5
Mediana	0.0
Máximo	39.4
Desviación Estándar	3.7
Varianza	14.0
Coefficiente de Variación	5.9E-04
Coefficiente de Asimetría	4.3
Coefficiente de Curtosis	27.6

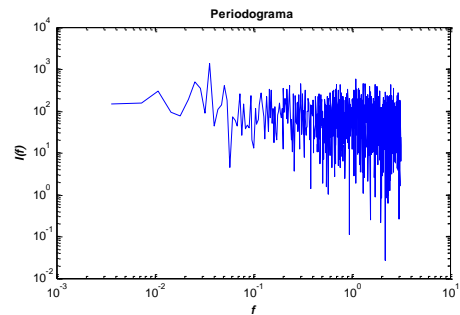
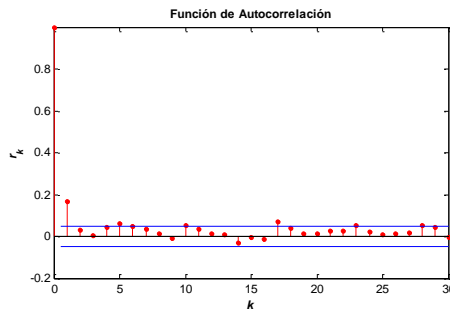
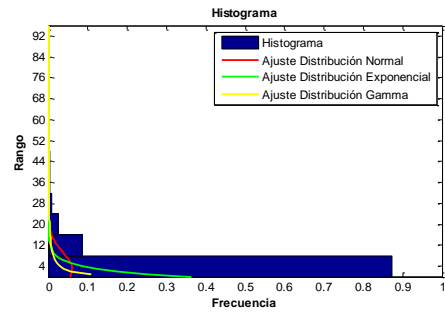
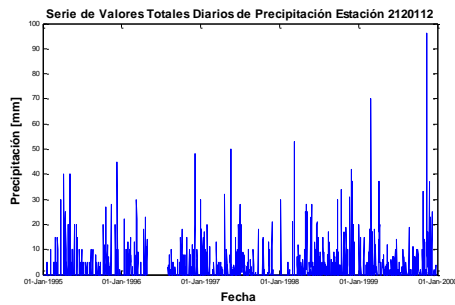


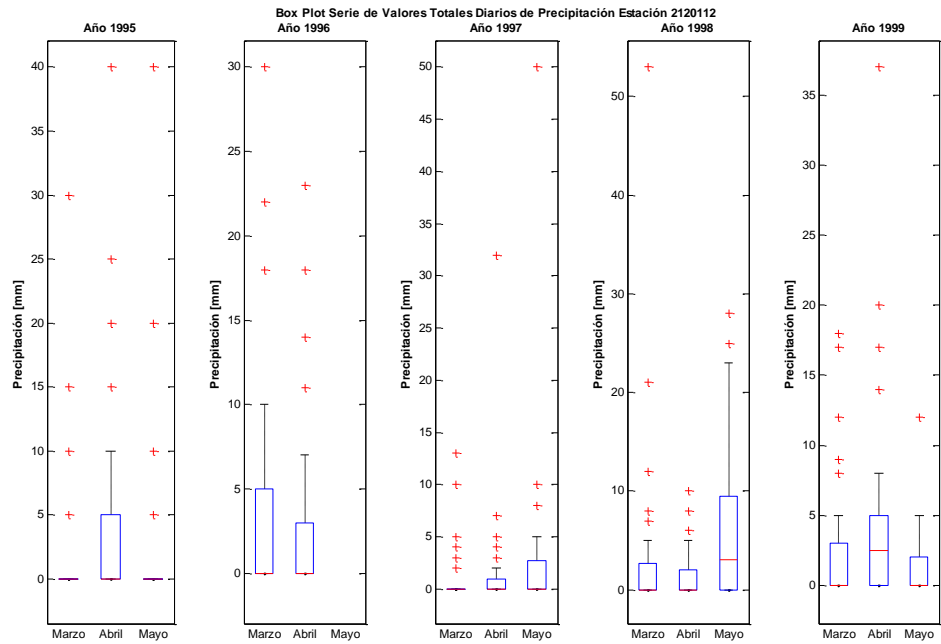
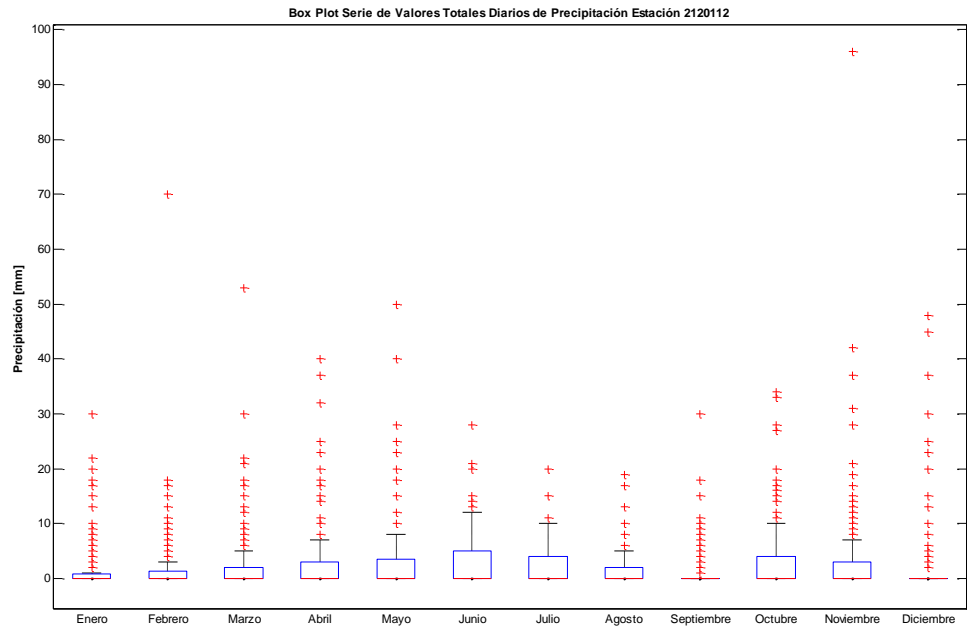


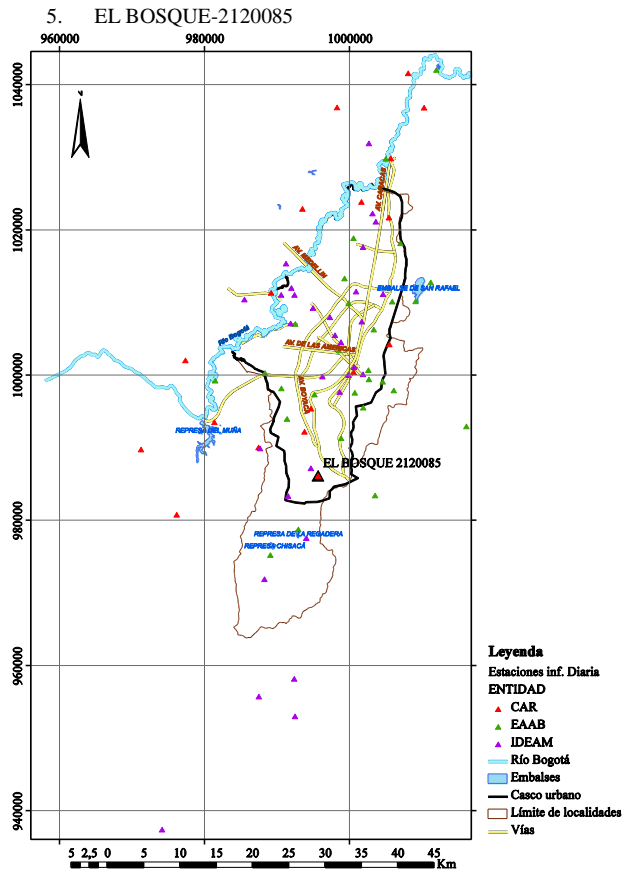
4. LA CASITA-2120112



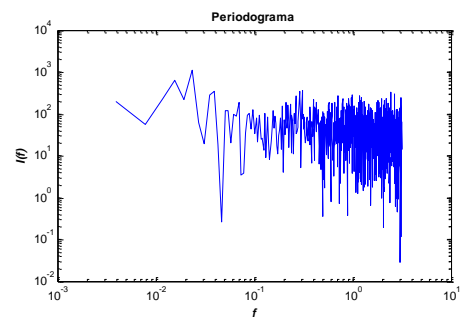
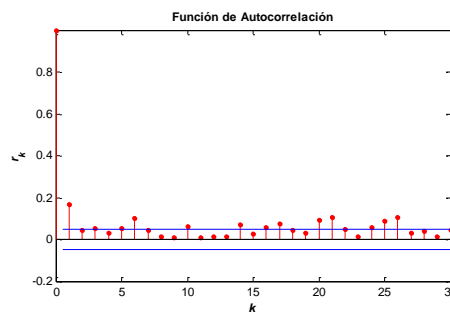
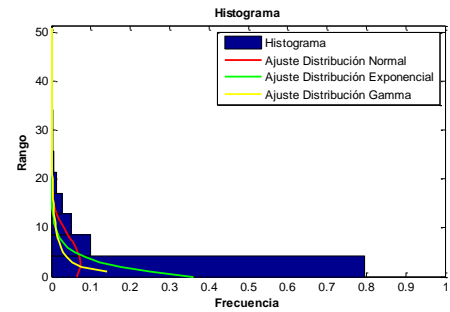
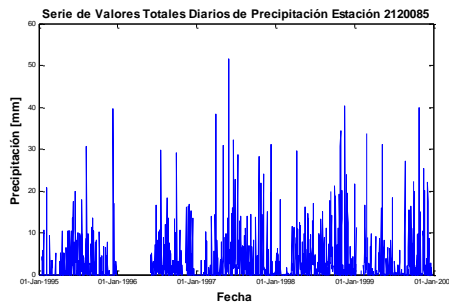
Este	1005400
Norte	1004320
Cota	3045
Número de Datos	1734
Número de Datos Sin Registro	92
Media	2.8
Mediana	0.0
Máximo	96.0
Desviación Estándar	6.6
Varianza	43.8
Coefficiente de Variación	9.1E-04
Coefficiente de Asimetría	4.9
Coefficiente de Curtosis	42.3

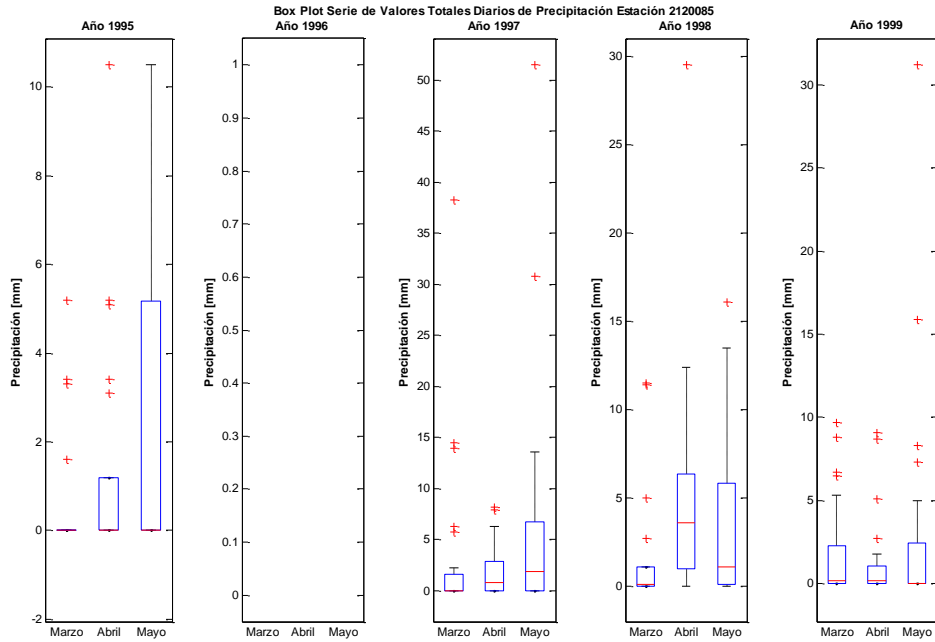
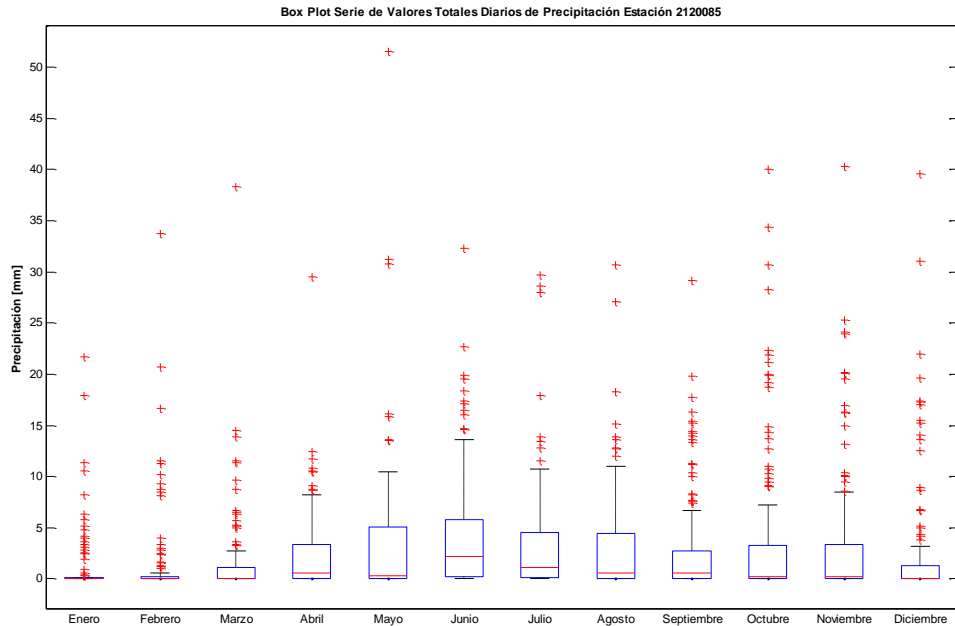


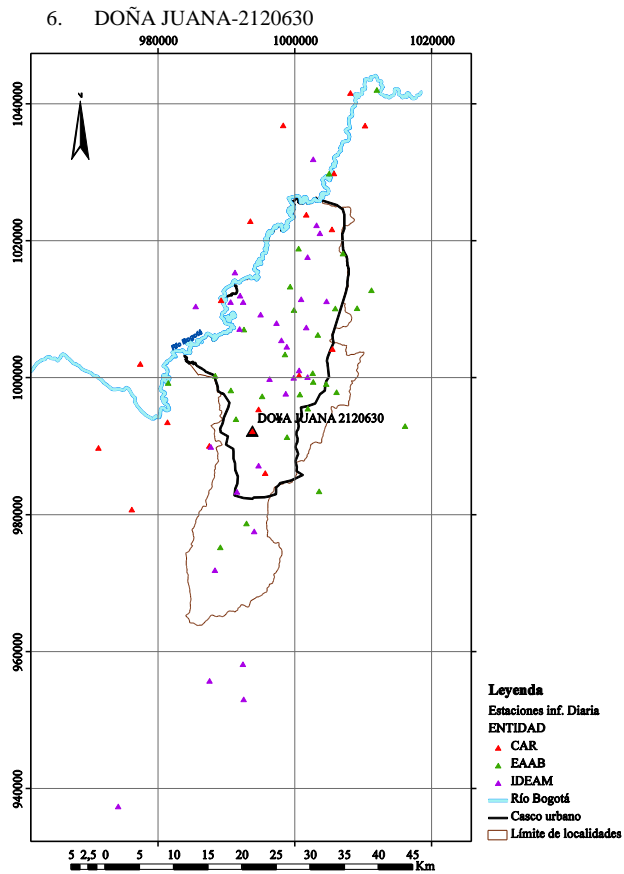




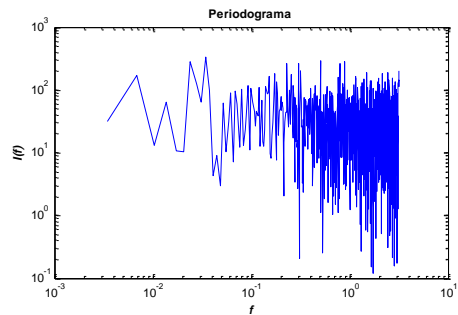
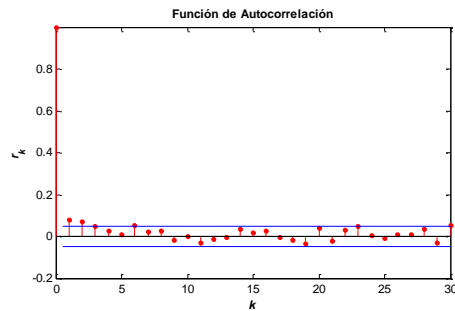
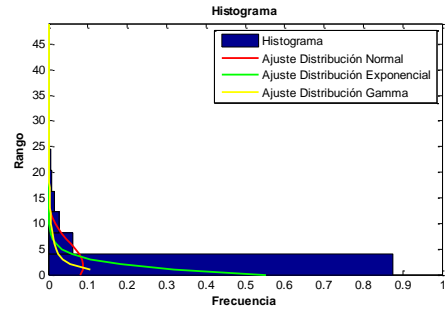
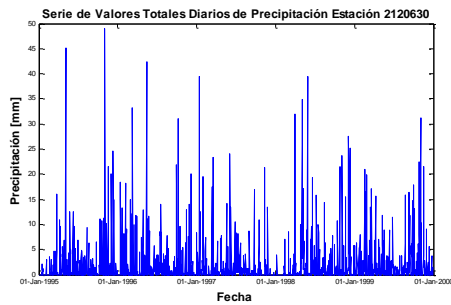
Este	995560
Norte	986200
Cota	2880
Número de Datos	1618
Número de Datos Sin Registro	208
Media	2.8
Mediana	0.2
Máximo	51.5
Desviación Estándar	5.4
Varianza	29.7
Coefficiente de Variación	9.6E-04
Coefficiente de Asimetría	3.4
Coefficiente de Curtosis	18.3

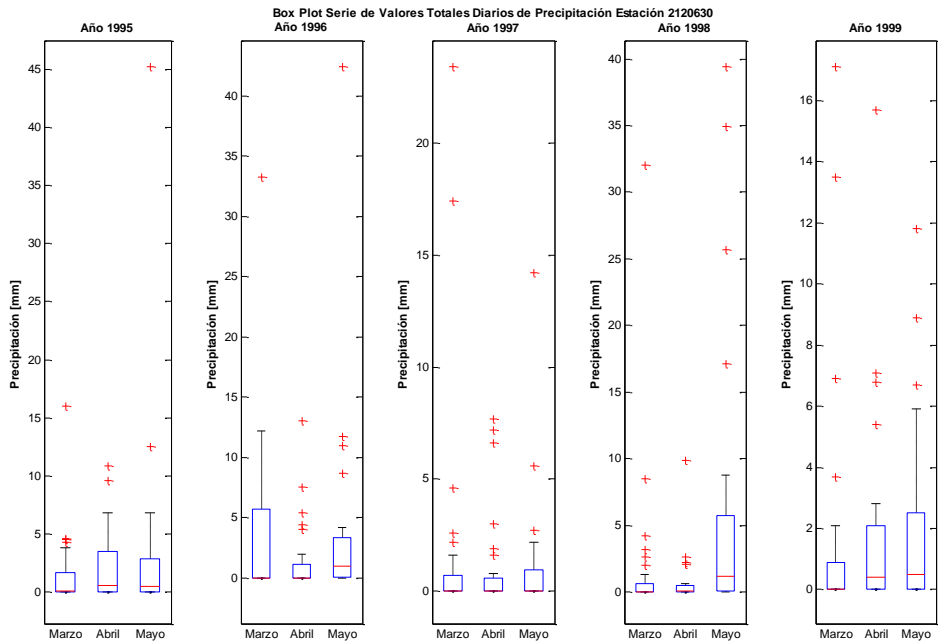
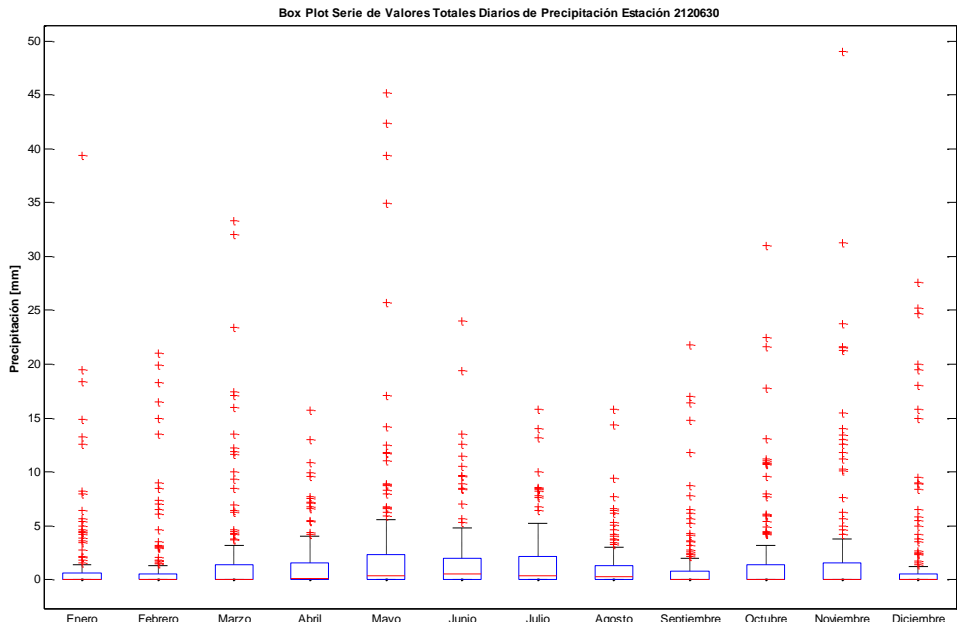


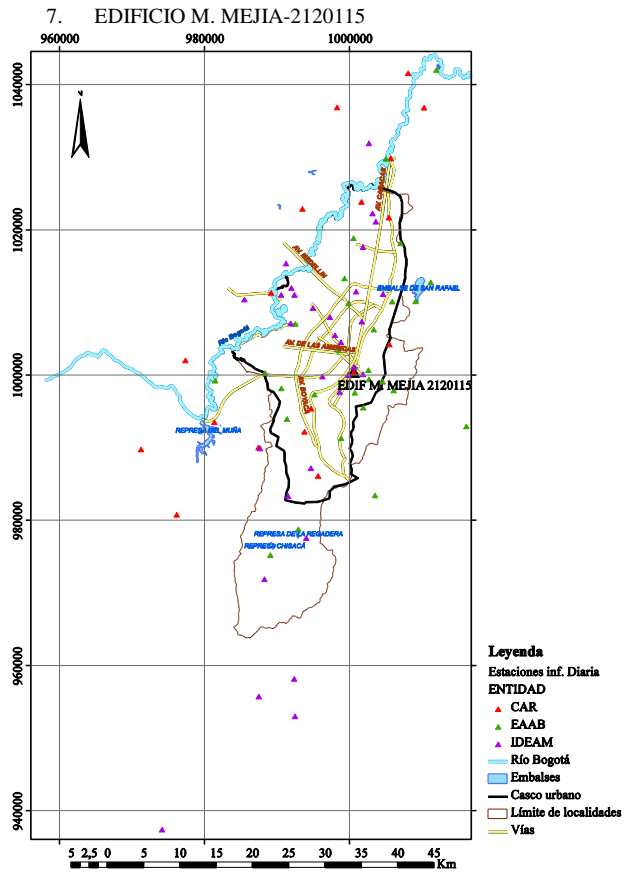




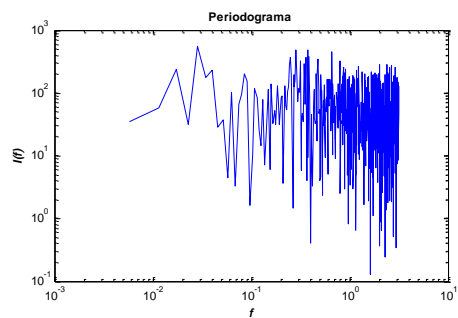
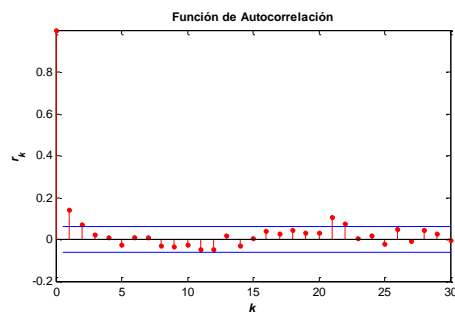
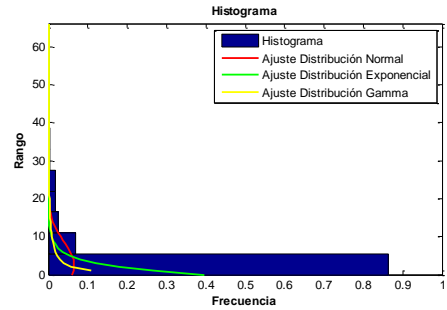
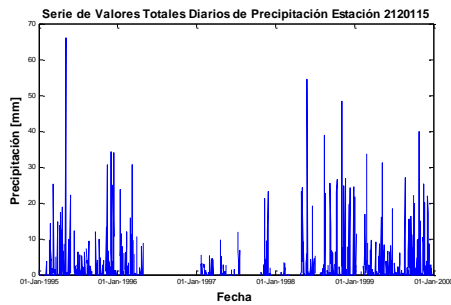
Este	993700
Norte	992300
Cota	2700
Número de Datos	1826
Número de Datos Sin Registro	0
Media	1.8
Mediana	0.0
Máximo	49.0
Desviación Estándar	4.5
Varianza	20.4
Coefficiente de Variación	6.7E-04
Coefficiente de Asimetría	4.7
Coefficiente de Curtosis	32.7

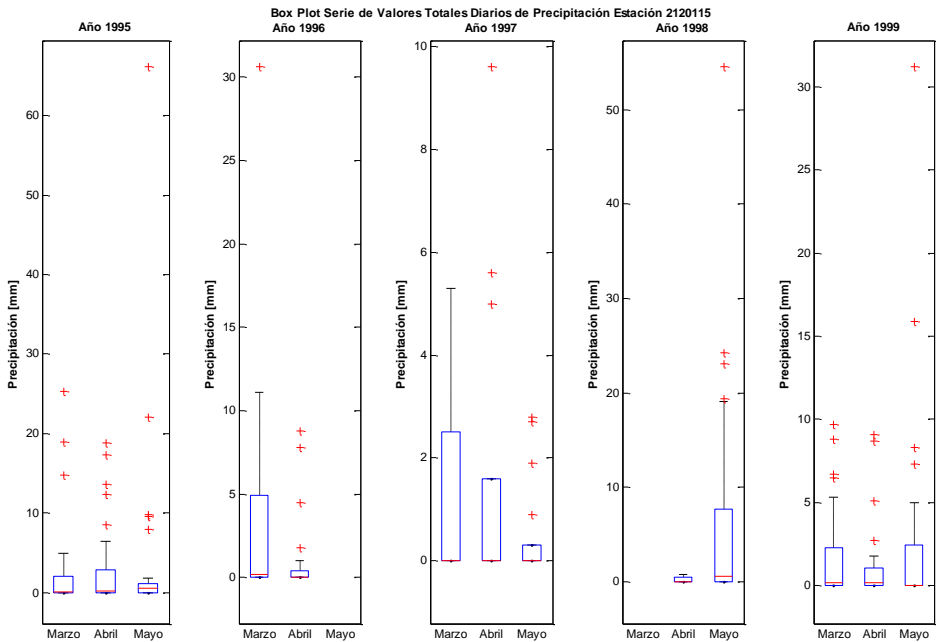
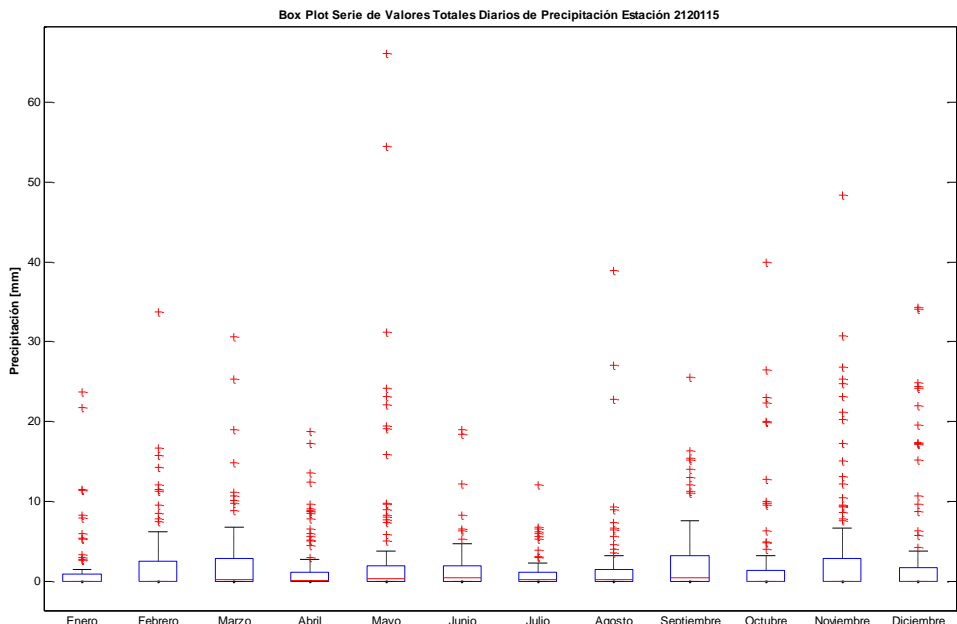


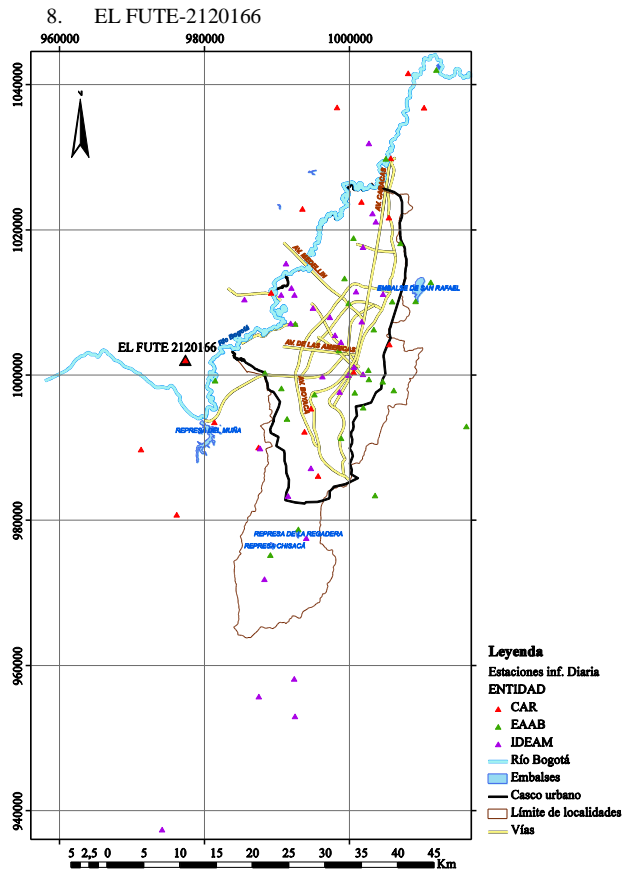




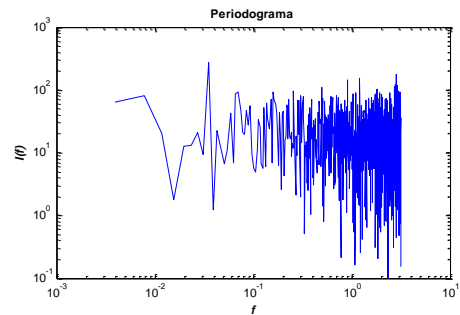
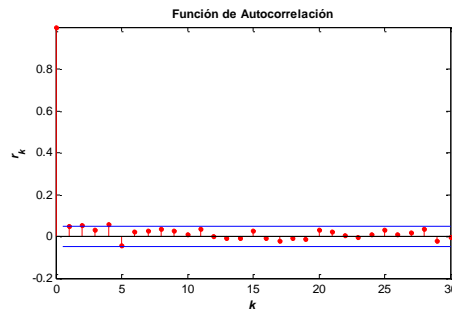
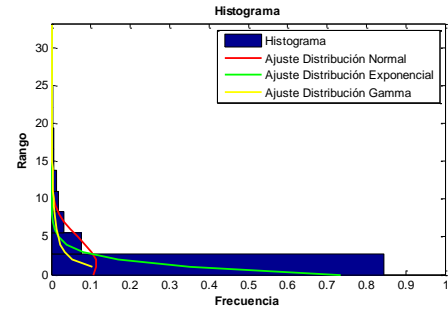
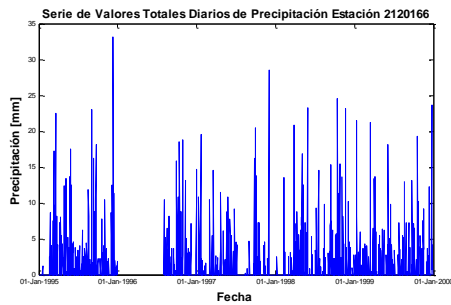
Este	1000500
Norte	1000500
Cota	2580
Número de Datos	1097
Número de Datos Sin Registro	729
Media	2.5
Mediana	0.1
Máximo	66.1
Desviación Estándar	6.1
Varianza	37.2
Coefficiente de Variación	9.8E-04
Coefficiente de Asimetría	4.3
Coefficiente de Curtosis	28.2

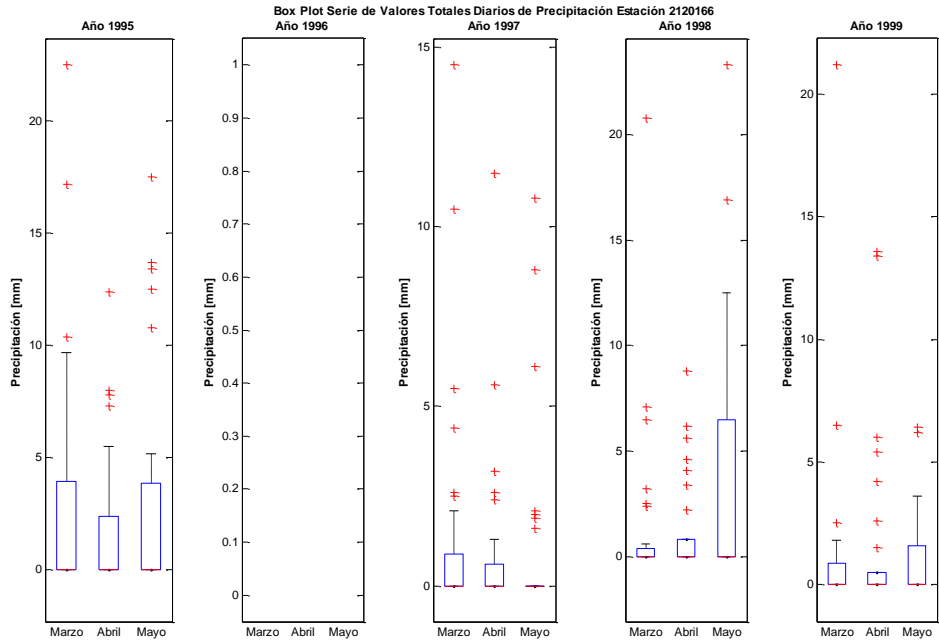
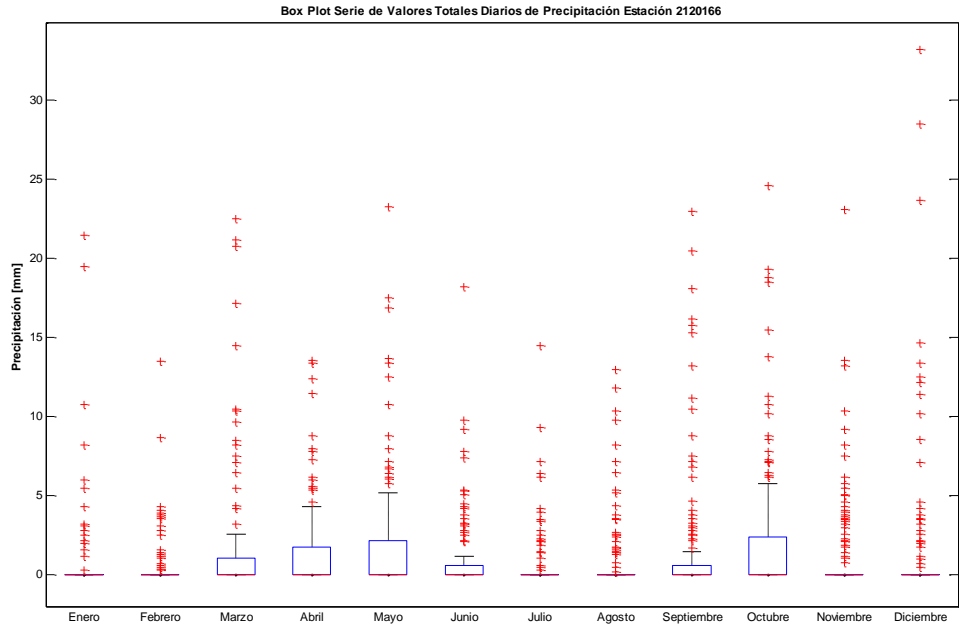


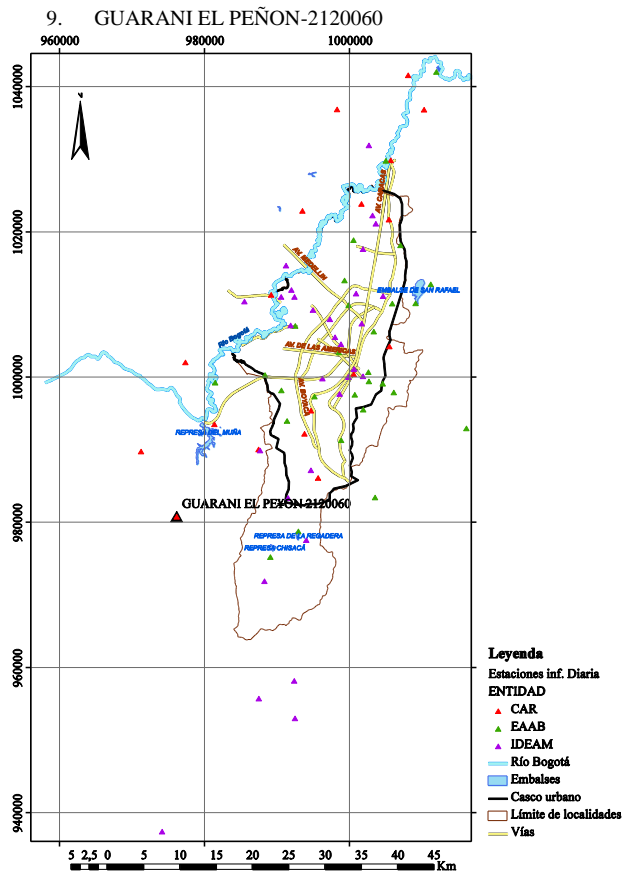




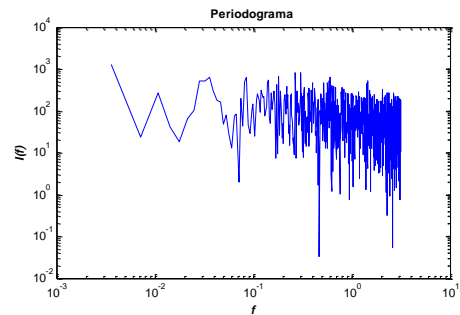
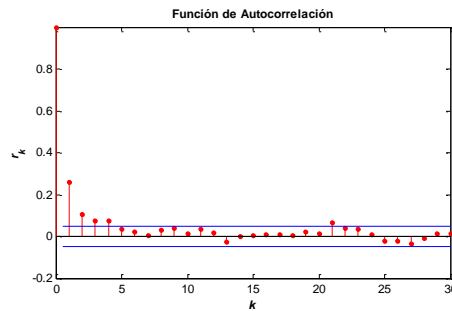
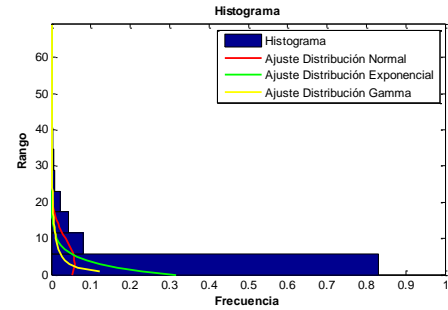
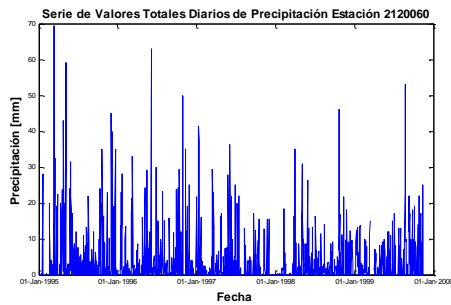
Este	977280
Norte	1002150
Cota	2607
Número de Datos	1610
Número de Datos Sin Registro	216
Media	1.4
Mediana	0.0
Máximo	33.2
Desviación Estándar	3.5
Varianza	12.2
Coefficiente de Variación	5.2E-04
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	21.6

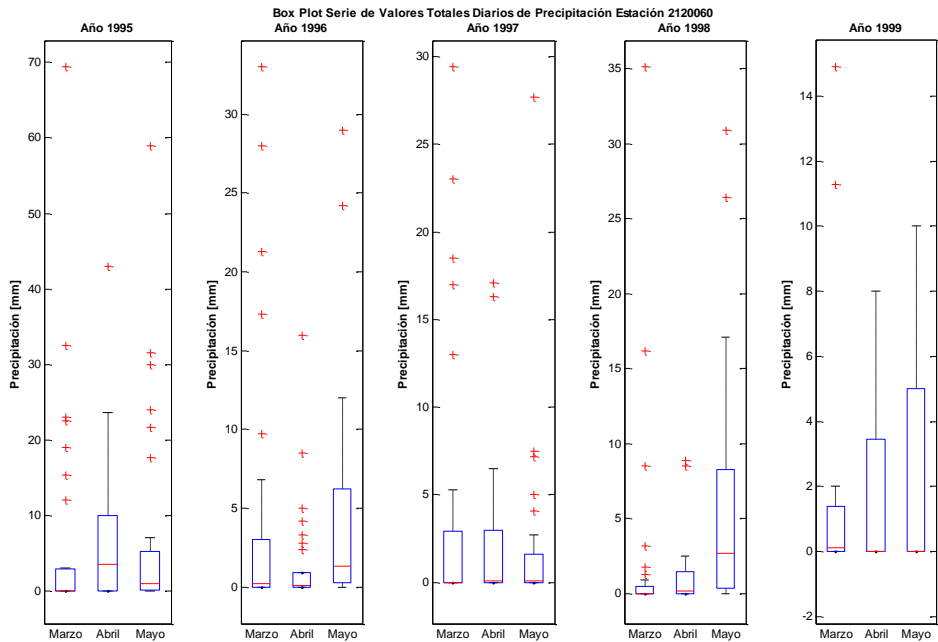
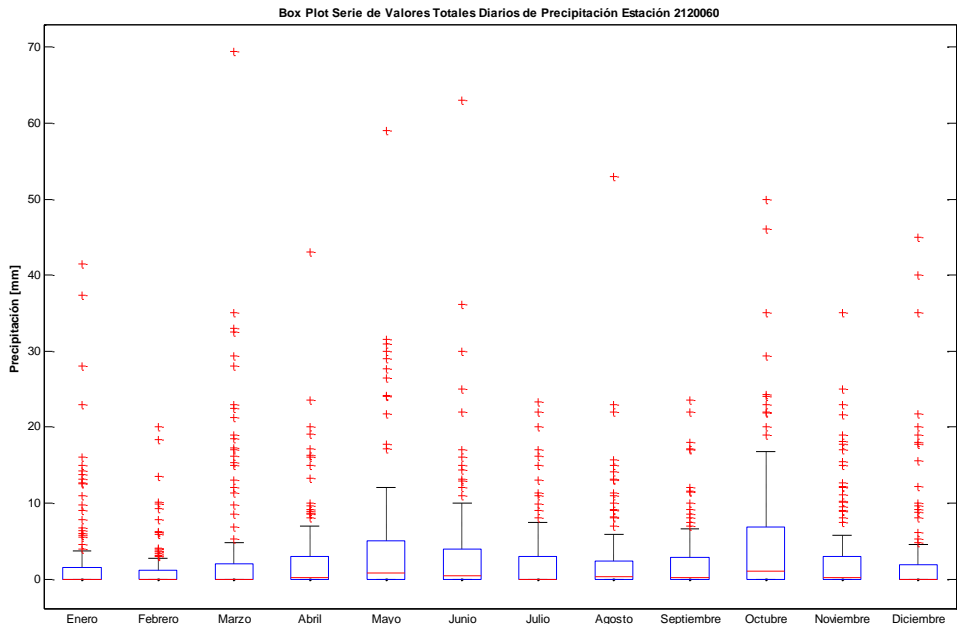


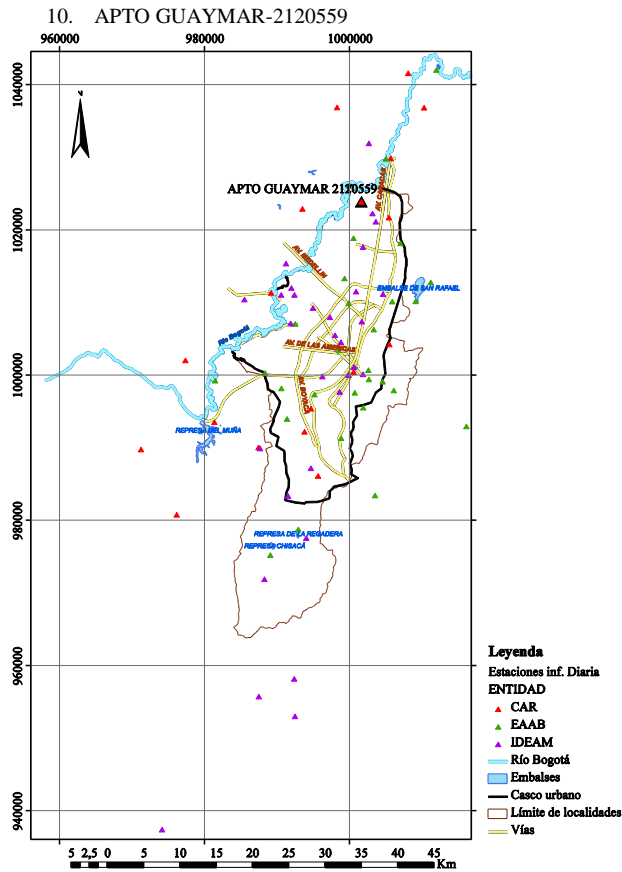




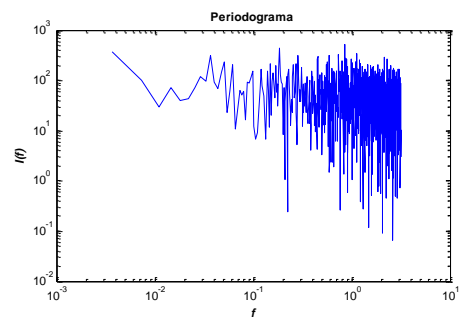
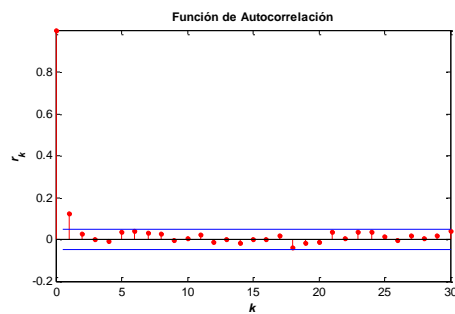
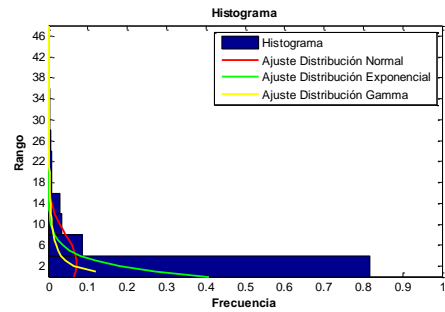
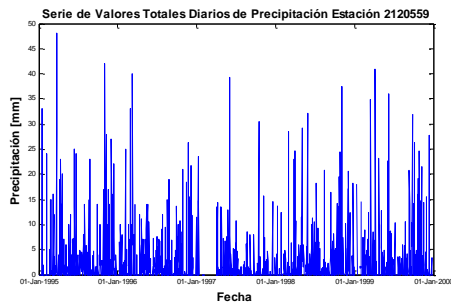
Este	976070
Norte	980870
Cota	2800
Número de Datos	1761
Número de Datos Sin Registro	65
Media	3.2
Mediana	0.2
Máximo	69.4
Desviación Estándar	6.8
Varianza	46.1
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	3.8
Coefficiente de Curtosis	23.3

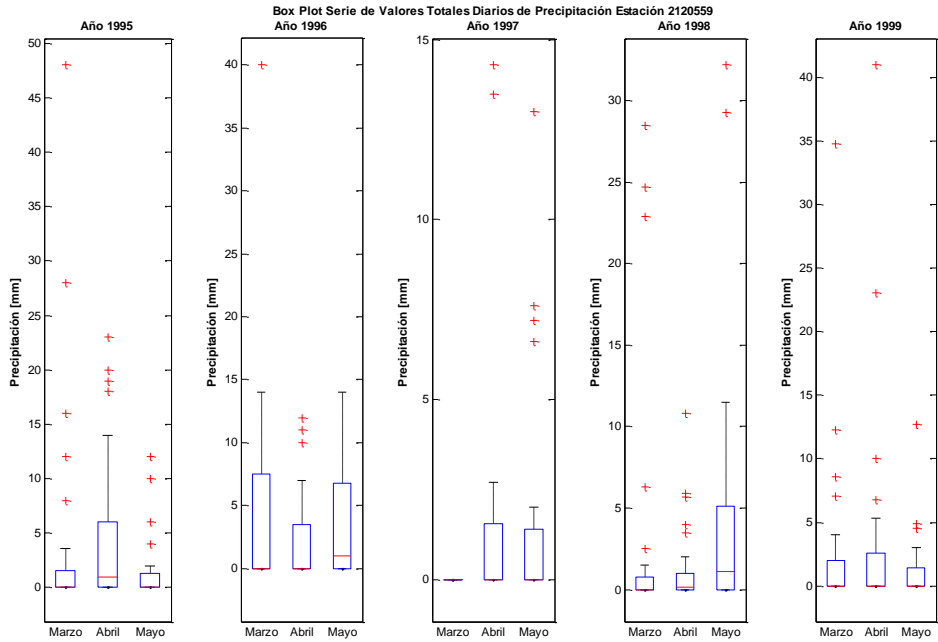
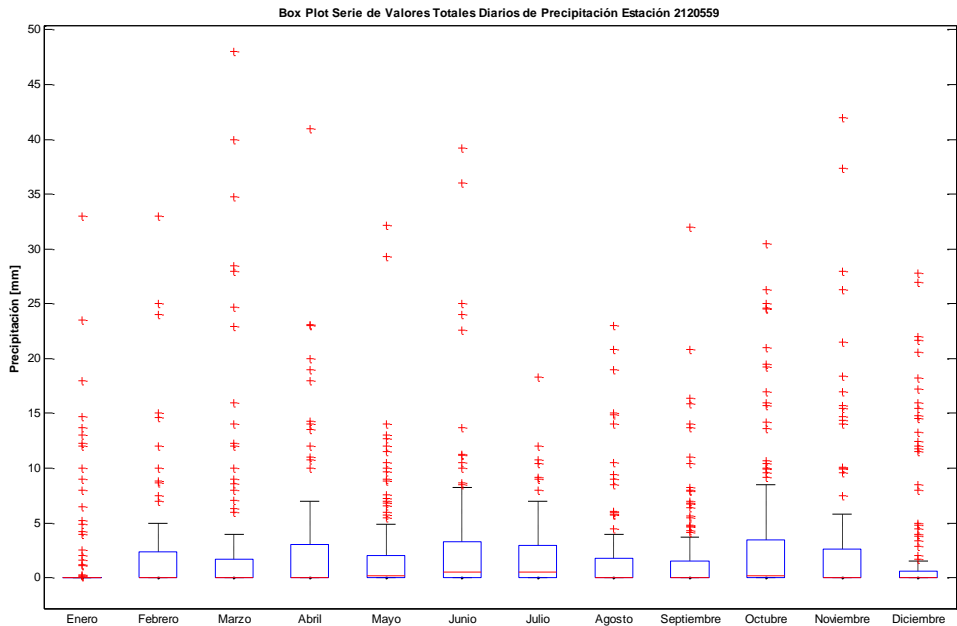




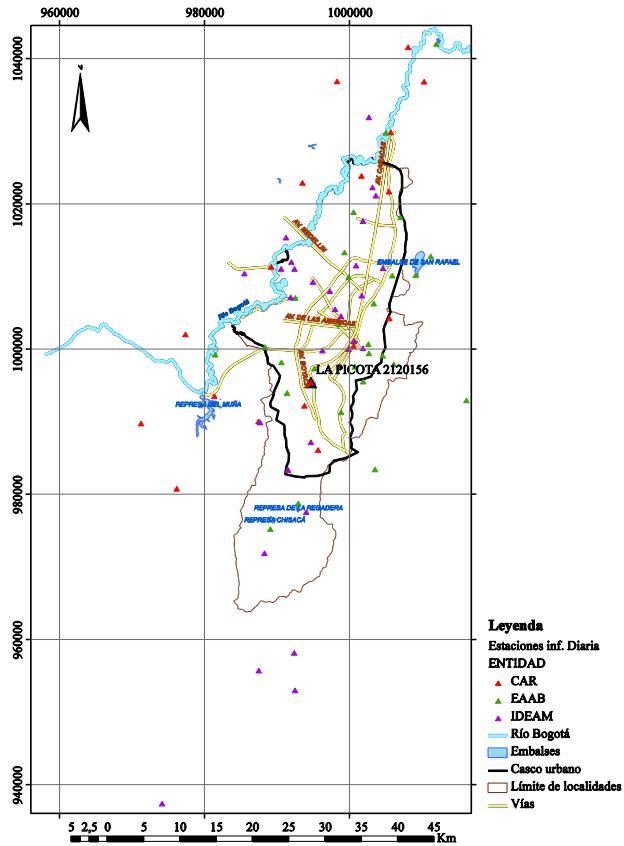


Este	1001550
Norte	1023950
Cota	2560
Número de Datos	1723
Número de Datos Sin Registro	103
Media	2.5
Mediana	0.0
Máximo	48.0
Desviación Estándar	5.5
Varianza	30.7
Coefficiente de Variación	9.6E-04
Coefficiente de Asimetría	3.7
Coefficiente de Curtosis	19.8

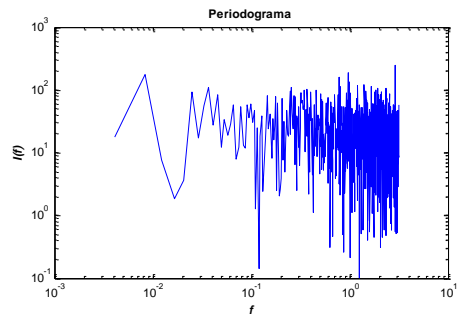
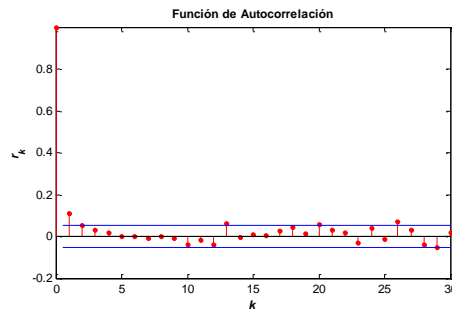
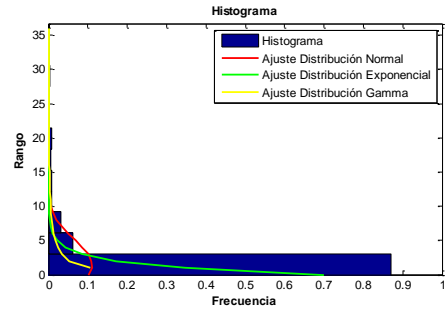
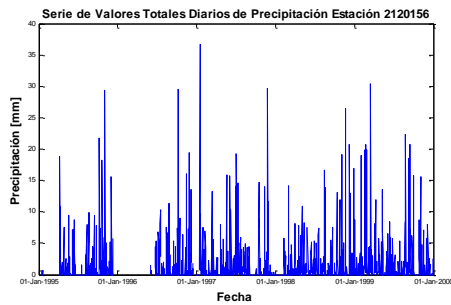


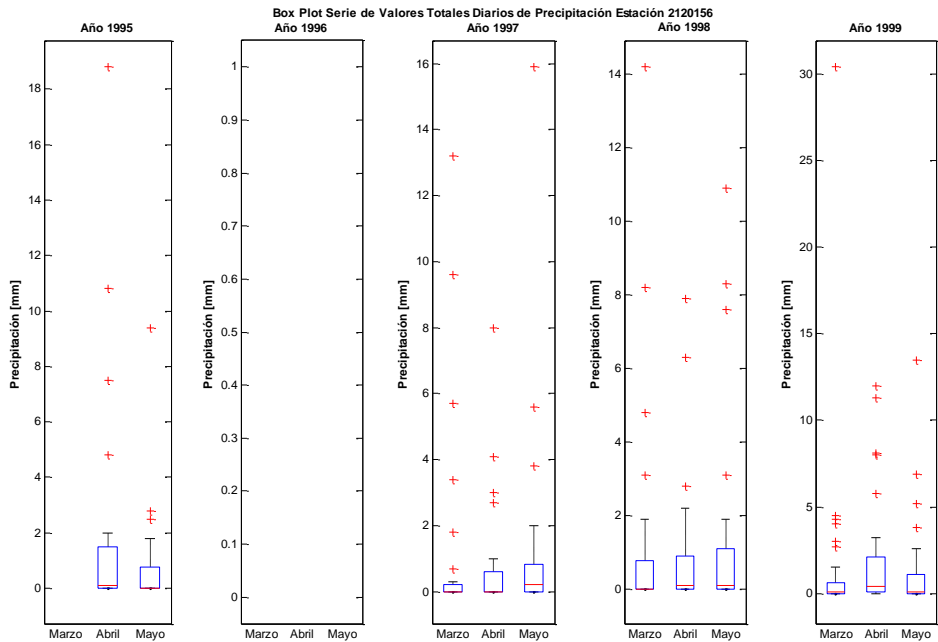
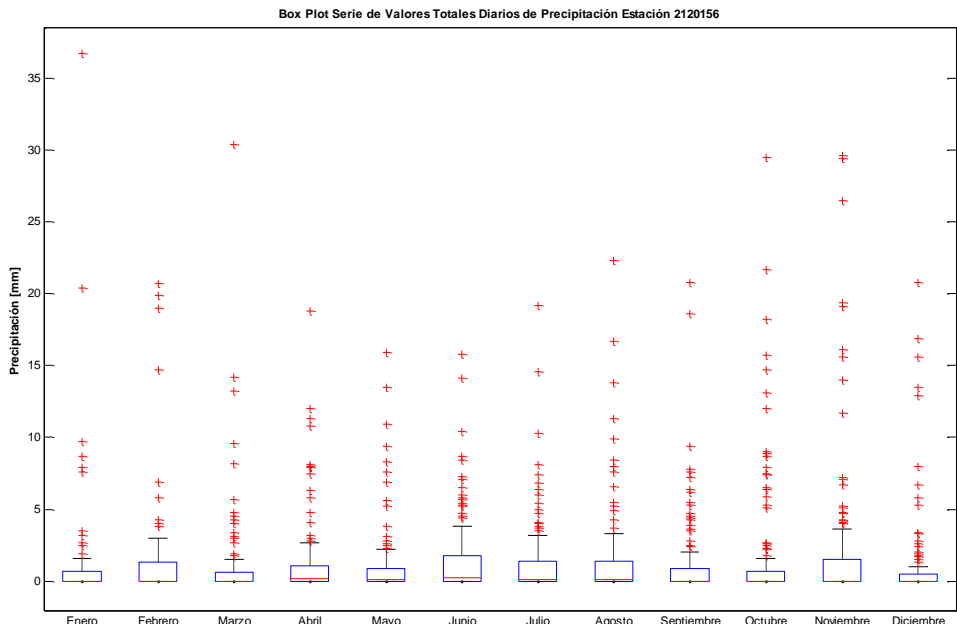


11. LA PICOTA-2120156

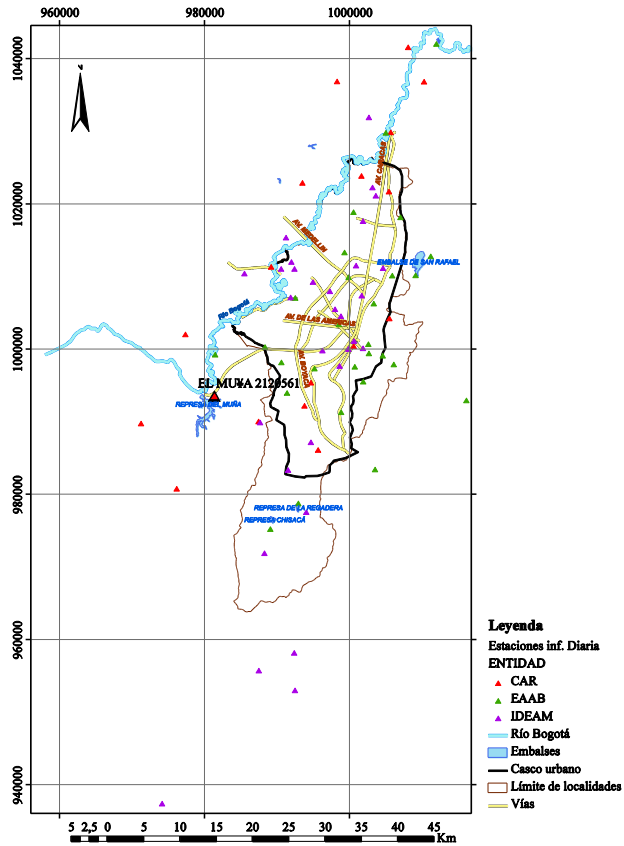


Este	994600
Norte	995500
Cota	2580
Número de Datos	1525
Número de Datos Sin Registro	301
Media	1.4
Mediana	0.0
Máximo	36.7
Desviación Estándar	3.6
Varianza	13.0
Coefficiente de Variación	5.5E-04
Coefficiente de Asimetría	4.4
Coefficiente de Curtosis	27.8

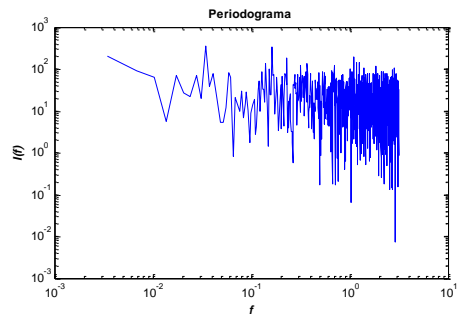
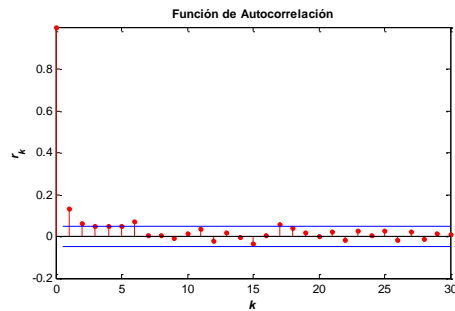
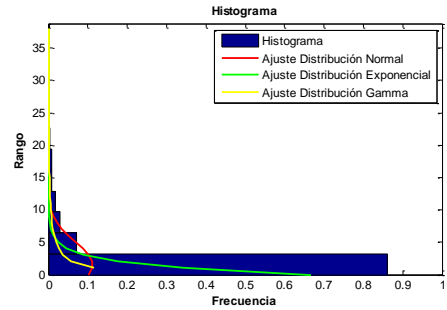
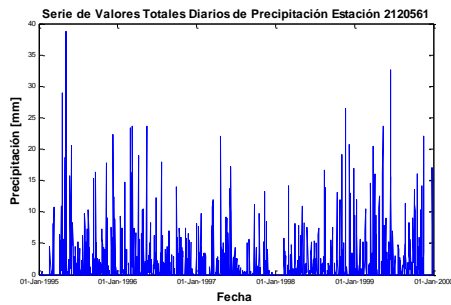


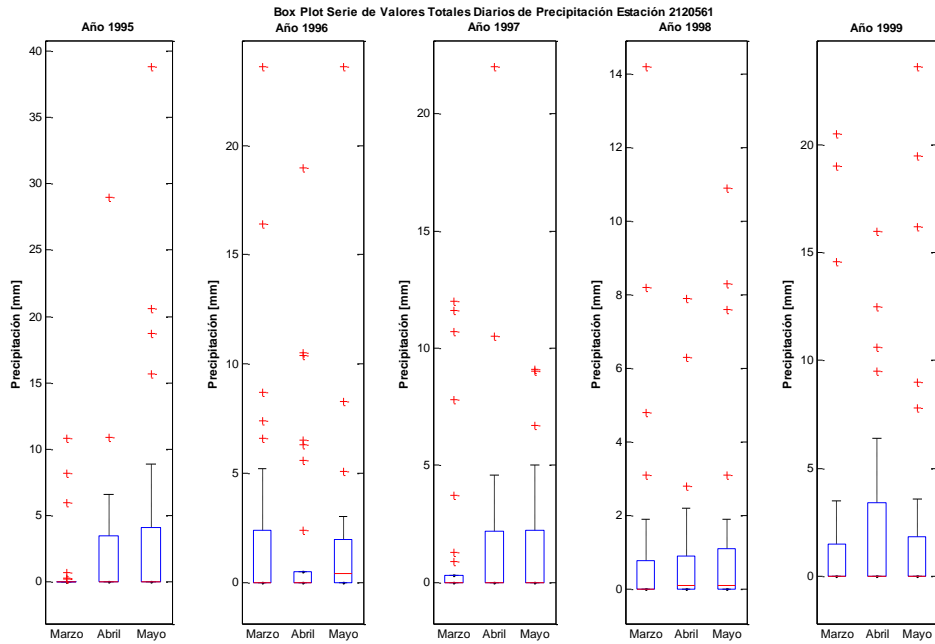
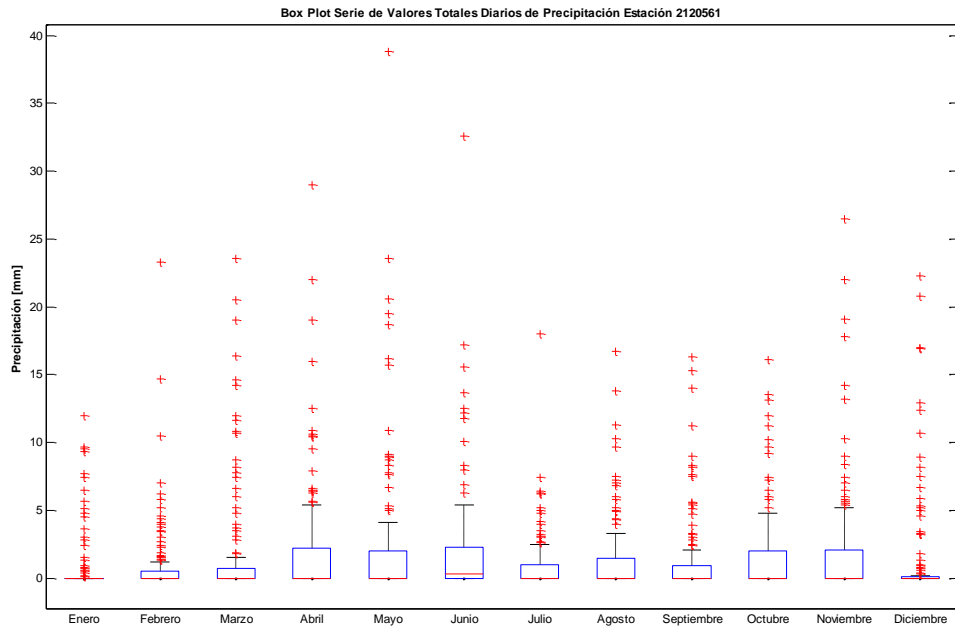


12. EL MUÑA-2120561

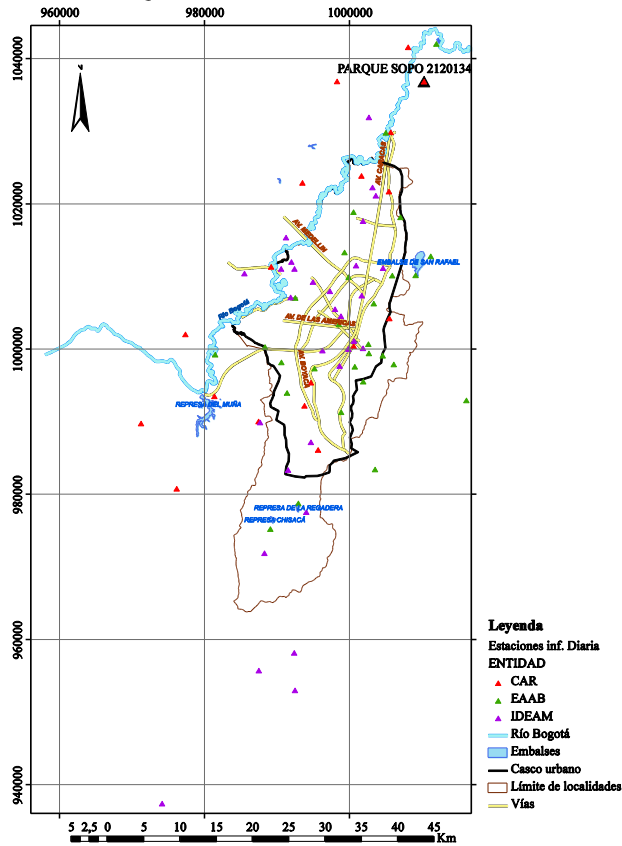


Este	981280
Norte	993600
Cota	2565
Número de Datos	1826
Número de Datos Sin Registro	0
Media	1.5
Mediana	0.0
Máximo	38.8
Desviación Estándar	3.6
Varianza	12.9
Coefficiente de Variación	5.9E-04
Coefficiente de Asimetría	4.1
Coefficiente de Curtosis	24.8

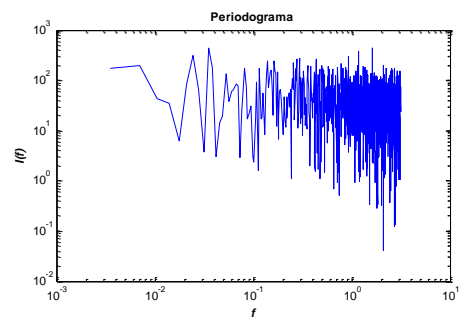
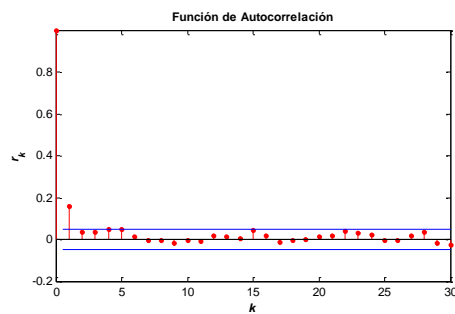
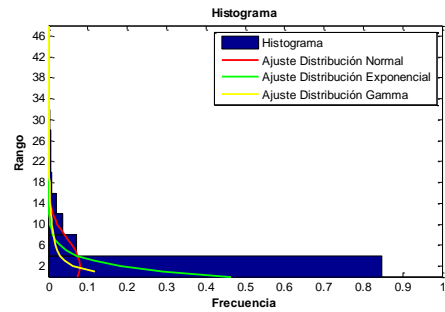
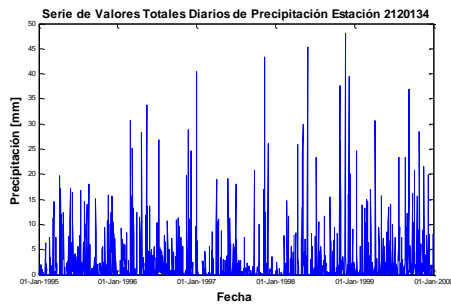


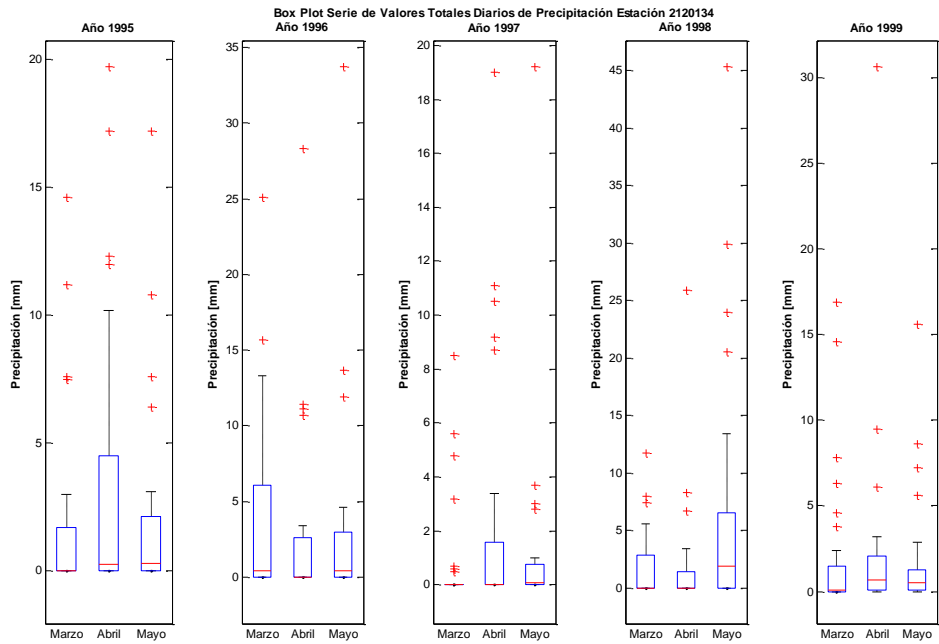
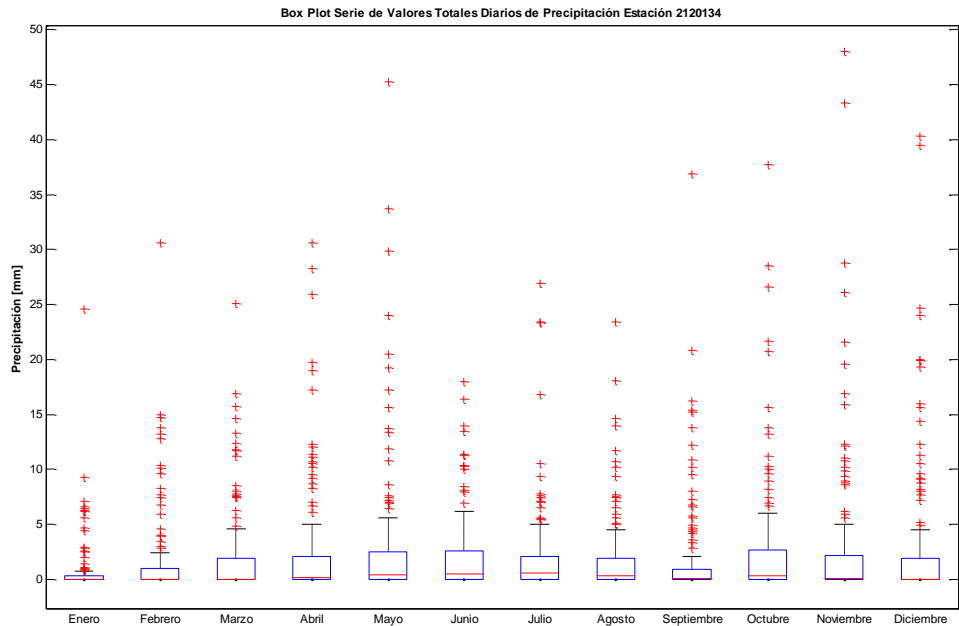


13. PARQUE SOPO-2120134

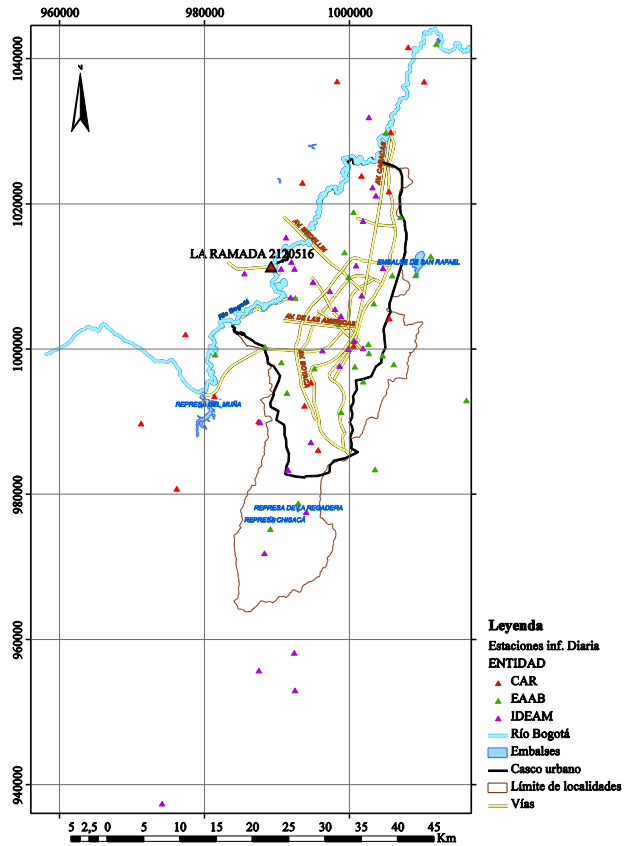


Este	1010170
Norte	1036950
Cota	2540
Número de Datos	1790
Número de Datos Sin Registro	36
Media	2.2
Mediana	0.1
Máximo	48.0
Desviación Estándar	5.0
Varianza	24.8
Coefficiente de Variación	8.5E-04
Coefficiente de Asimetría	4.2
Coefficiente de Curtosis	25.6

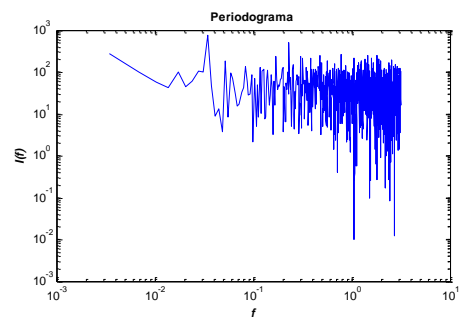
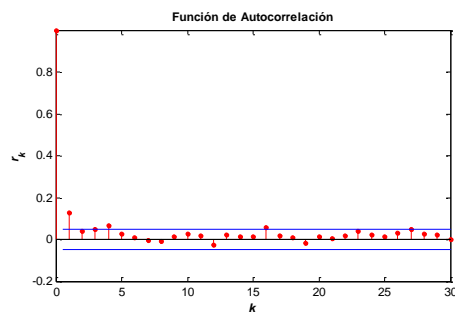
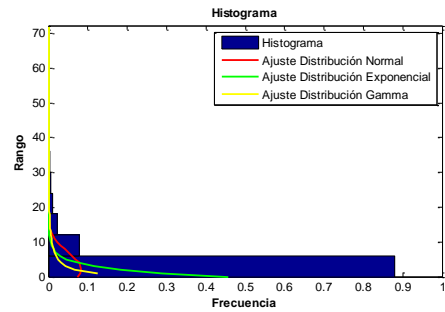
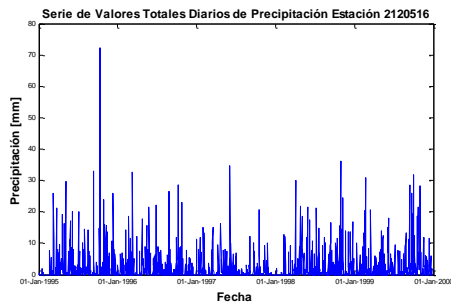


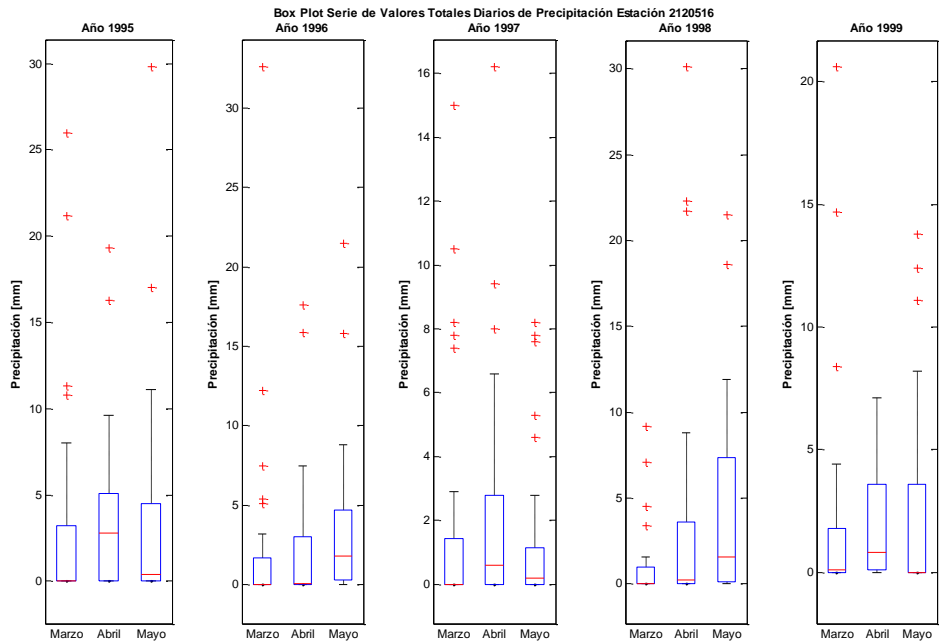
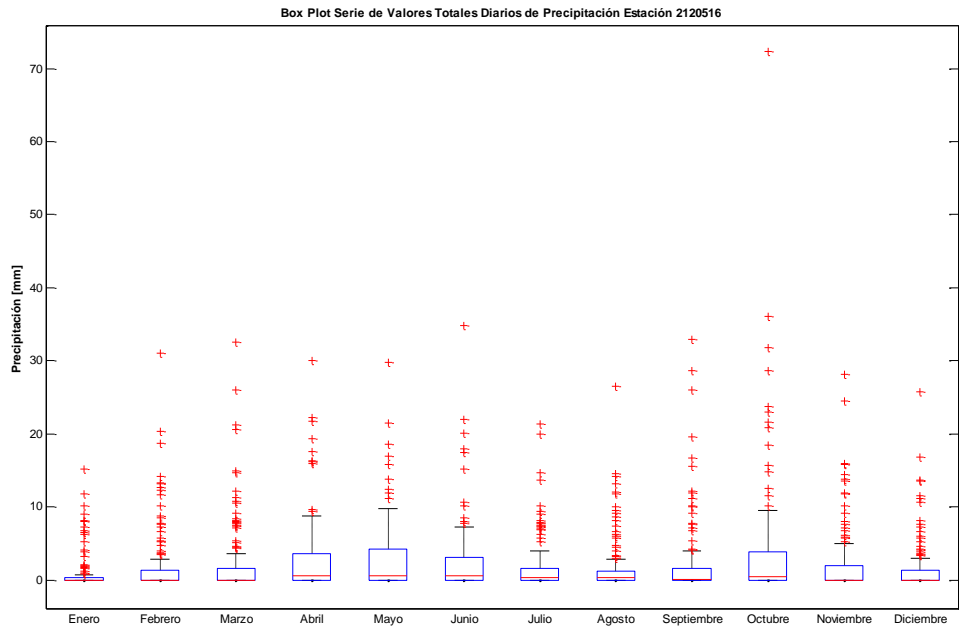


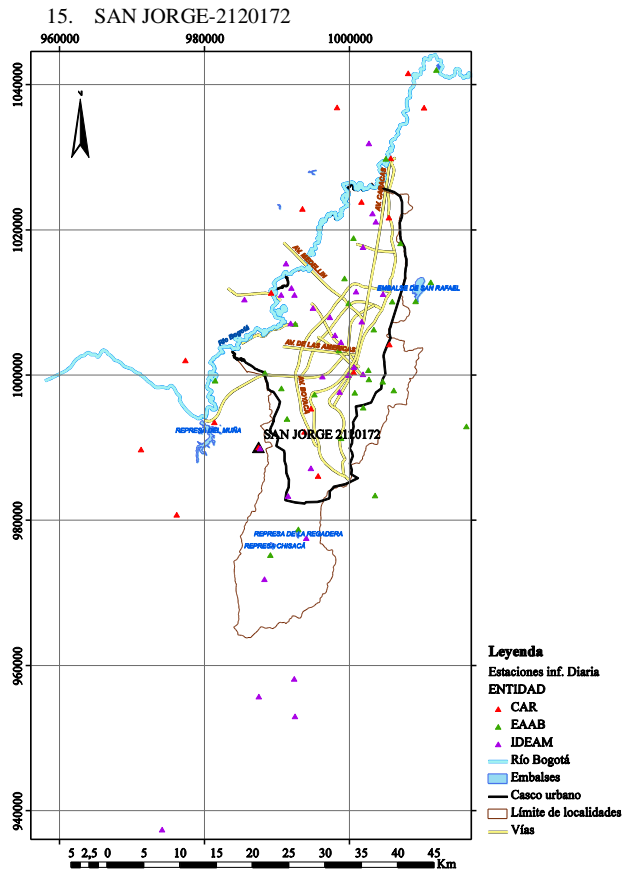
14. LA RAMADA-2120516



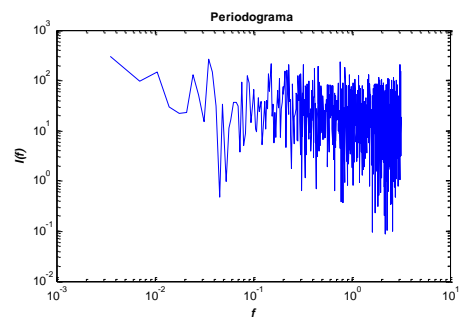
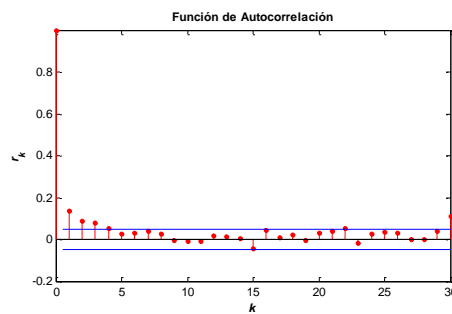
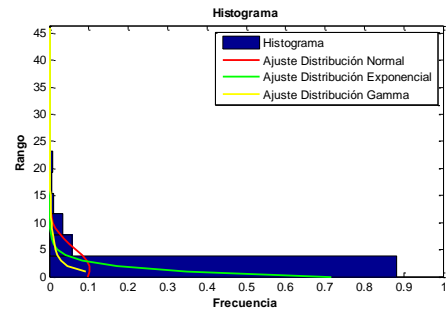
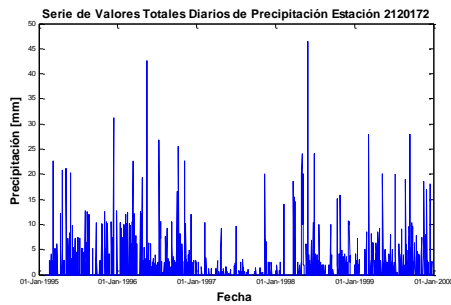
Este	989110
Norte	1011430
Cota	2545
Número de Datos	1826
Número de Datos Sin Registro	0
Media	2.2
Mediana	0.1
Máximo	72.3
Desviación Estándar	4.8
Varianza	23.2
Coefficiente de Variación	8.6E-04
Coefficiente de Asimetría	4.6
Coefficiente de Curtosis	38.6

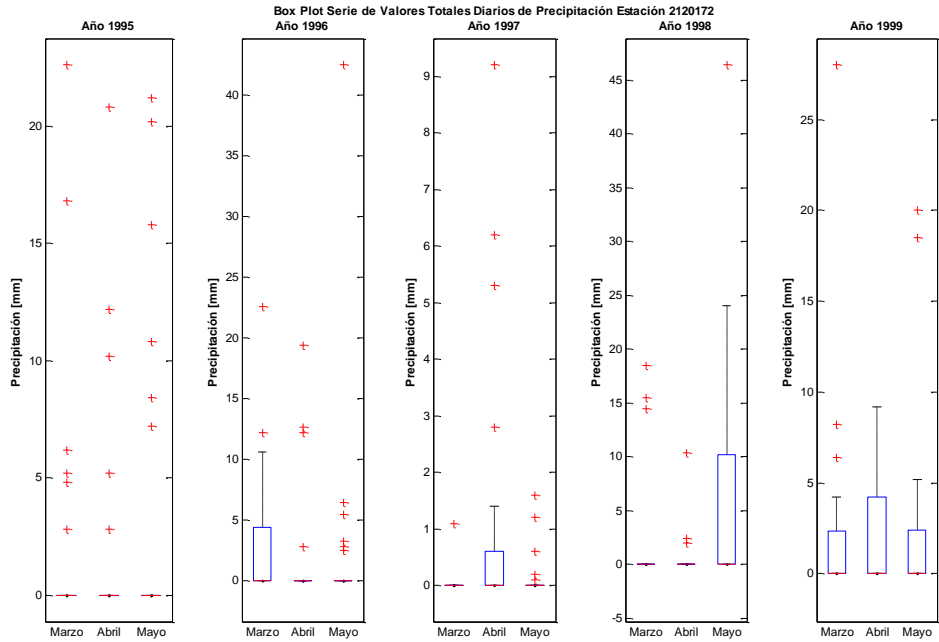
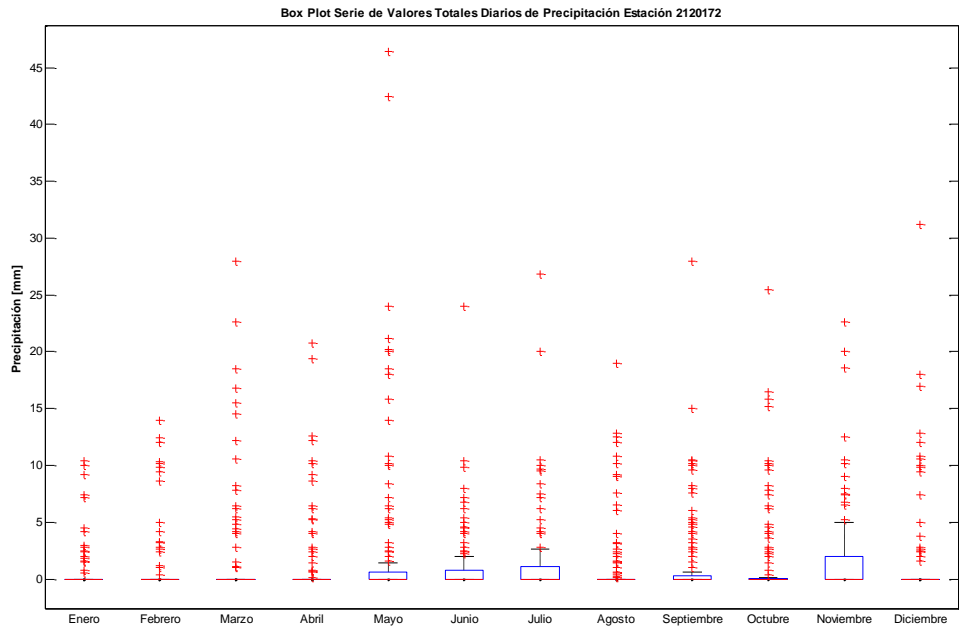


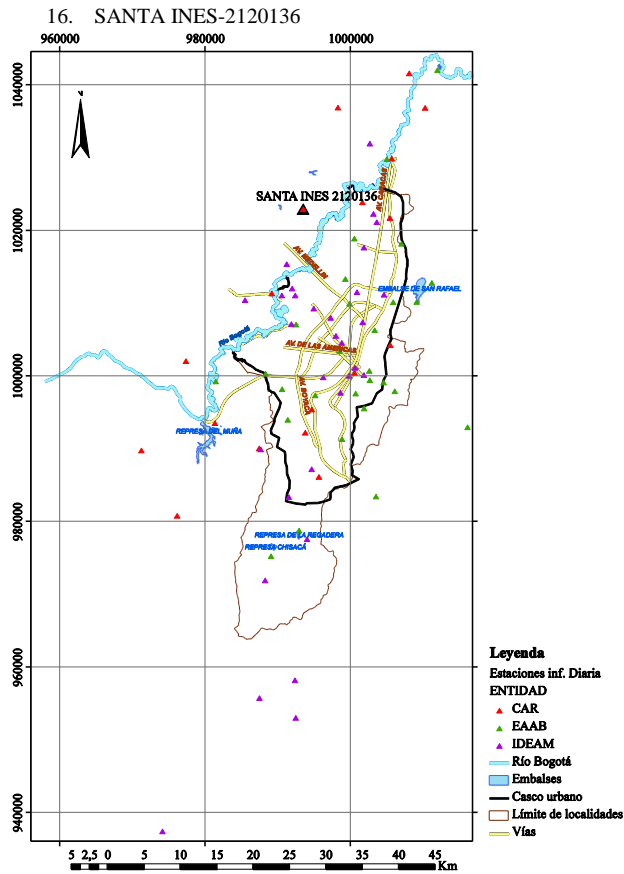




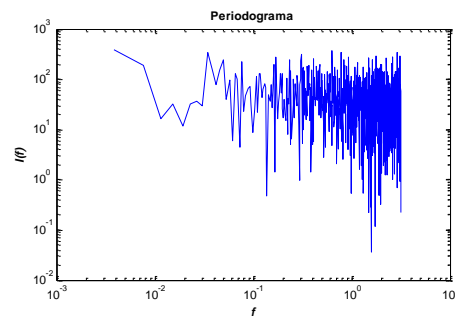
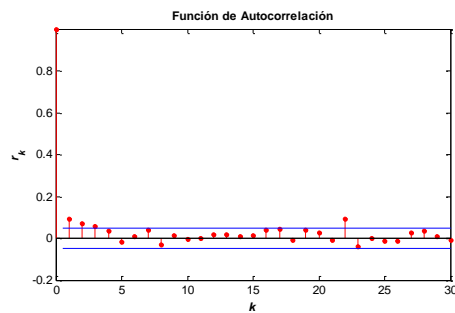
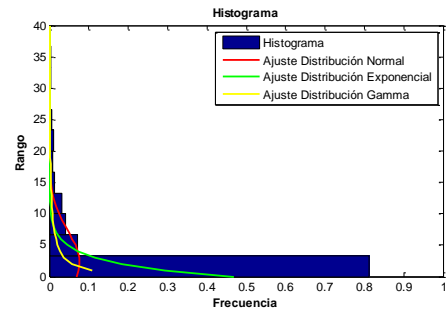
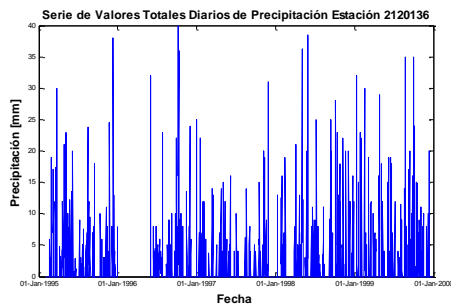
Este	987380
Norte	990160
Cota	2890
Número de Datos	1795
Número de Datos Sin Registro	31
Media	1.4
Mediana	0.0
Máximo	46.4
Desviación Estándar	3.9
Varianza	15.2
Coefficiente de Variación	4.8E-04
Coefficiente de Asimetría	4.6
Coefficiente de Curtosis	32.2

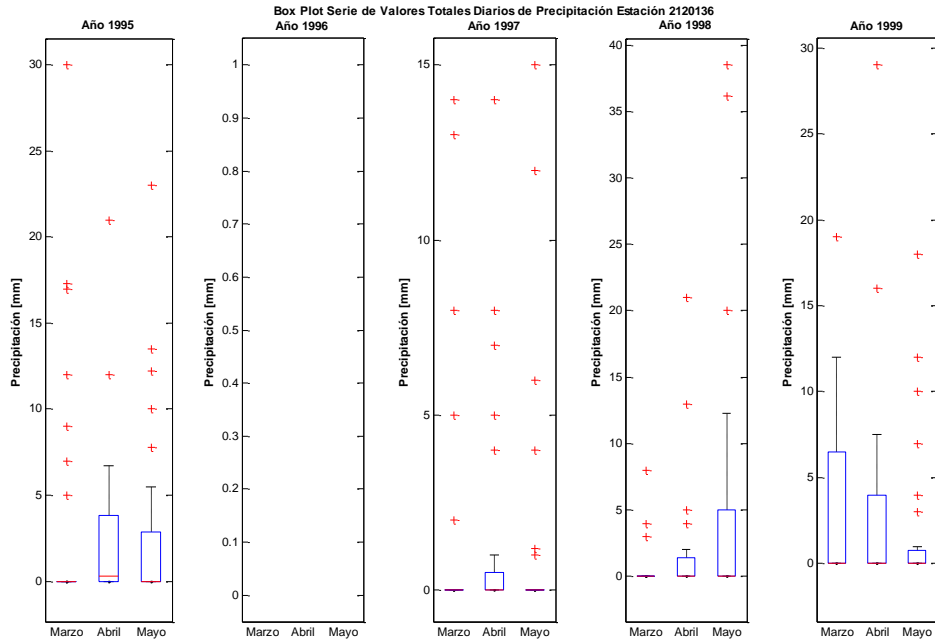
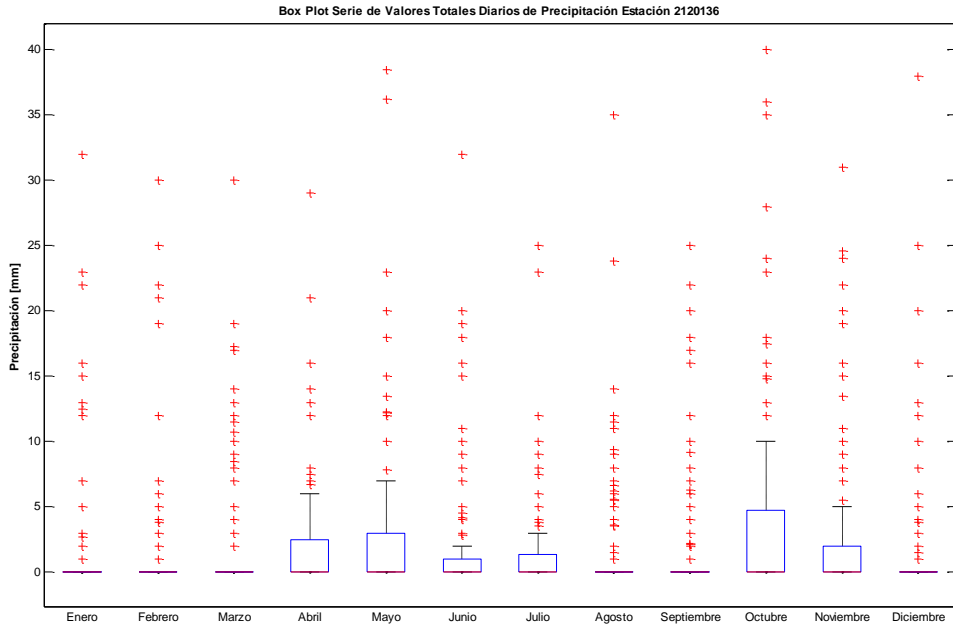


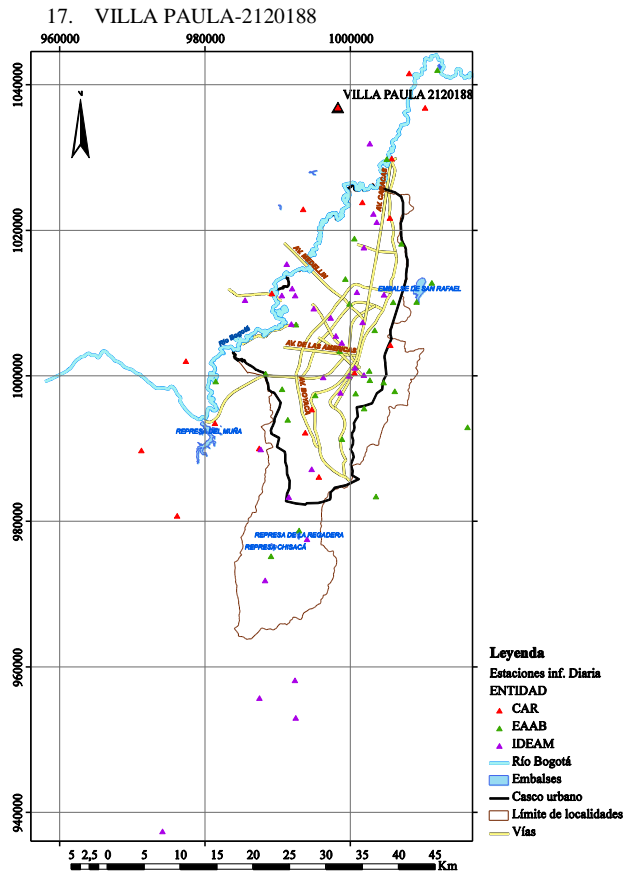




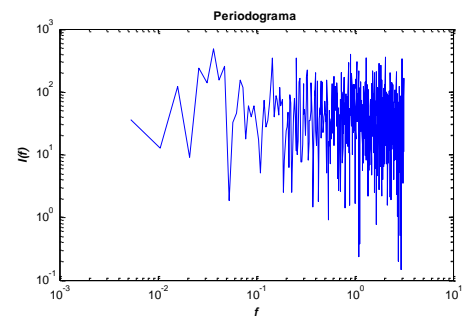
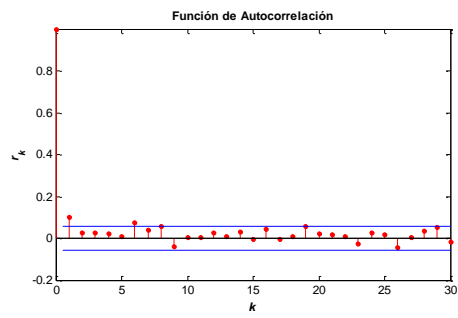
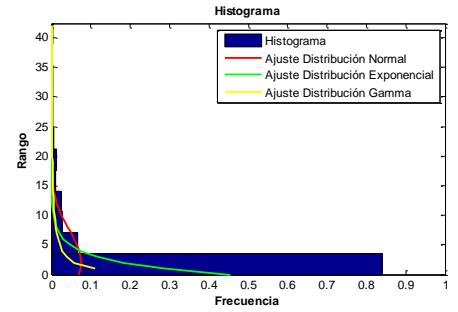
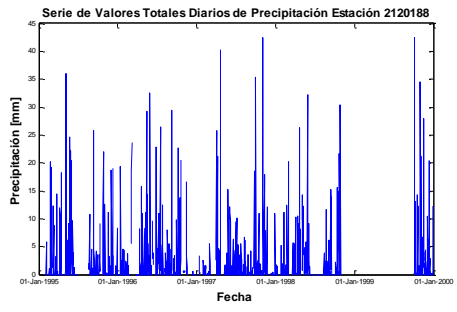
Este	993400
Norte	1023000
Cota	2550
Número de Datos	1642
Número de Datos Sin Registro	184
Media	2.1
Mediana	0.0
Máximo	40.0
Desviación Estándar	5.2
Varianza	27.4
Coefficiente de Variación	8.4E-04
Coefficiente de Asimetría	3.5
Coefficiente de Curtosis	17.4

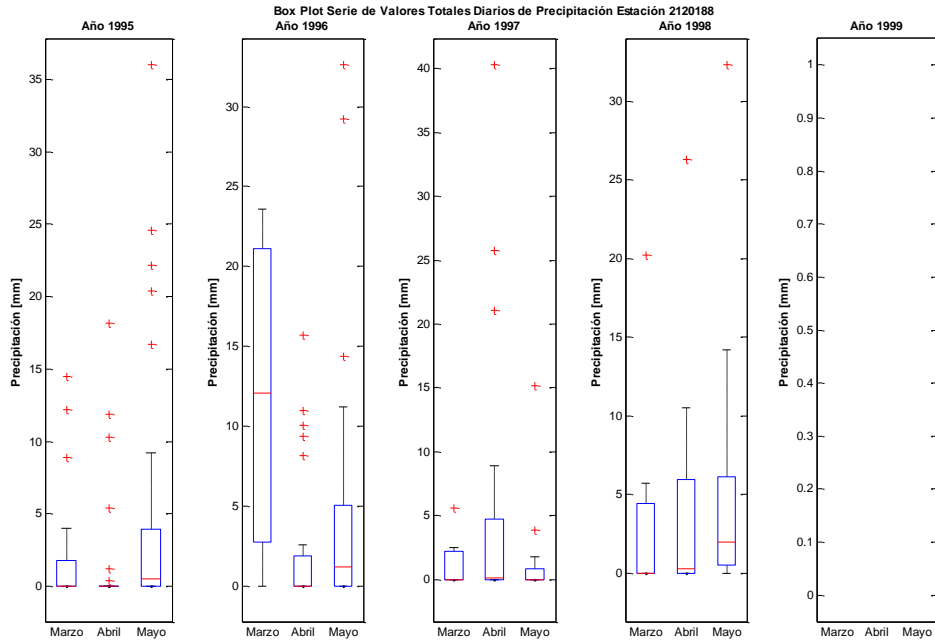
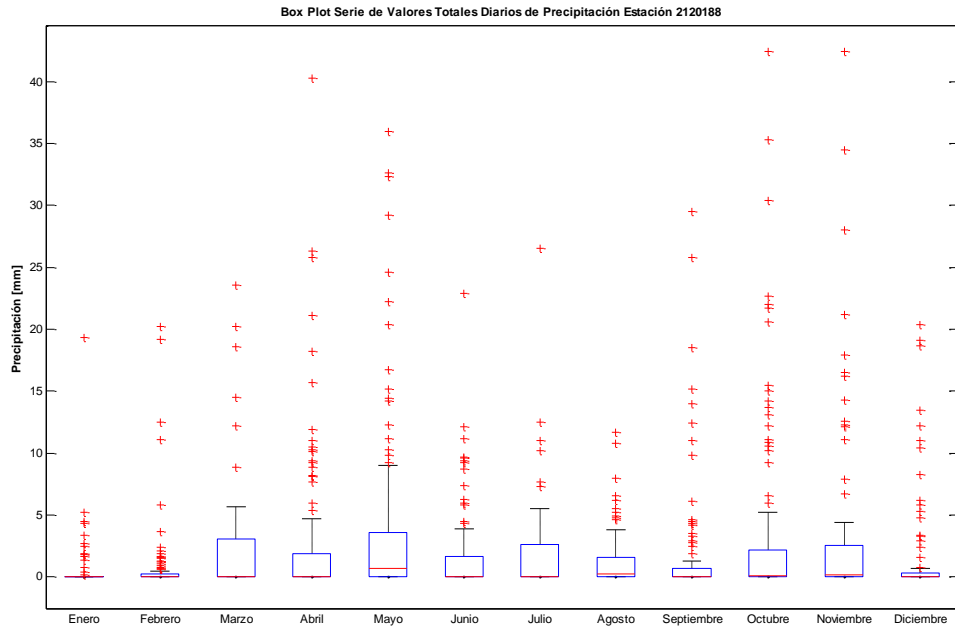


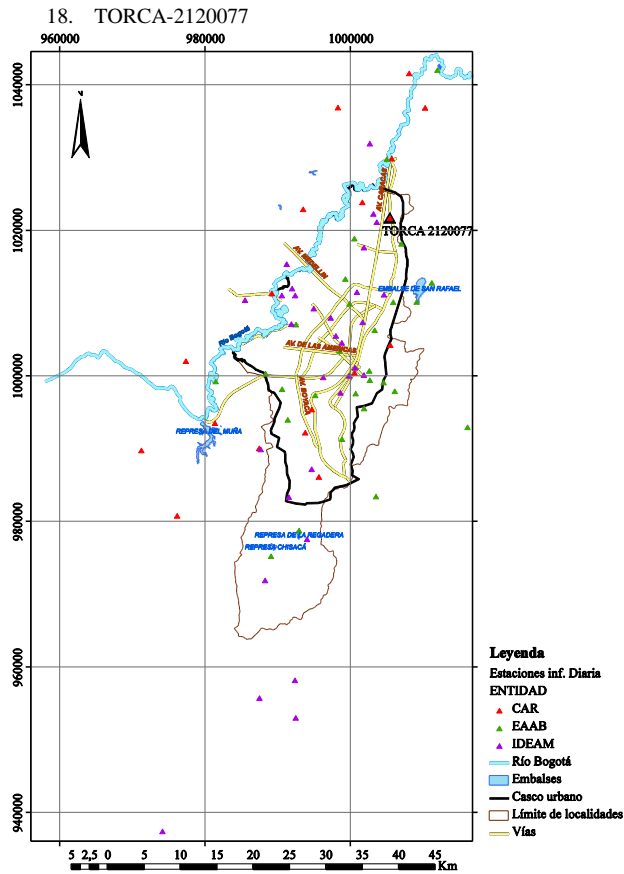




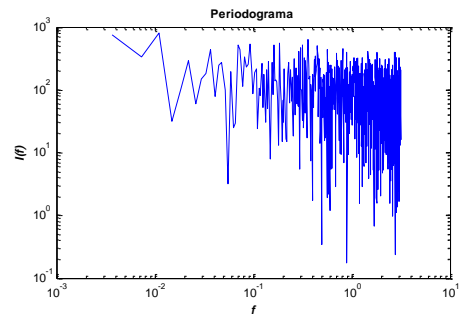
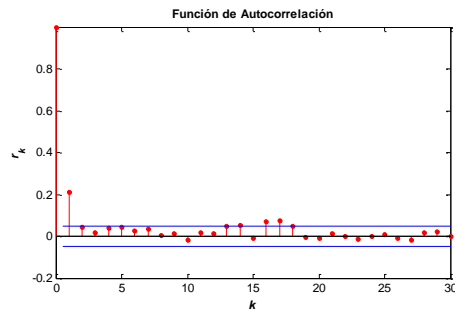
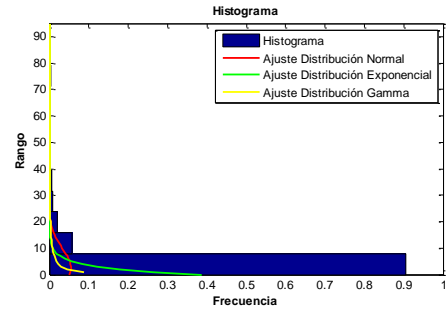
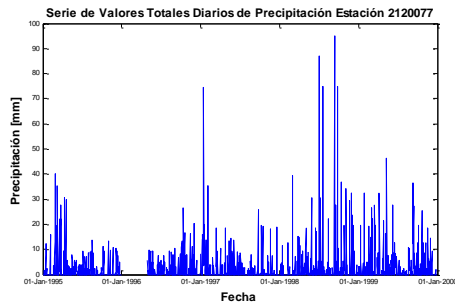
Este	998200
Norte	1036980
Cota	2750
Número de Datos	1195
Número de Datos Sin Registro	631
Media	2.2
Mediana	0.0
Máximo	42.4
Desviación Estándar	5.3
Varianza	28.1
Coefficiente de Variación	8.0E-04
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	20.7

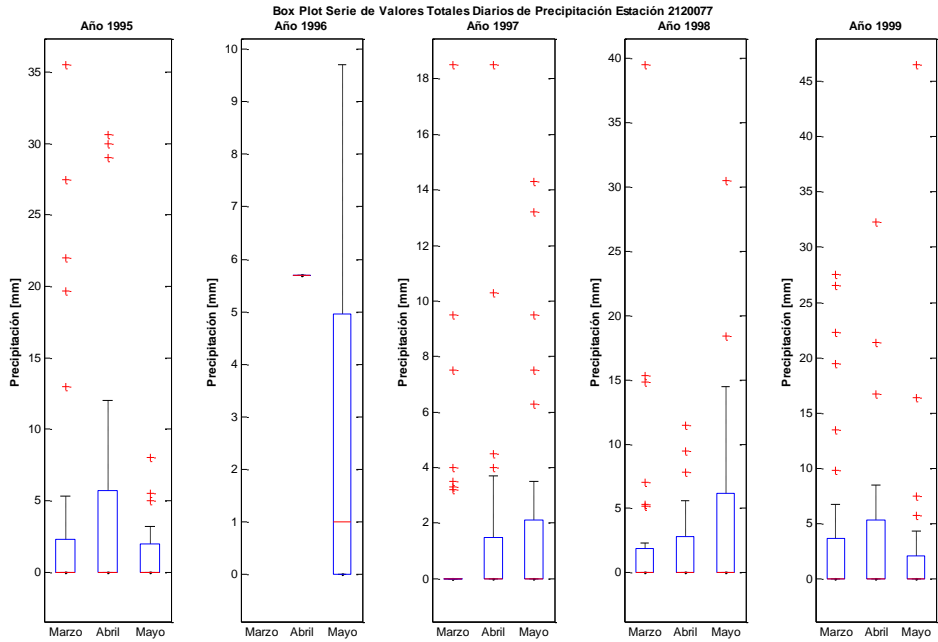
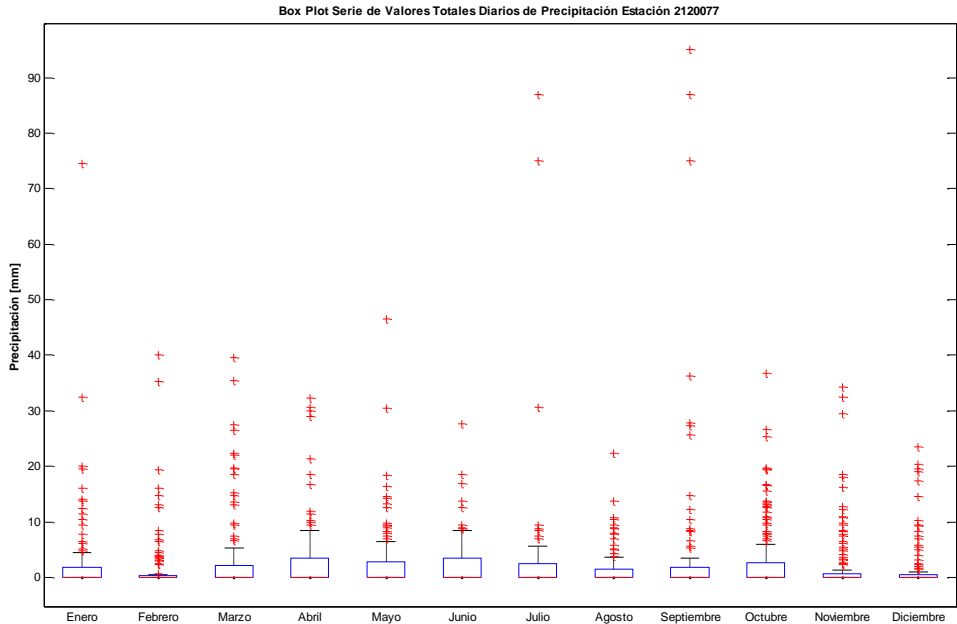


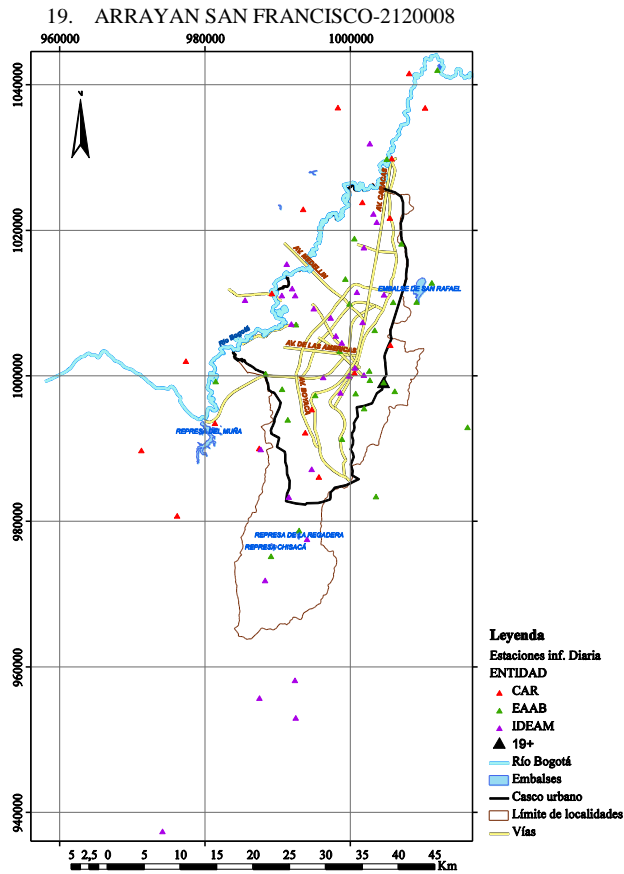




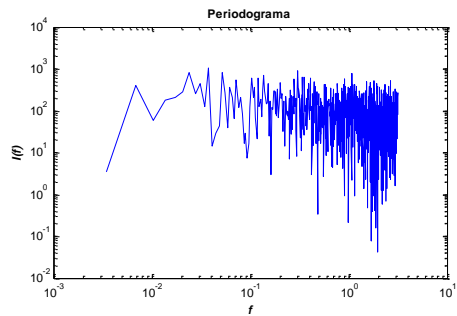
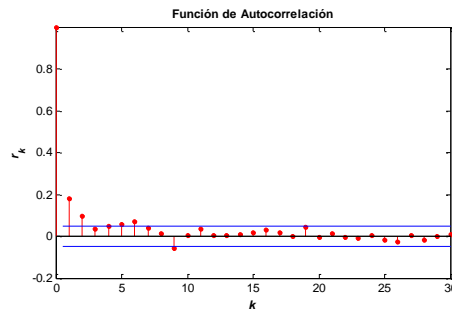
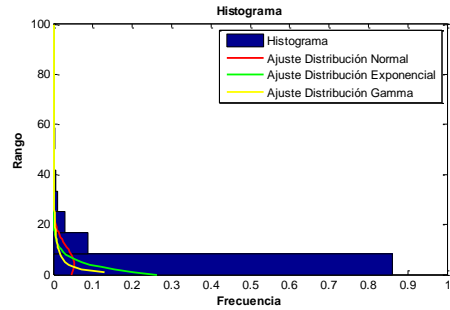
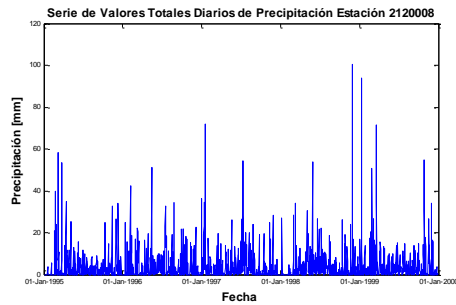
Este	1005320
Norte	1021800
Cota	2579
Número de Datos	1704
Número de Datos Sin Registro	122
Media	2.6
Mediana	0.0
Máximo	95.0
Desviación Estándar	7.3
Varianza	53.8
Coefficiente de Variación	1.0E-03
Coefficiente de Asimetría	6.6
Coefficiente de Curtosis	62.9

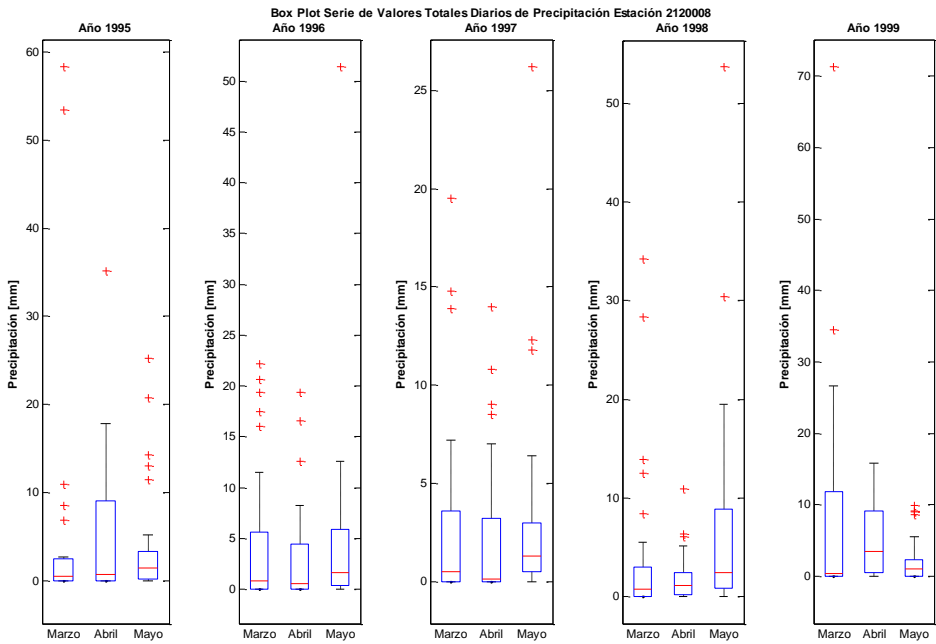
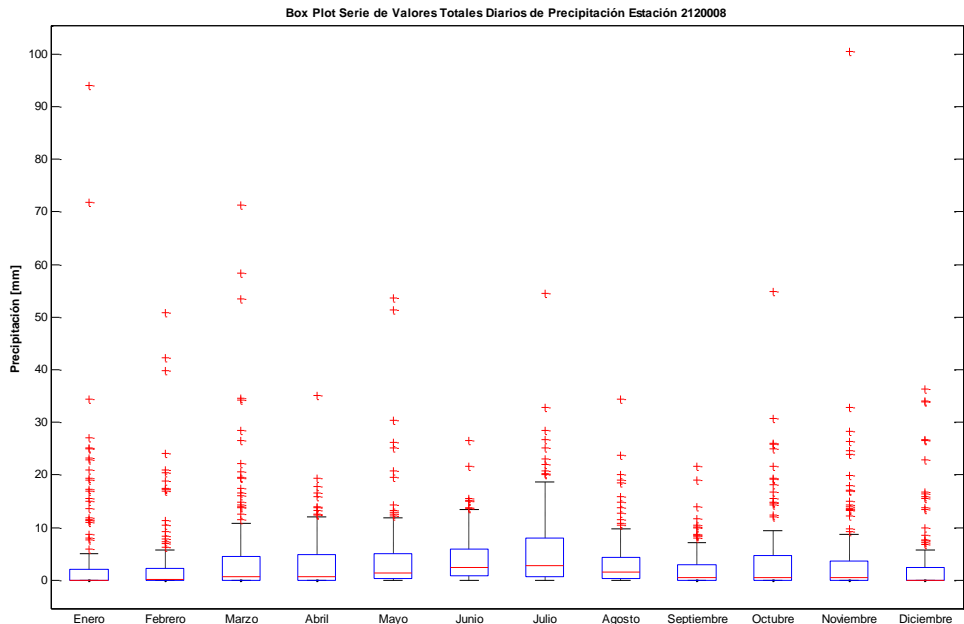


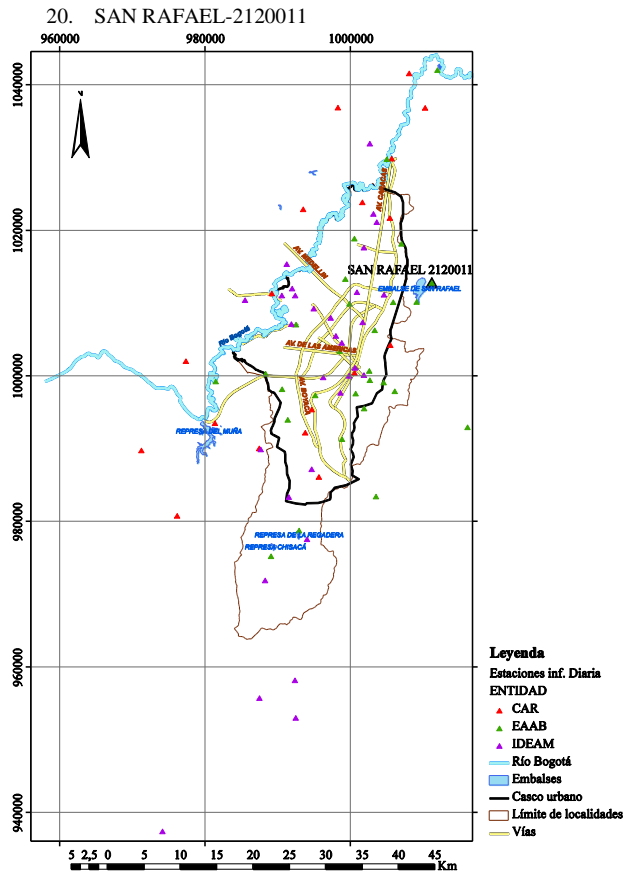




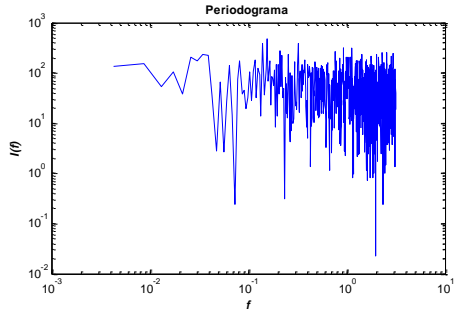
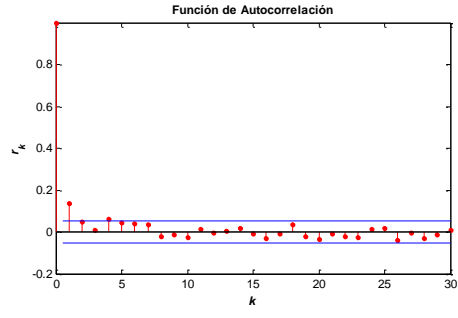
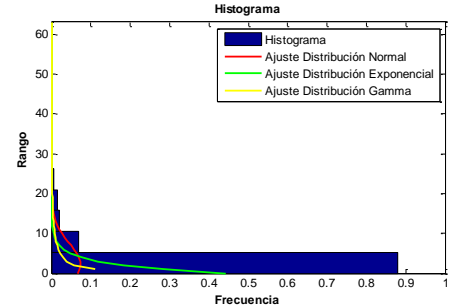
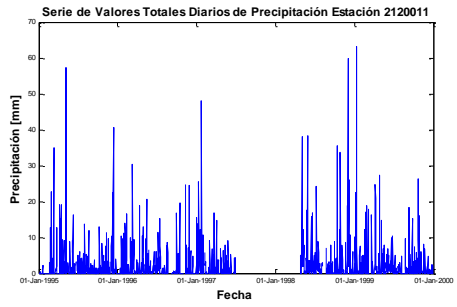
Este	1004500
Norte	999200
Cota	3047
Número de Datos	1826
Número de Datos Sin Registro	0
Media	3.8
Mediana	0.8
Máximo	100.4
Desviación Estándar	7.6
Varianza	58.3
Coefficiente de Variación	1.2E-03
Coefficiente de Asimetría	5.0
Coefficiente de Curtosis	42.5

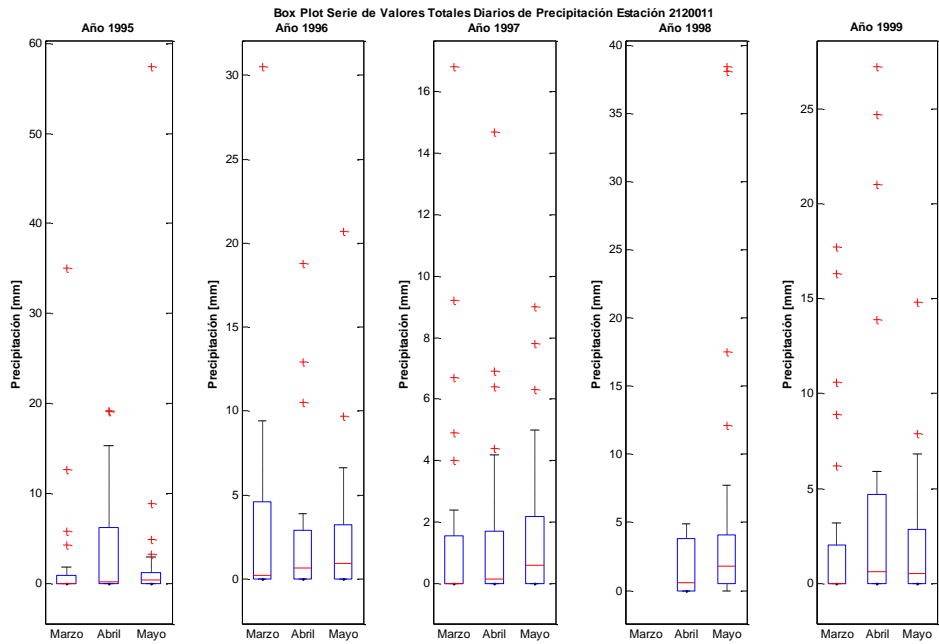
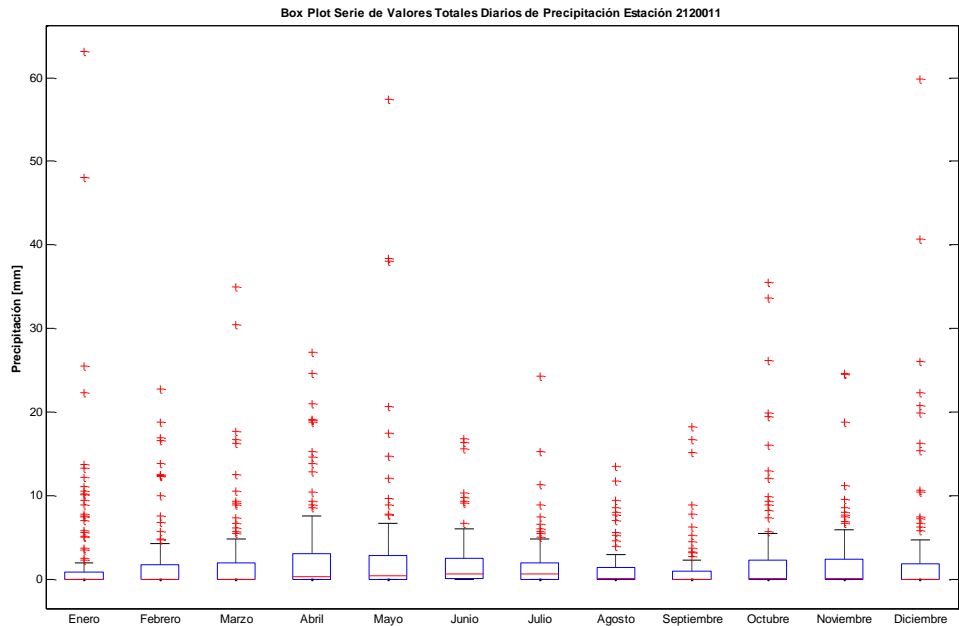




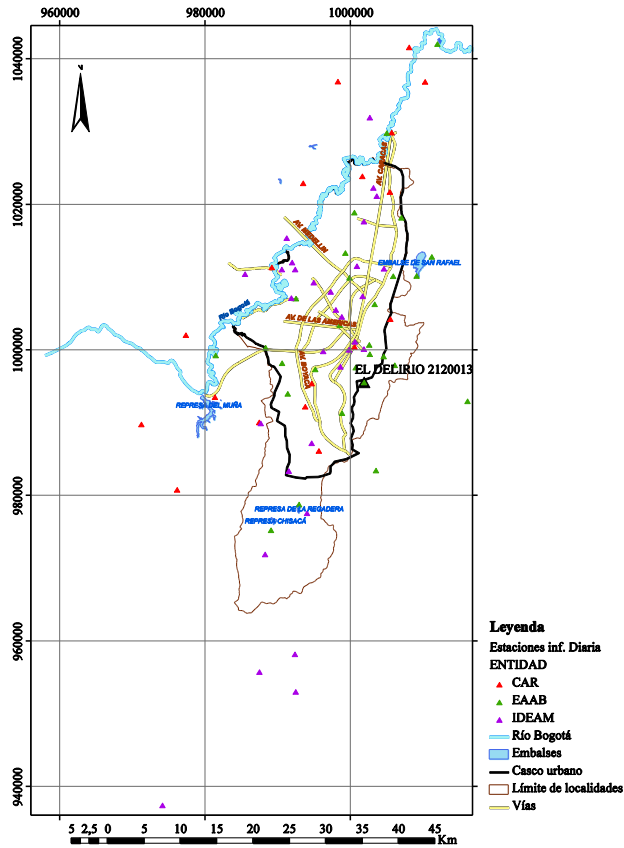


Este	1011100
Norte	1012902
Cota	2750
Número de Datos	1455
Número de Datos Sin Registro	371
Media	2.3
Mediana	0.2
Máximo	63.1
Desviación Estándar	5.4
Varianza	29.5
Coefficiente de Variación	8.2E-04
Coefficiente de Asimetría	5.2
Coefficiente de Curtosis	41.7

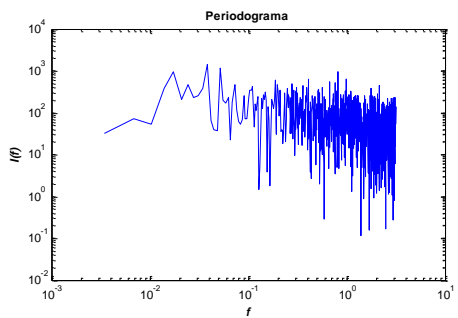
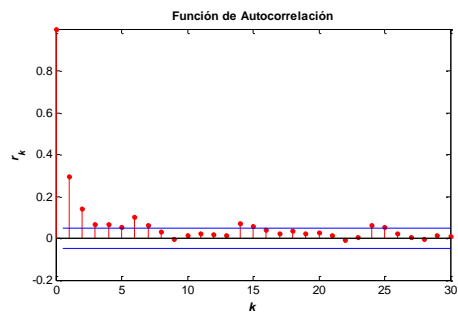
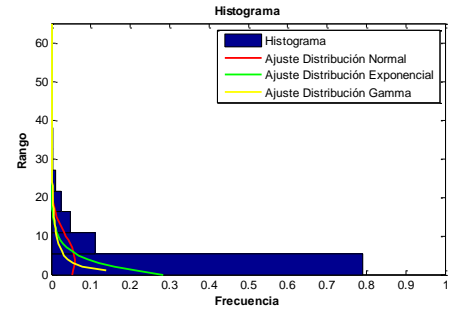
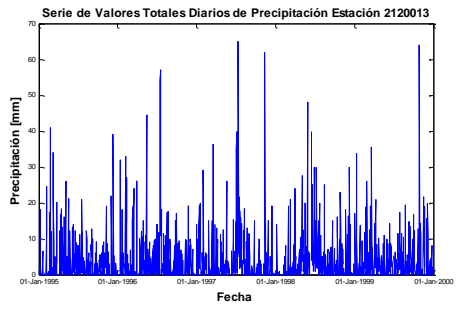


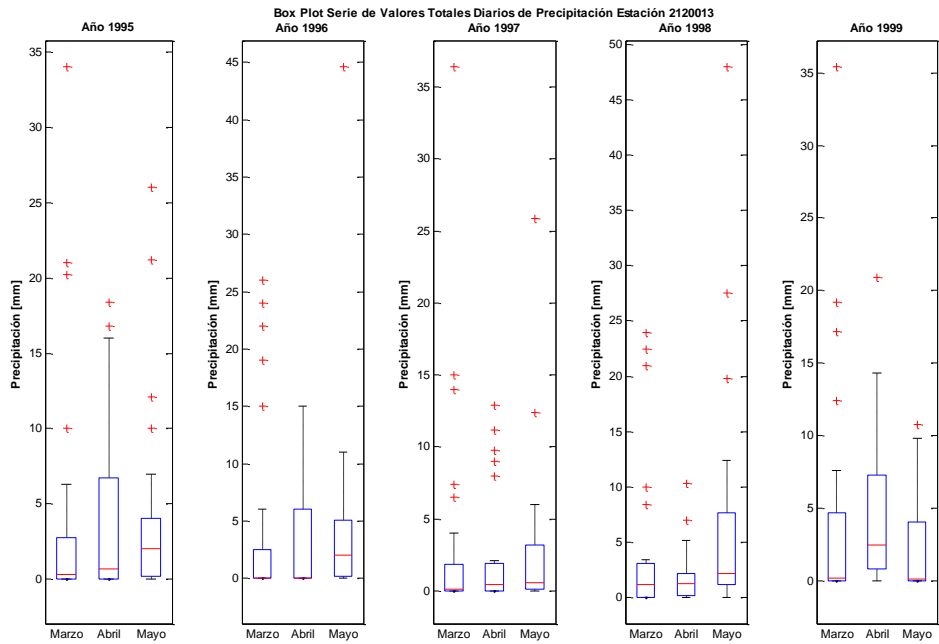
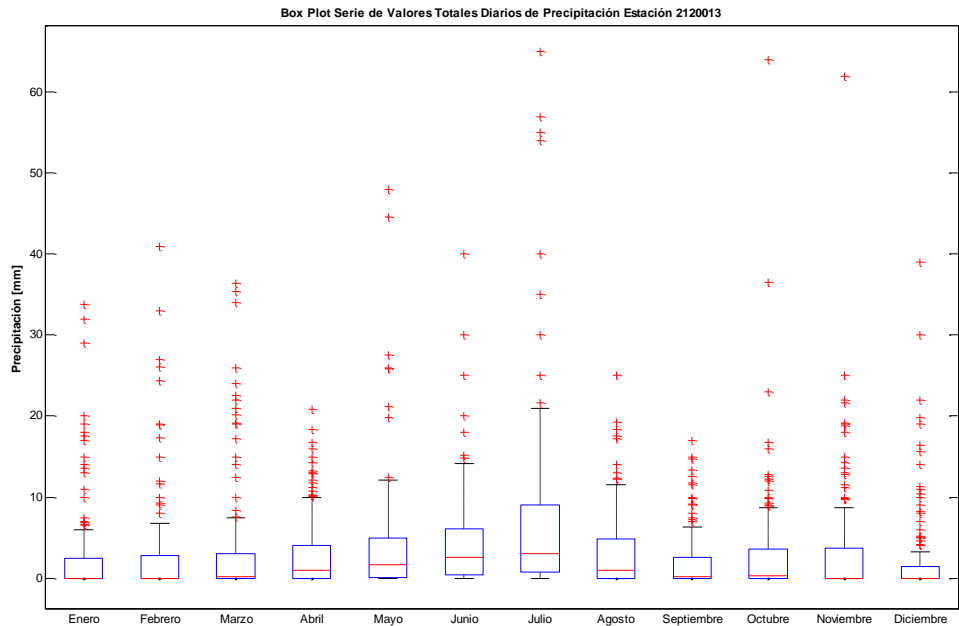


21. EL DELIRIO-2120013

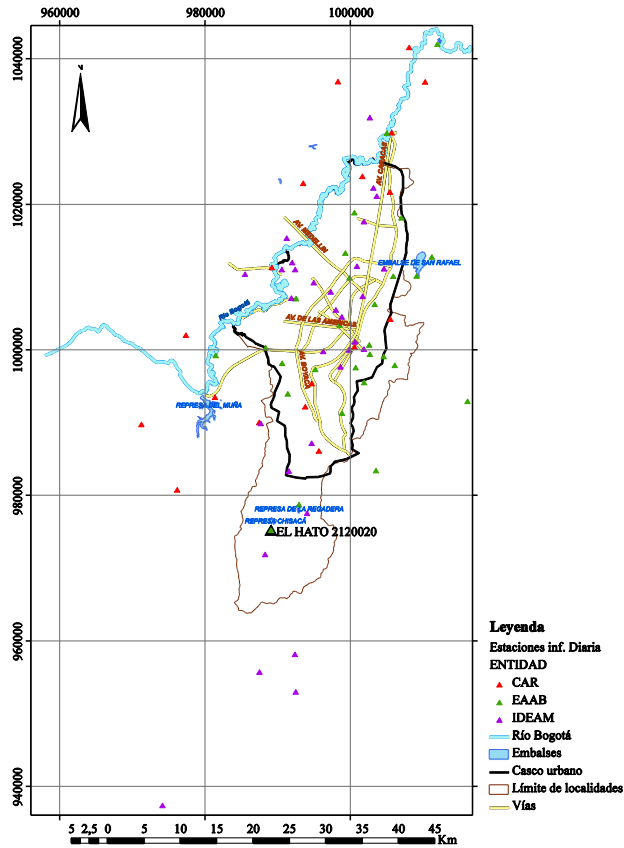


Este	1001802
Norte	995615
Cota	3000
Número de Datos	1812
Número de Datos Sin Registro	14
Media	3.5
Mediana	0.6
Máximo	65.0
Desviación Estándar	6.7
Varianza	45.5
Coefficiente de Variación	1.2E-03
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	24.5

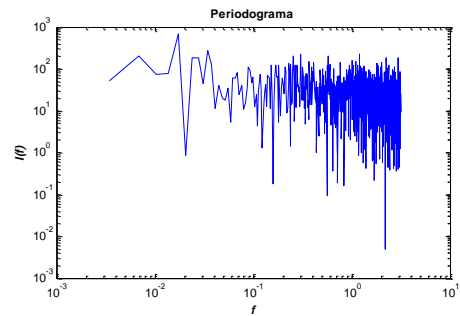
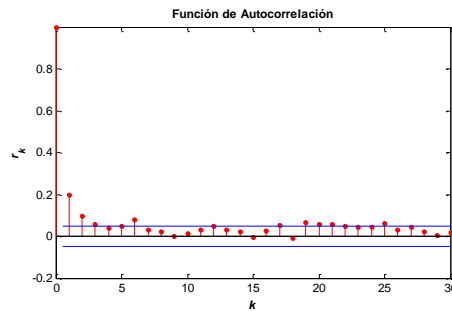
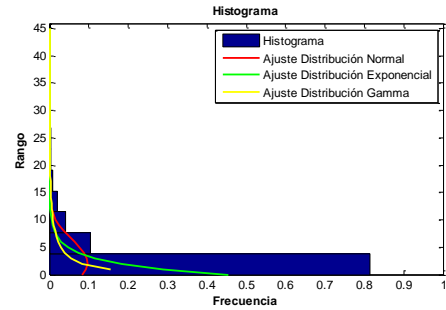
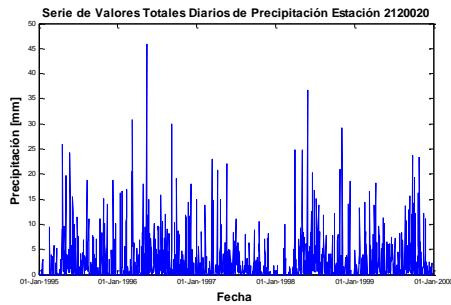


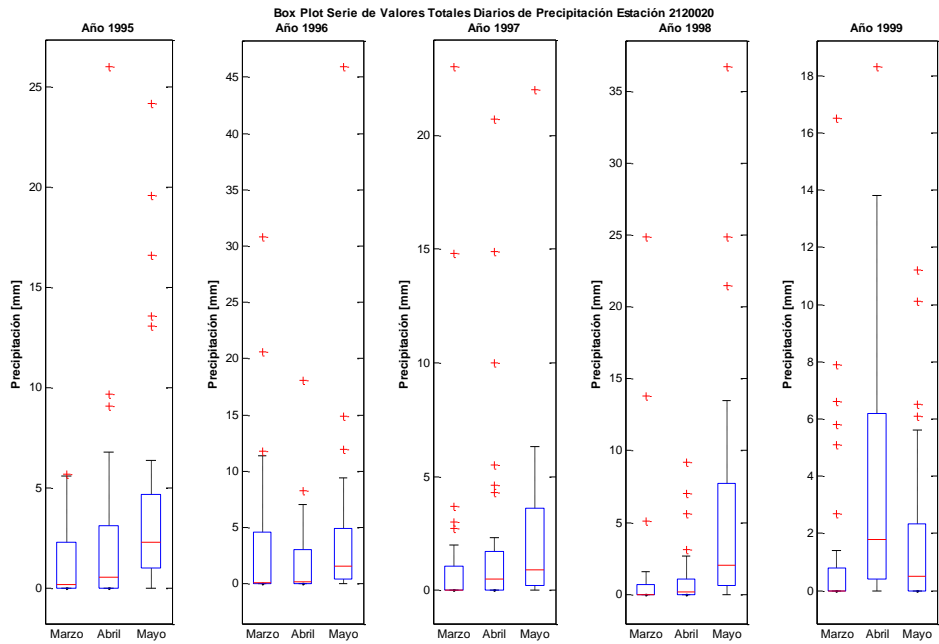
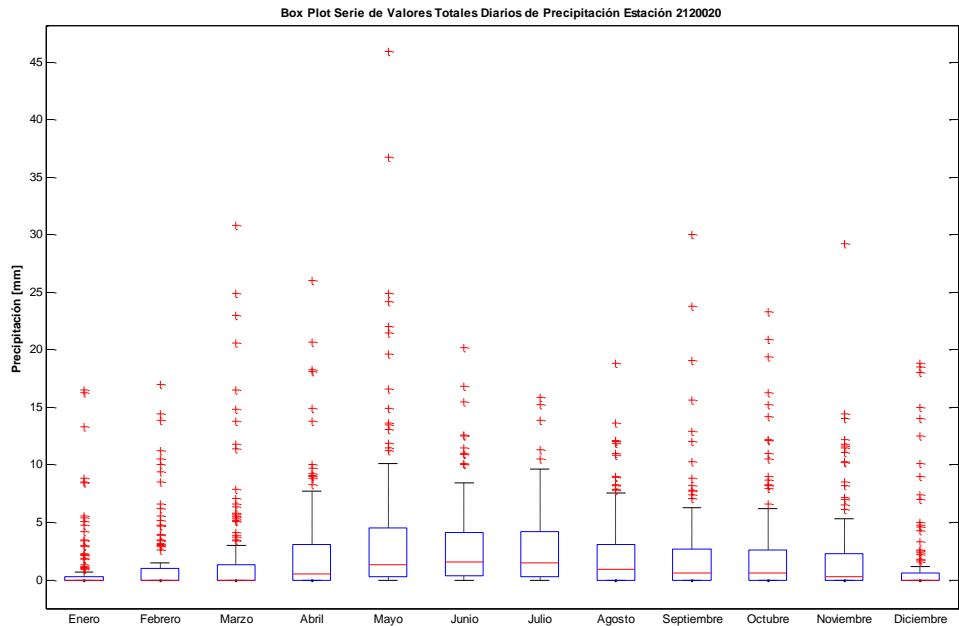


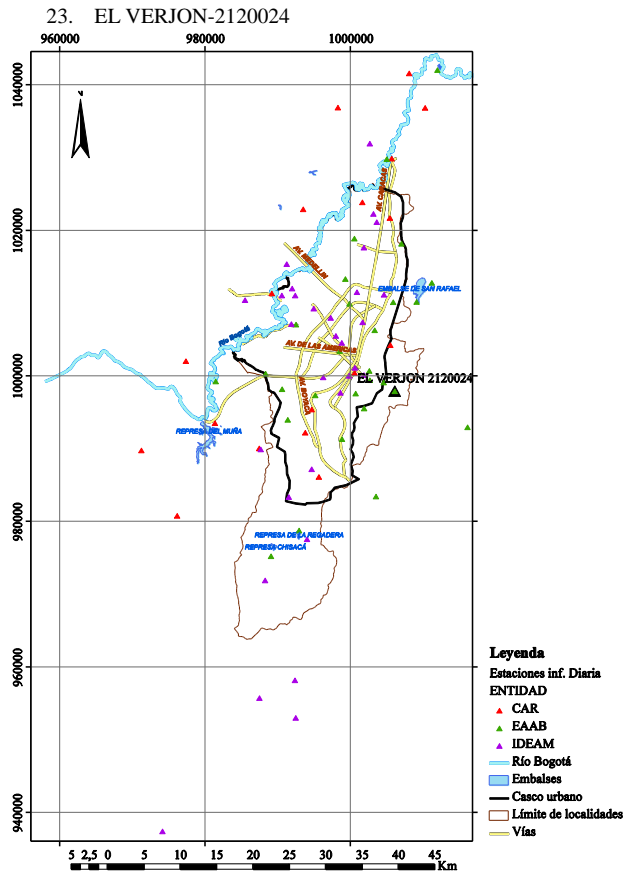
22. EL HATO-2120020



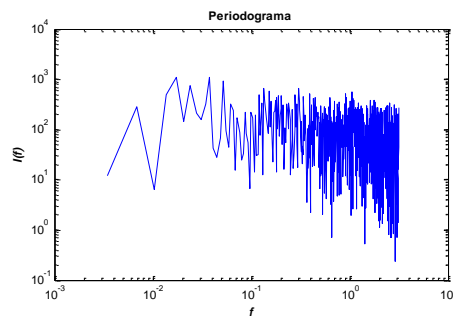
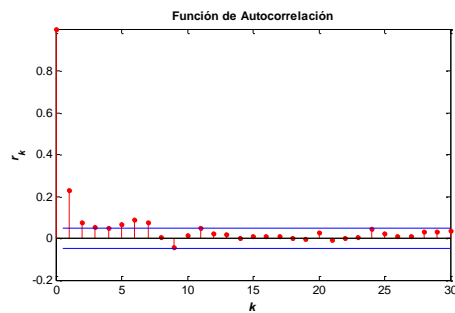
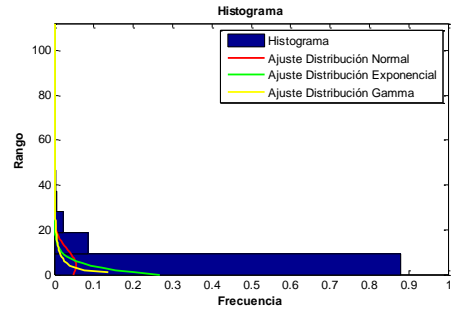
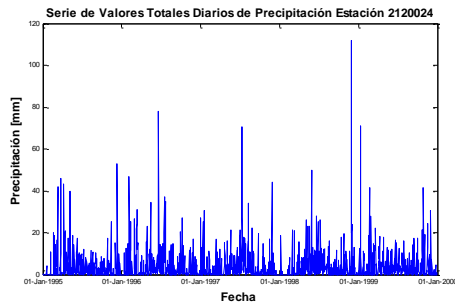
Este	988998
Norte	975335
Cota	3150
Número de Datos	1826
Número de Datos Sin Registro	0
Media	2.2
Mediana	0.5
Máximo	45.9
Desviación Estándar	4.1
Varianza	17.0
Coefficiente de Variación	7.0E-04
Coefficiente de Asimetría	3.6
Coefficiente de Curtosis	21.6

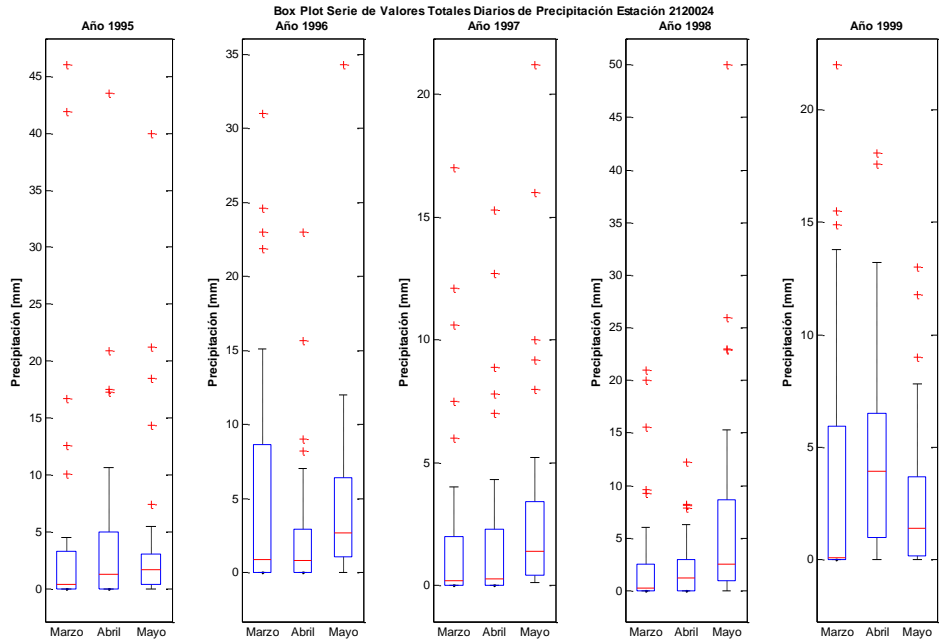
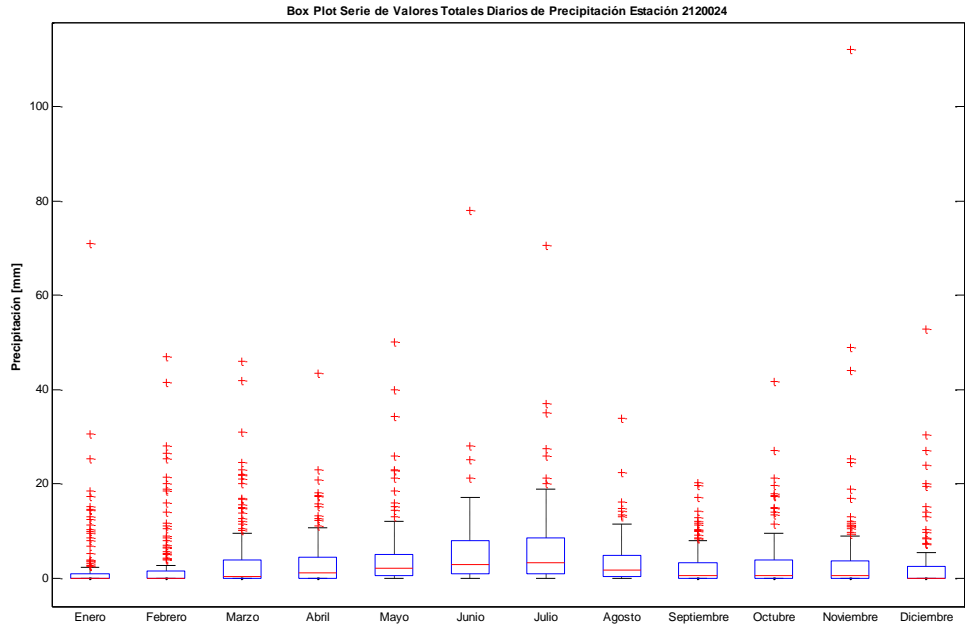


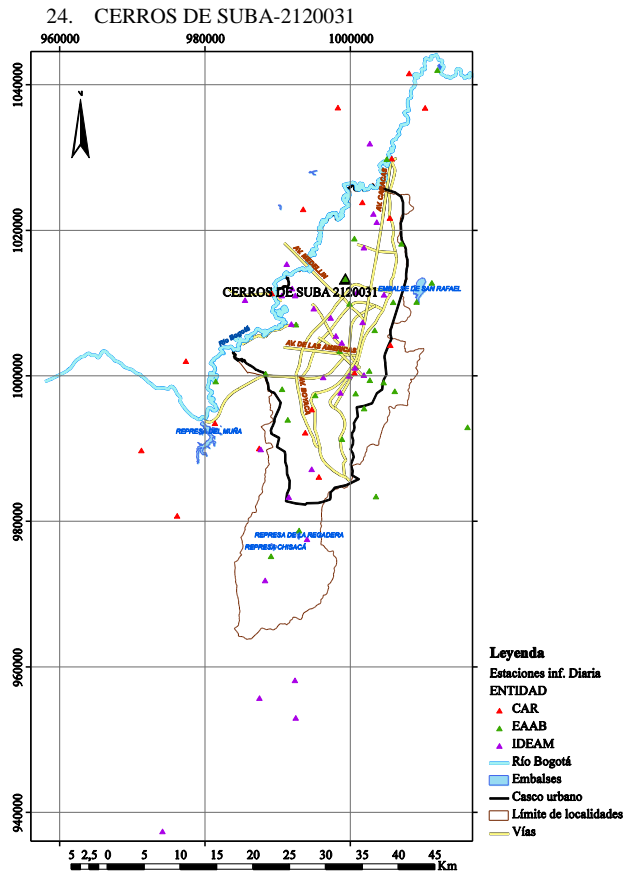




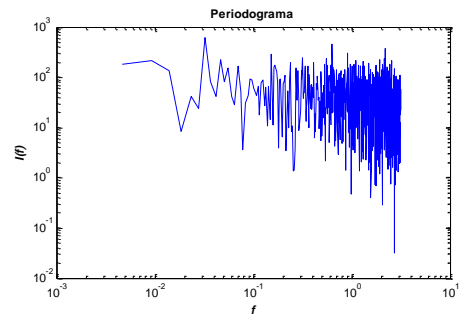
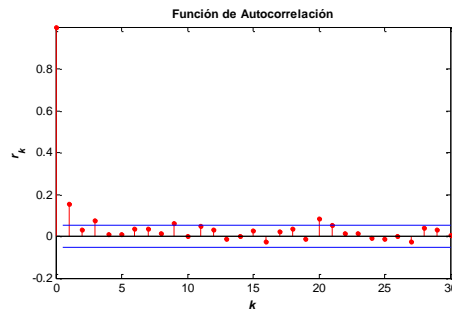
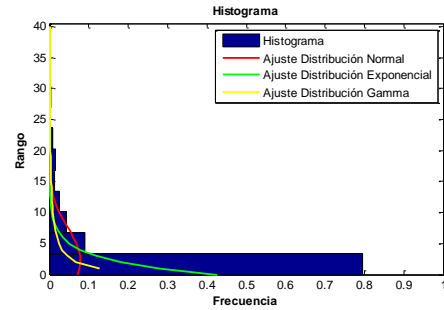
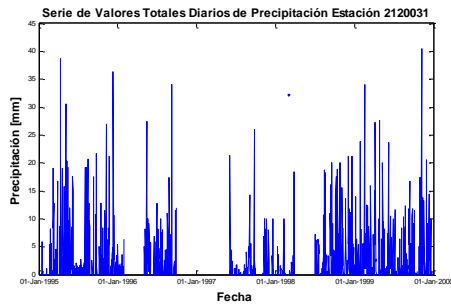
Este	1005995
Norte	998002
Cota	3250
Número de Datos	1825
Número de Datos Sin Registro	1
Media	3.7
Mediana	1.0
Máximo	112.0
Desviación Estándar	7.3
Varianza	52.9
Coefficiente de Variación	1.2E-03
Coefficiente de Asimetría	5.3
Coefficiente de Curtosis	50.2

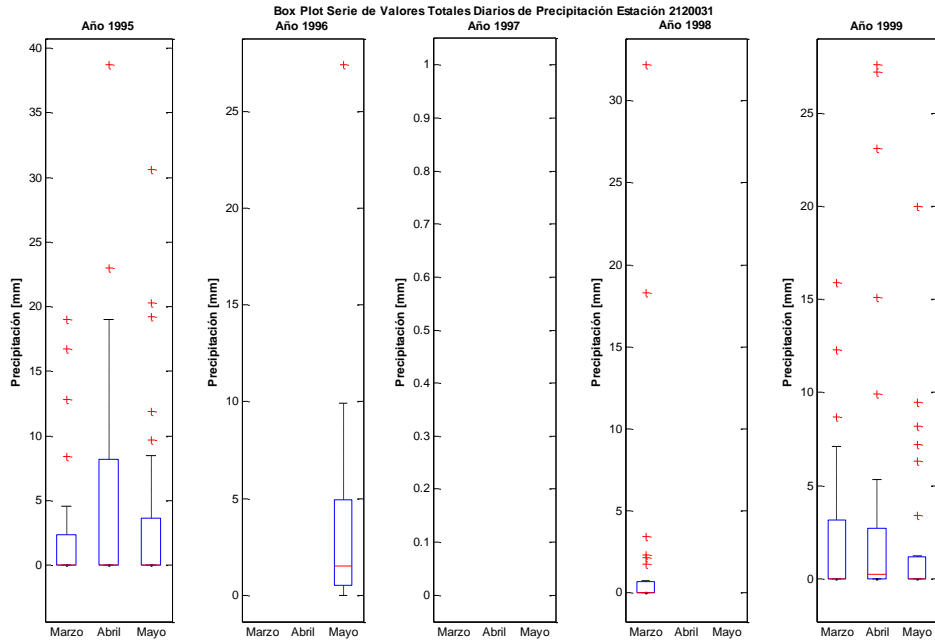
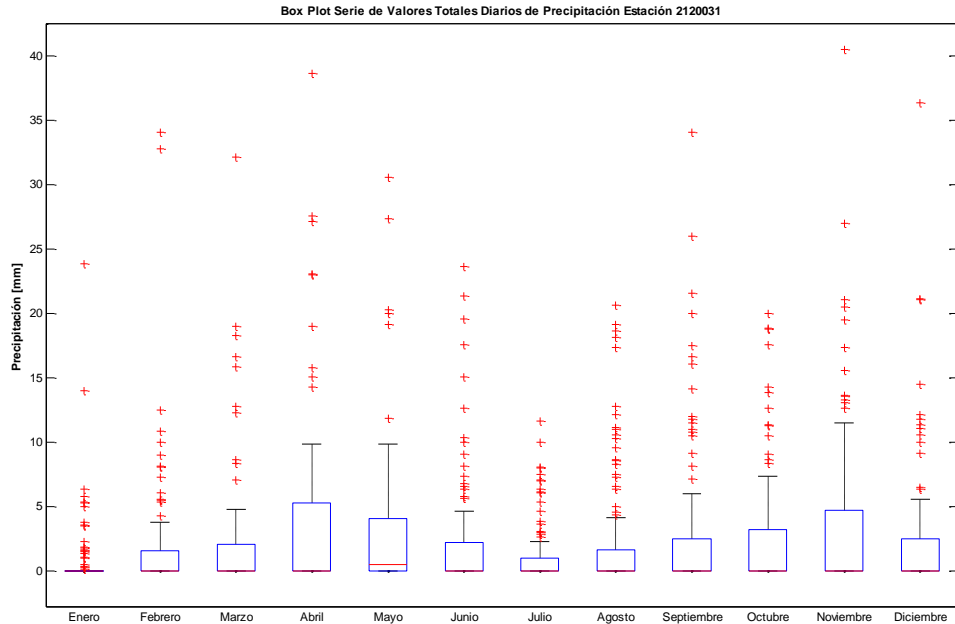




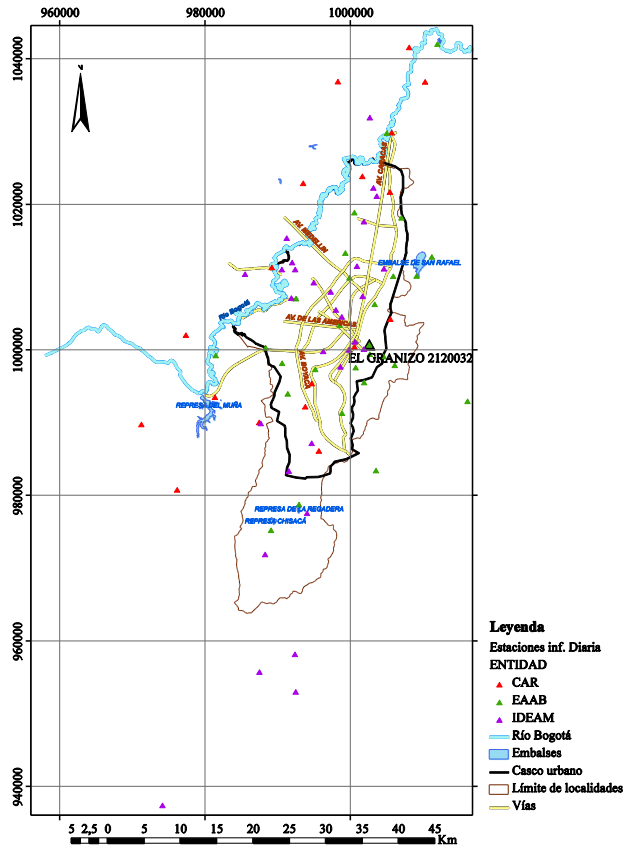


Este	999208
Norte	1013410
Cota	2691
Número de Datos	1354
Número de Datos Sin Registro	472
Media	2.3
Mediana	0.0
Máximo	40.5
Desviación Estándar	5.1
Varianza	26.0
Coefficiente de Variación	8.7E-04
Coefficiente de Asimetría	3.4
Coefficiente de Curtosis	16.9

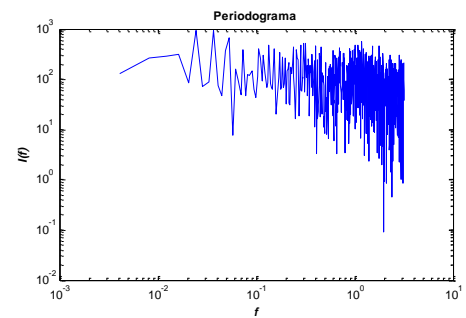
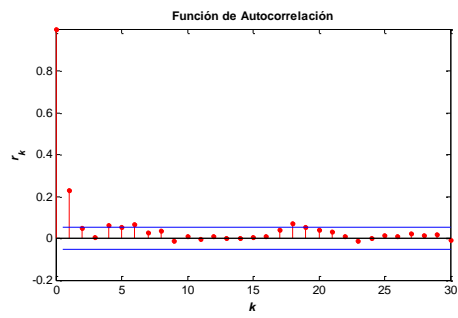
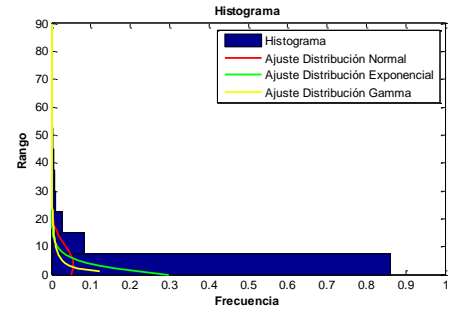
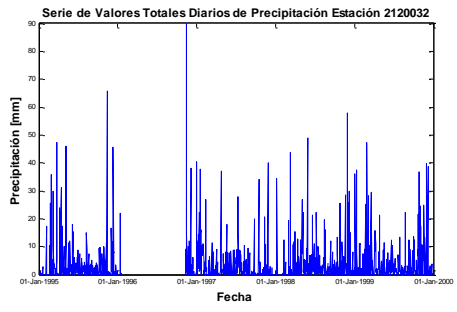


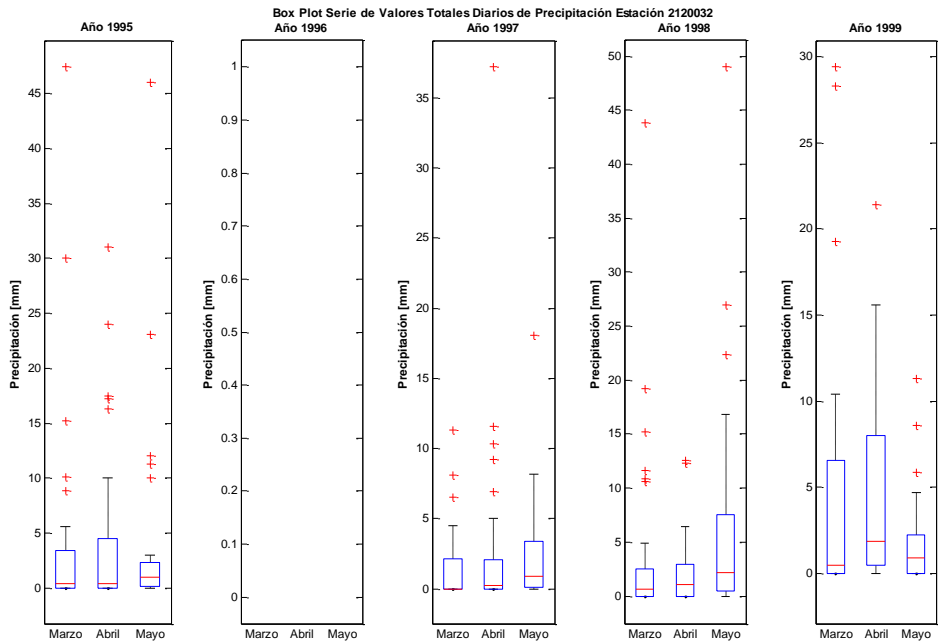
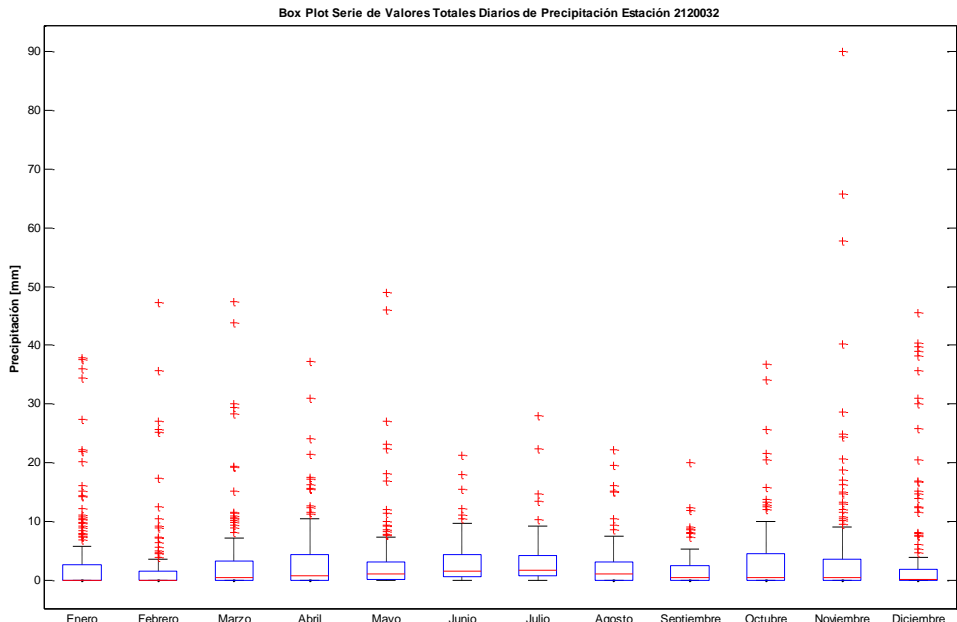


25. EL GRANIZO-2120032

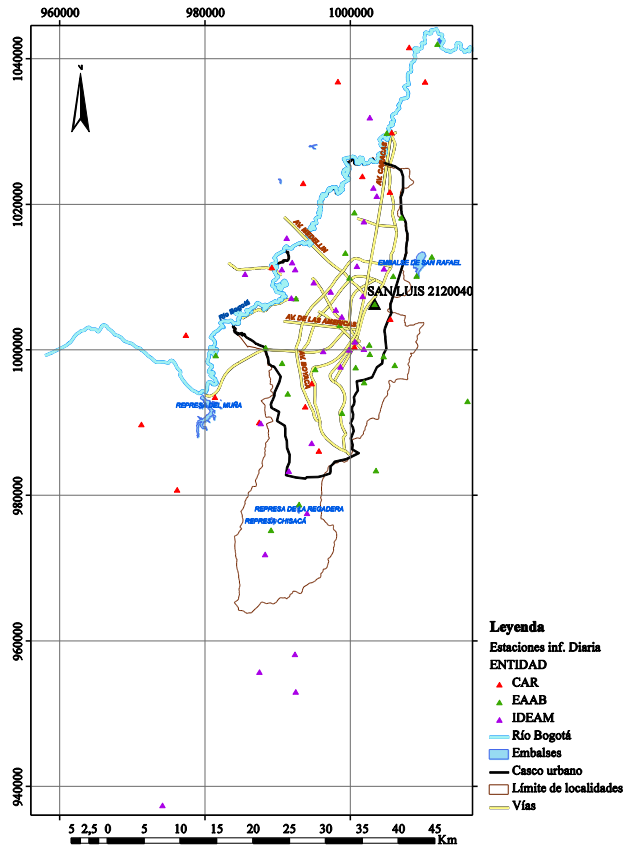


Este	1002517
Norte	1000782
Cota	3125
Número de Datos	1536
Número de Datos Sin Registro	290
Media	3.4
Mediana	0.6
Máximo	90.0
Desviación Estándar	7.2
Varianza	51.7
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	4.5
Coefficiente de Curtosis	31.8

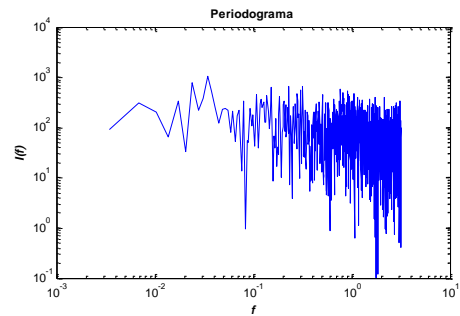
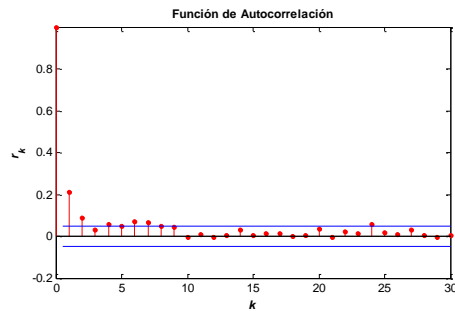
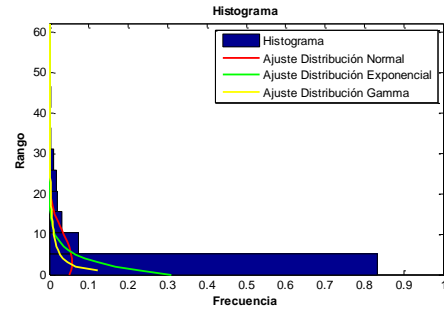
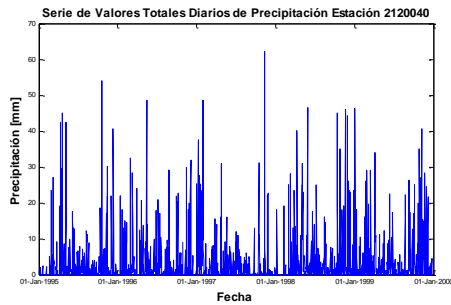


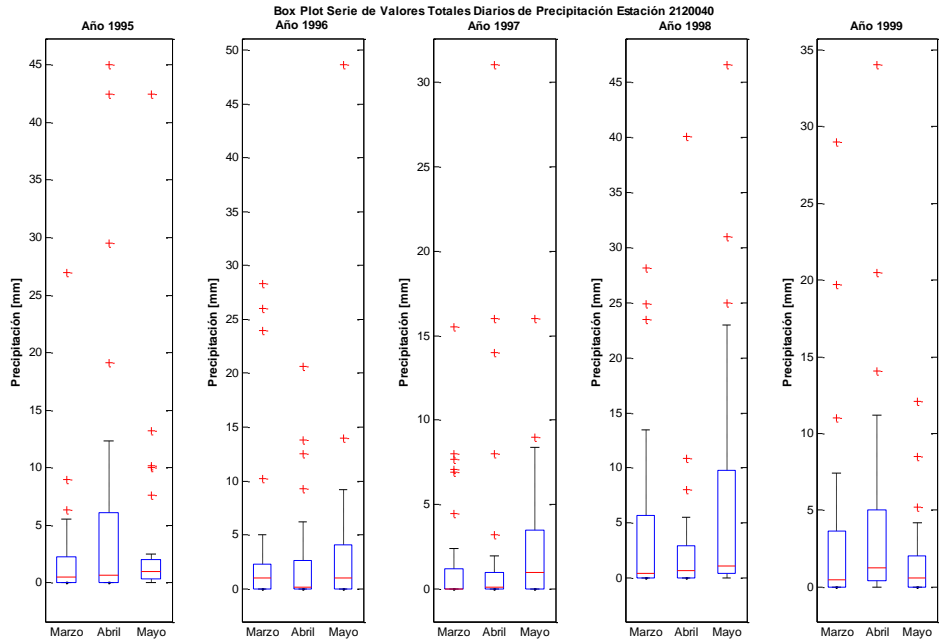
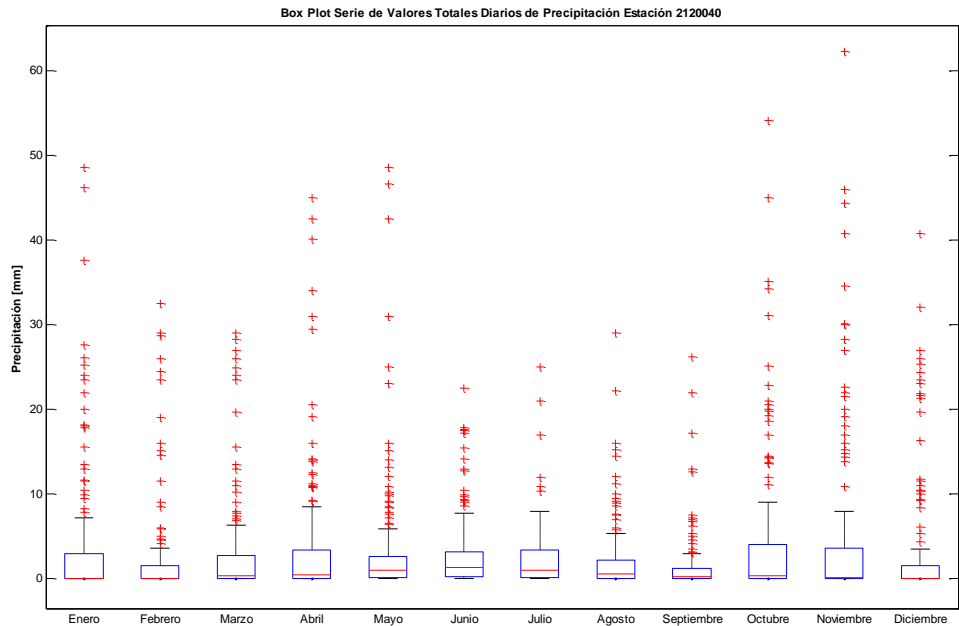


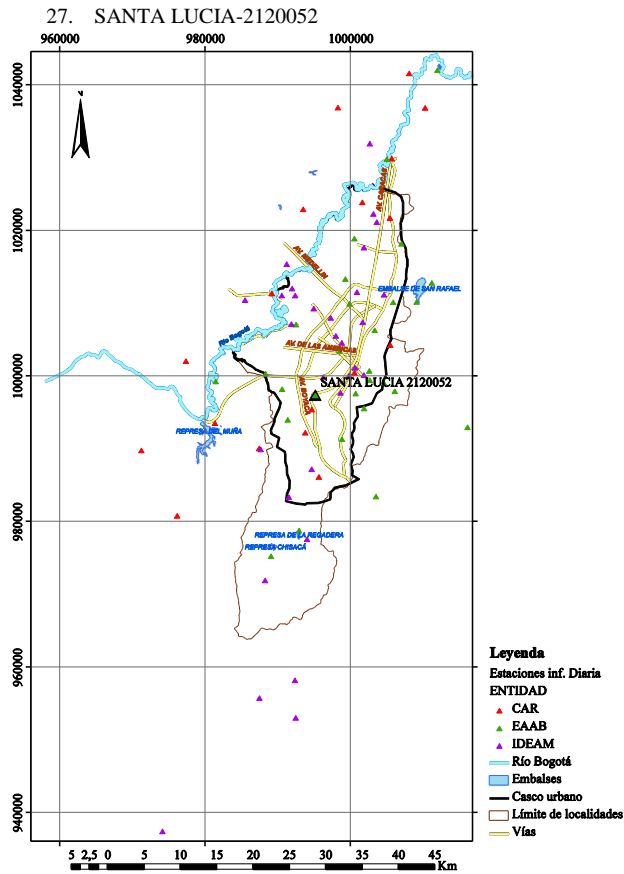
26. SAN LUIS-2120040



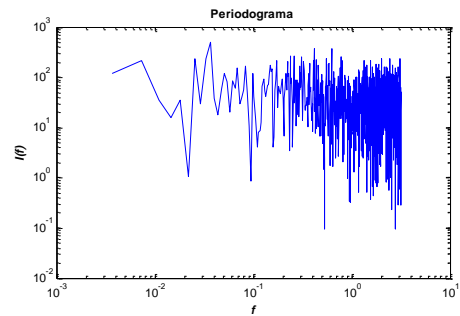
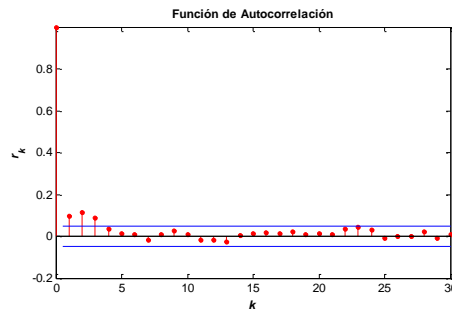
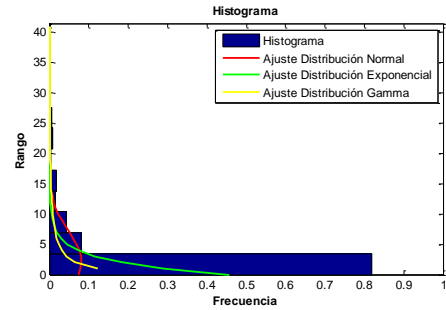
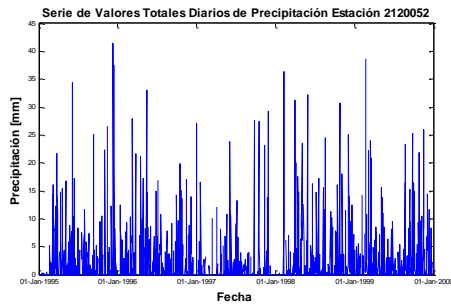
Este	1003254
Norte	1006379
Cota	2300
Número de Datos	1825
Número de Datos Sin Registro	1
Media	3.2
Mediana	0.4
Máximo	62.2
Desviación Estándar	7.0
Varianza	48.9
Coefficiente de Variación	1.4E-03
Coefficiente de Asimetría	3.6
Coefficiente de Curtosis	18.6

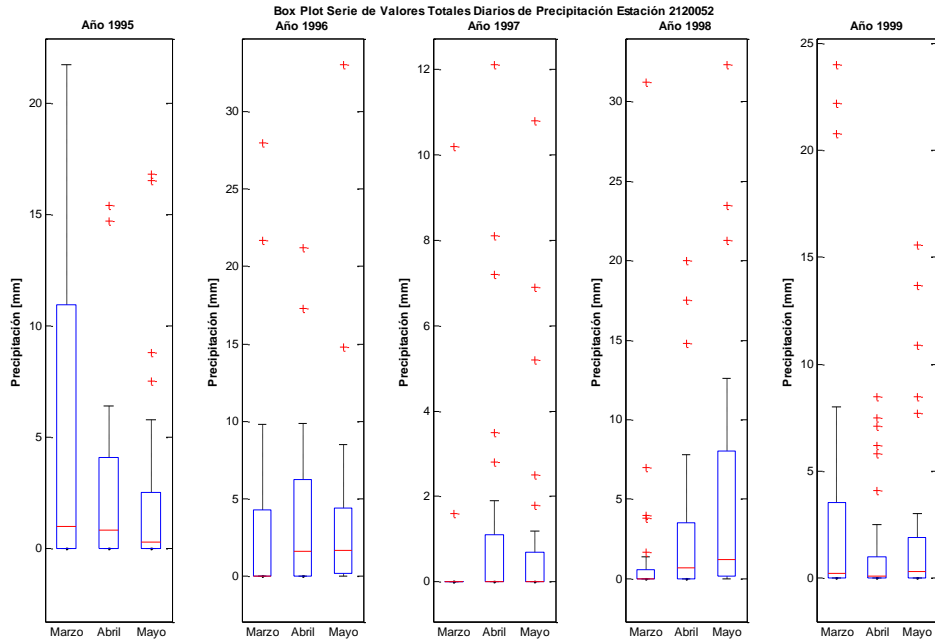
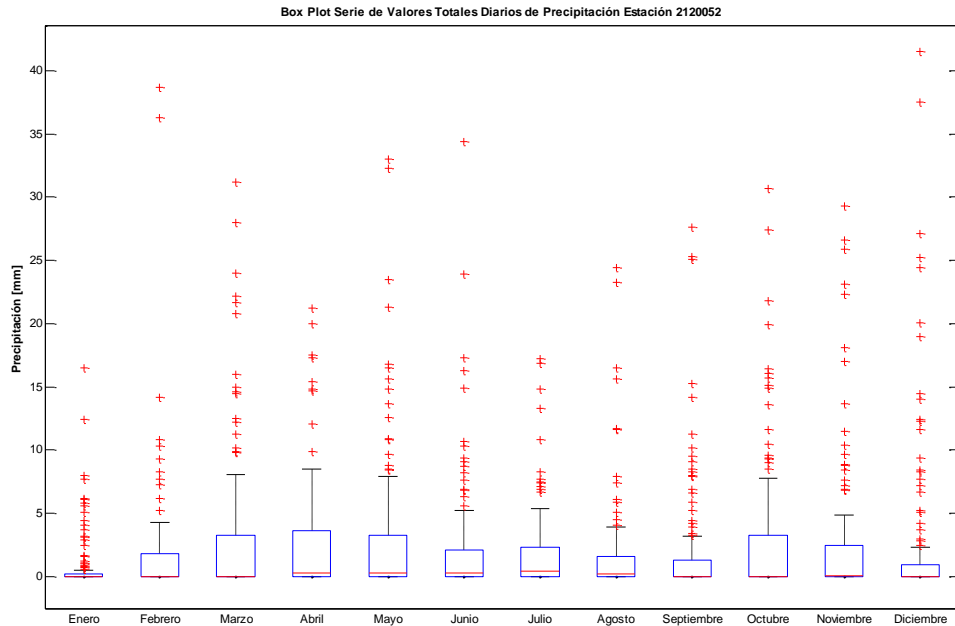


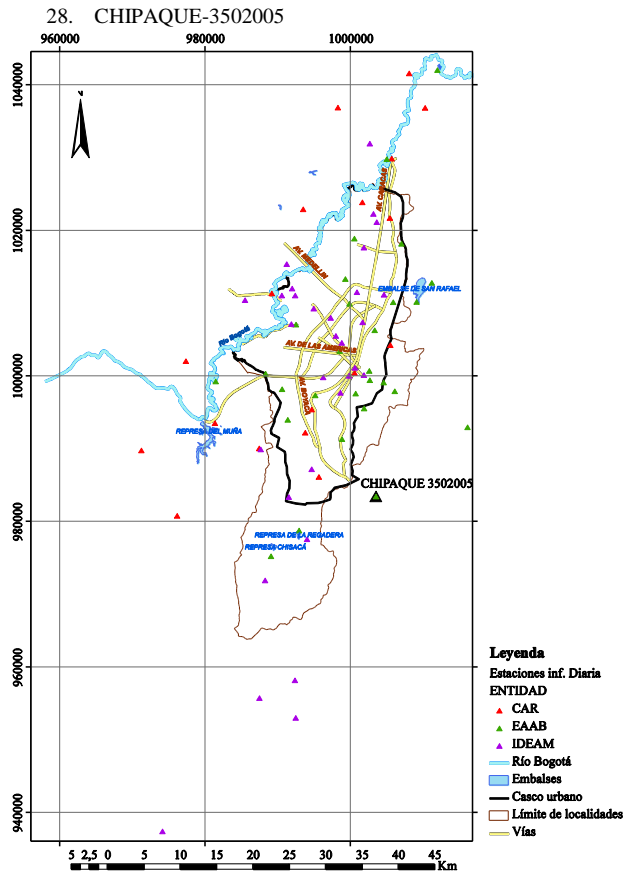




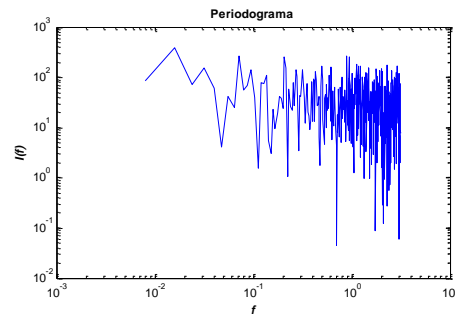
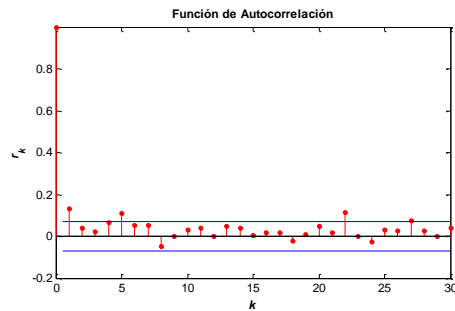
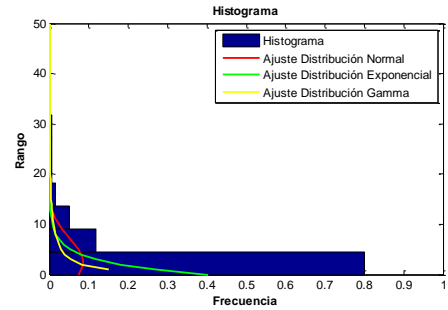
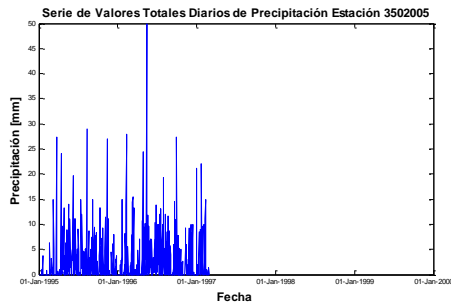
Este	995096
Norte	997414
Cota	2630
Número de Datos	1726
Número de Datos Sin Registro	100
Media	2.2
Mediana	0.1
Máximo	41.5
Desviación Estándar	4.9
Varianza	24.1
Coefficiente de Variación	8.3E-04
Coefficiente de Asimetría	3.7
Coefficiente de Curtosis	19.5

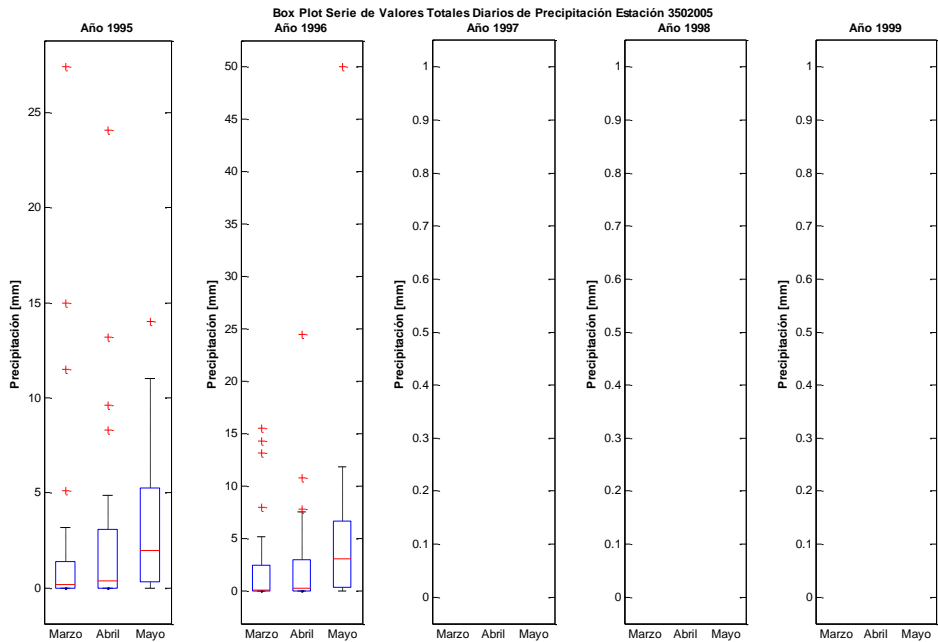
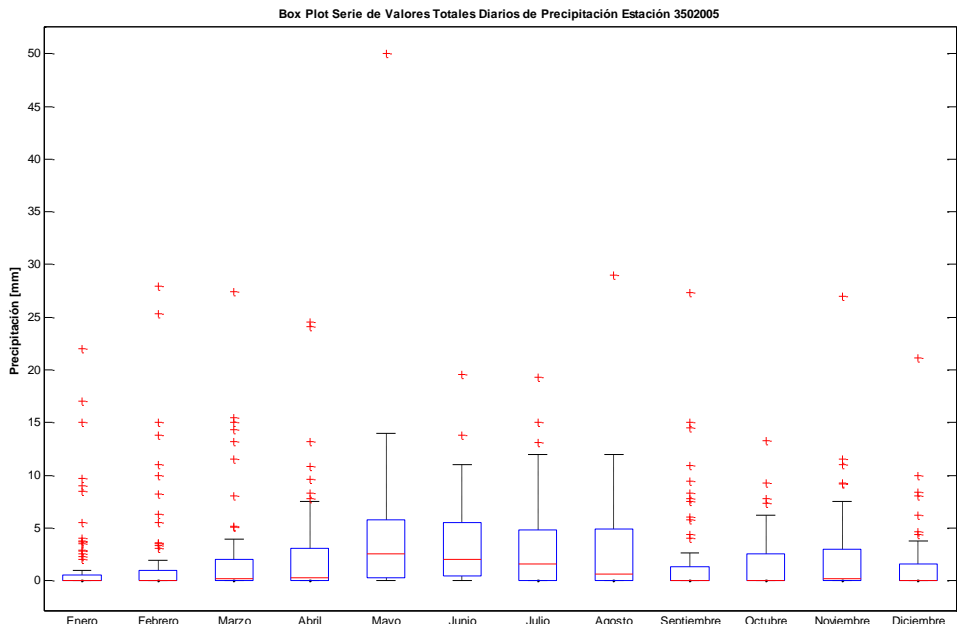


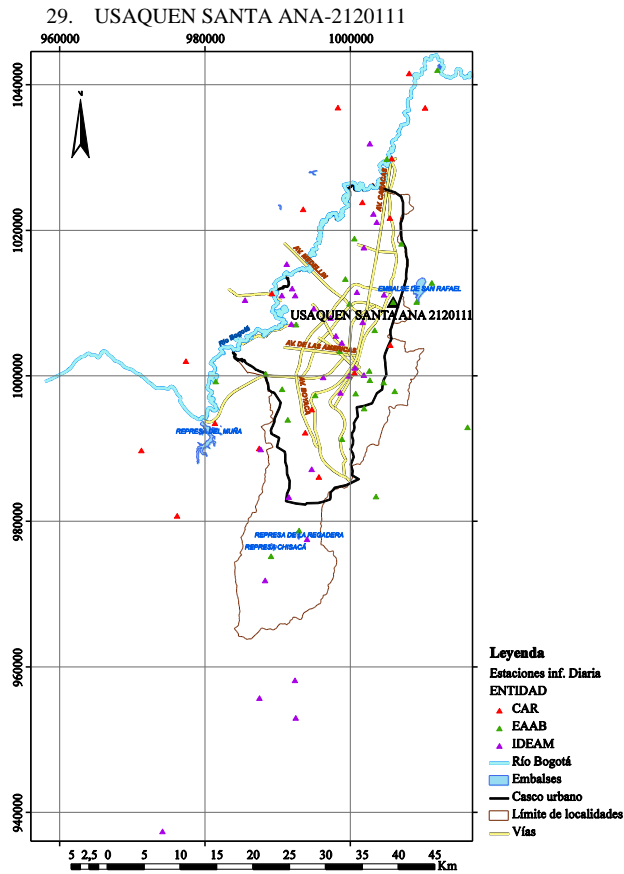




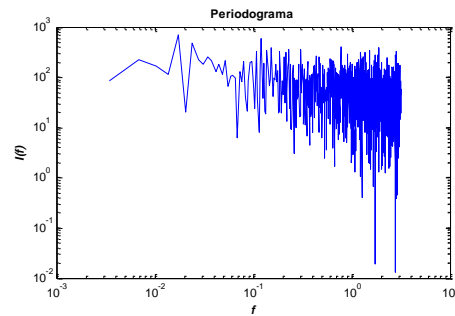
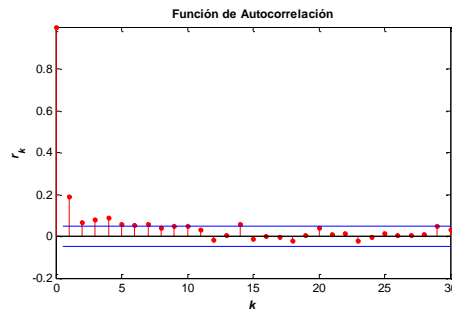
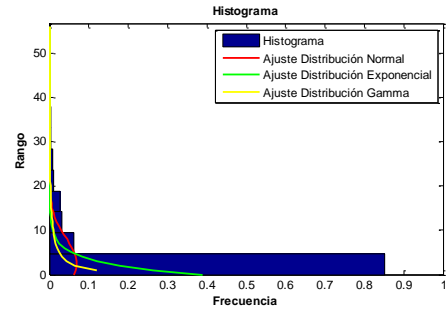
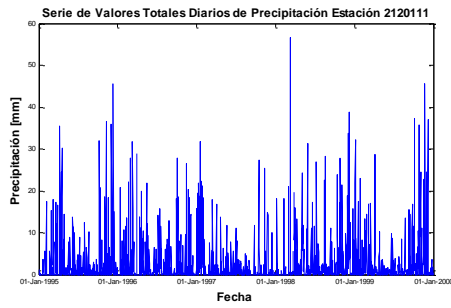
Este	1003431
Norte	983518
Cota	2850
Número de Datos	788
Número de Datos Sin Registro	1038
Media	2.5
Mediana	0.3
Máximo	50.0
Desviación Estándar	4.7
Varianza	21.9
Coefficiente de Variación	8.7E-04
Coefficiente de Asimetría	3.6
Coefficiente de Curtosis	23.5

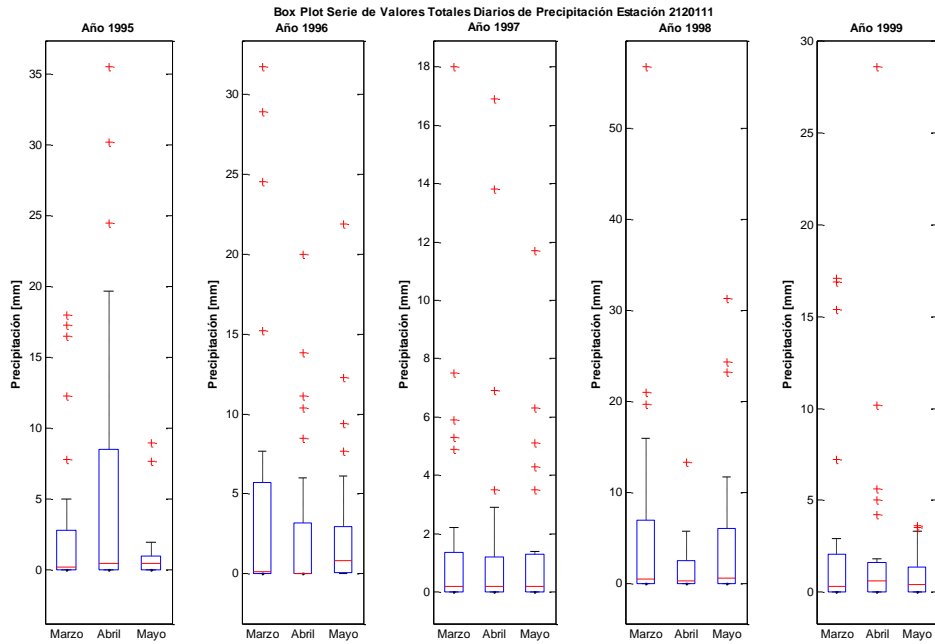
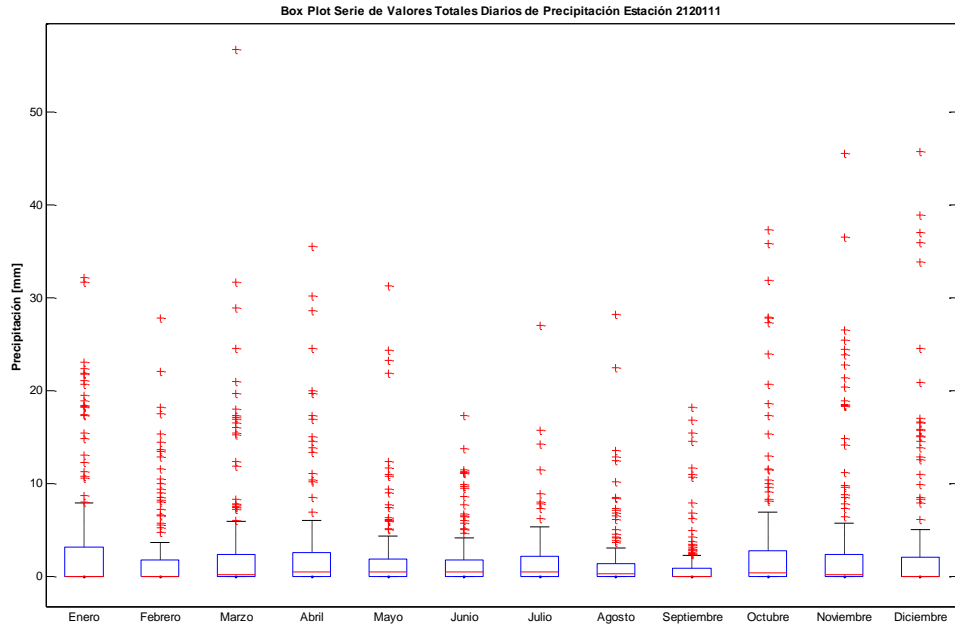


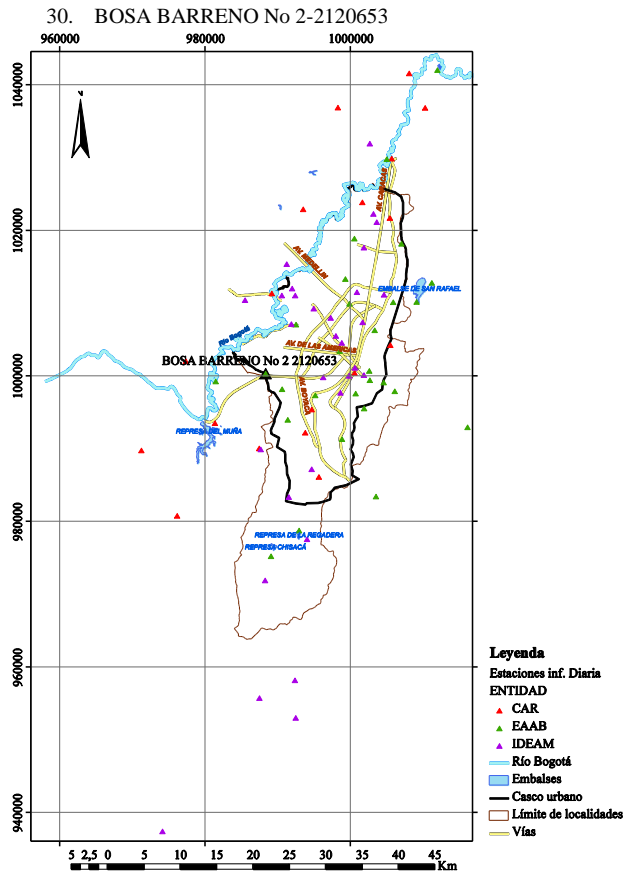




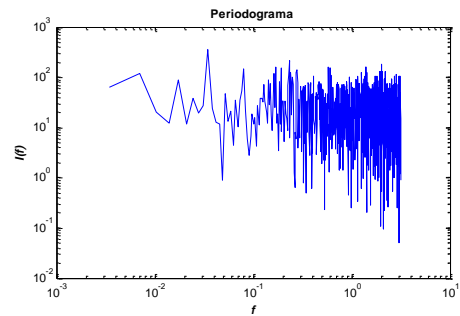
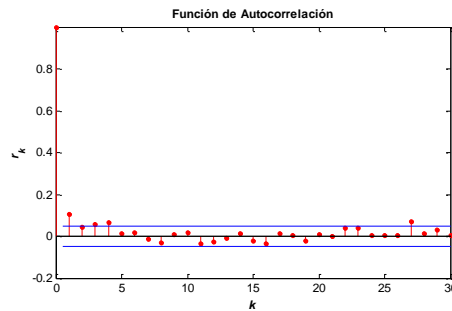
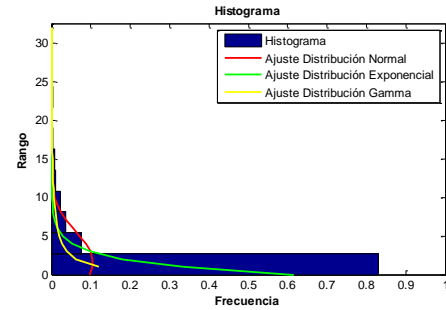
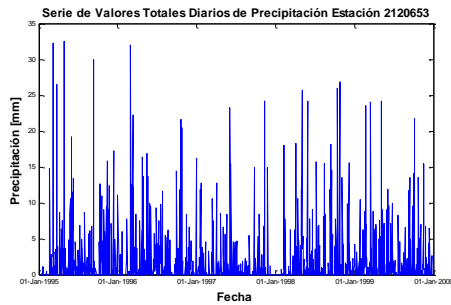
Este	1005805
Norte	1010244
Cota	2647
Número de Datos	1821
Número de Datos Sin Registro	5
Media	2.6
Mediana	0.2
Máximo	56.7
Desviación Estándar	5.8
Varianza	33.7
Coefficiente de Variación	9.7E-04
Coefficiente de Asimetría	3.7
Coefficiente de Curtosis	19.7

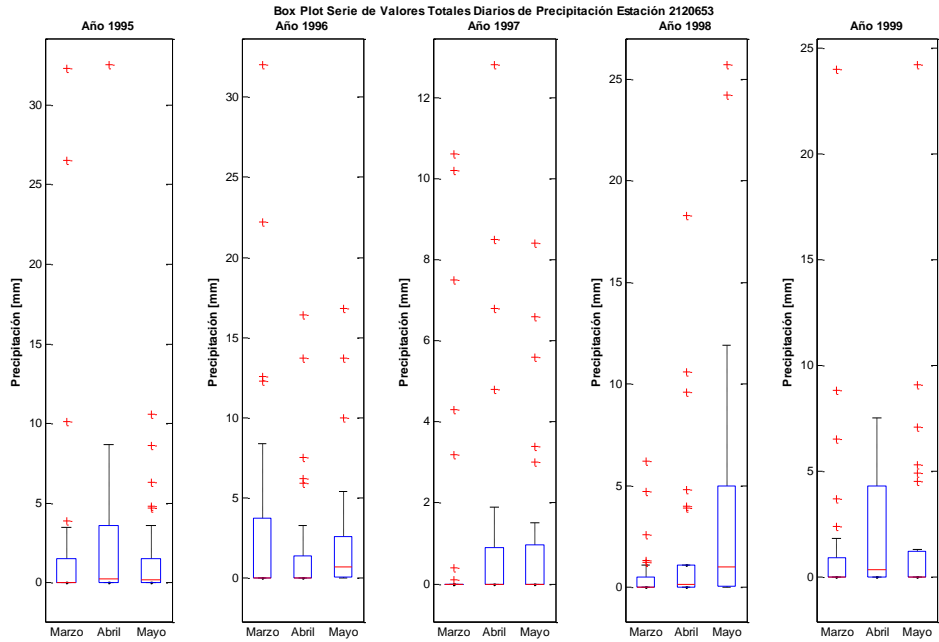
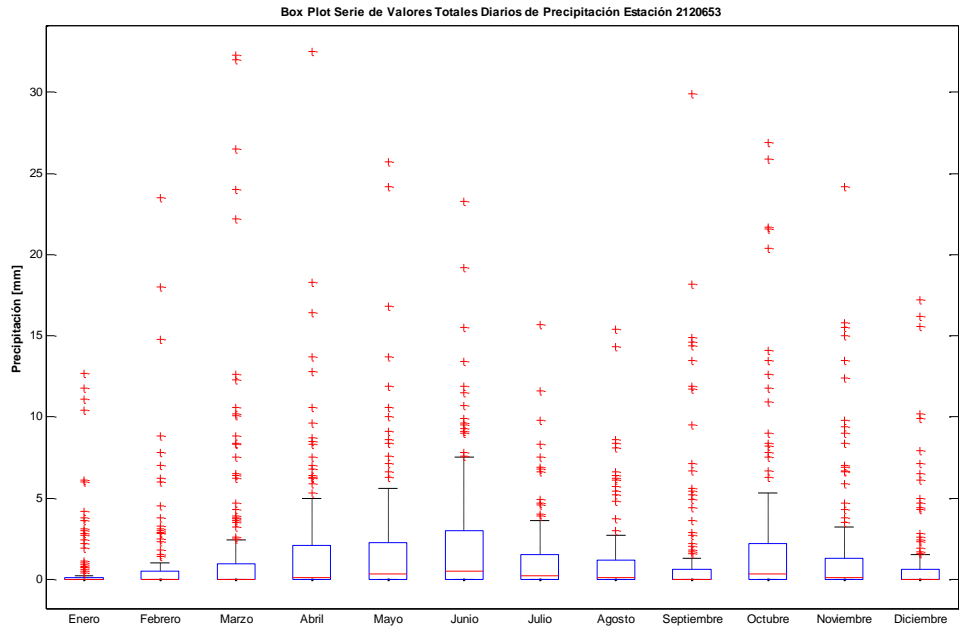


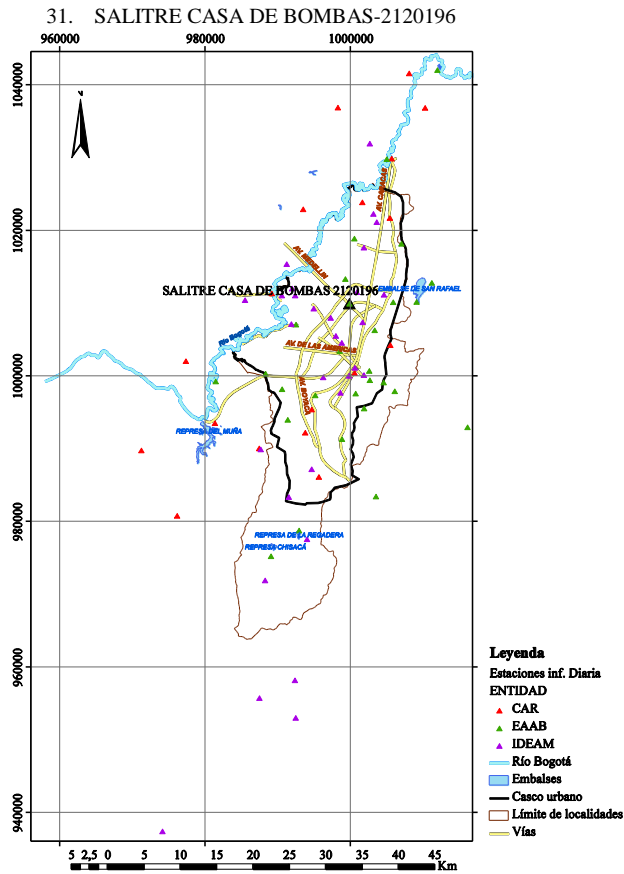




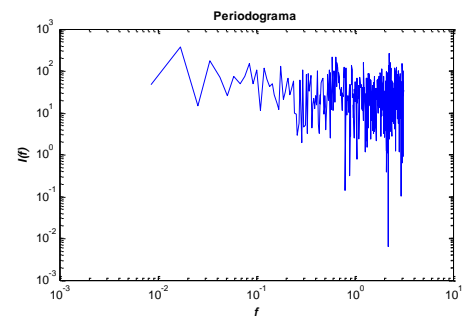
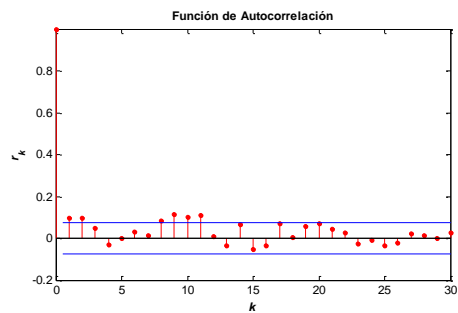
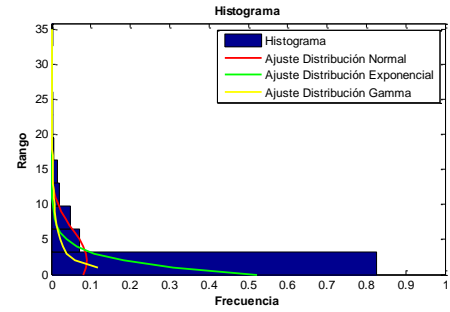
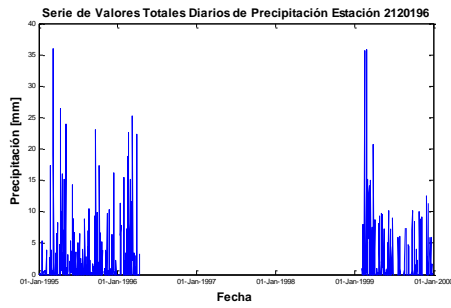
Este	988257
Norte	1000417
Cota	2550
Número de Datos	1808
Número de Datos Sin Registro	18
Media	1.6
Mediana	0.0
Máximo	32.5
Desviación Estándar	3.8
Varianza	14.3
Coefficiente de Variación	6.4E-04
Coefficiente de Asimetría	4.0
Coefficiente de Curtosis	22.7

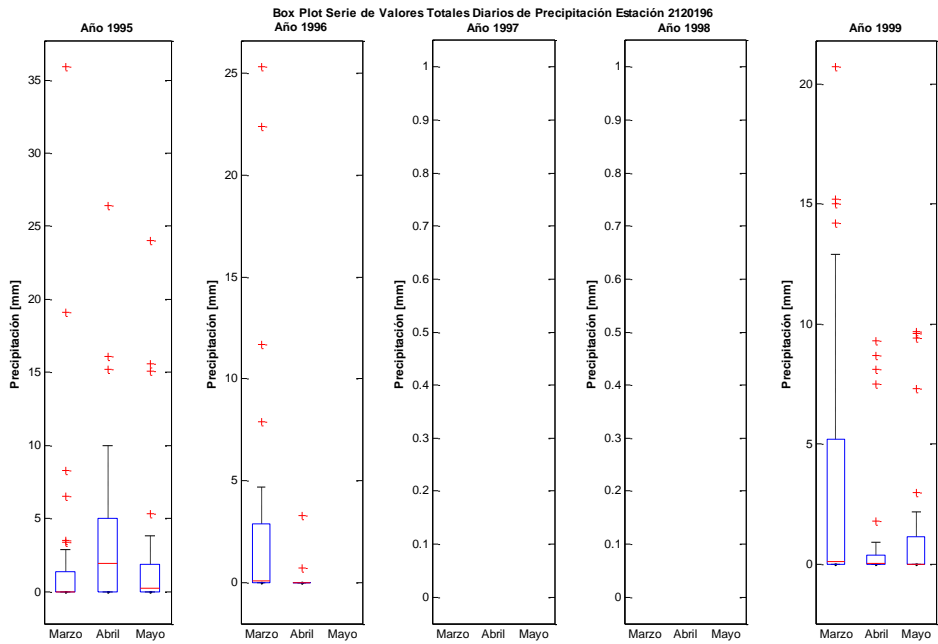
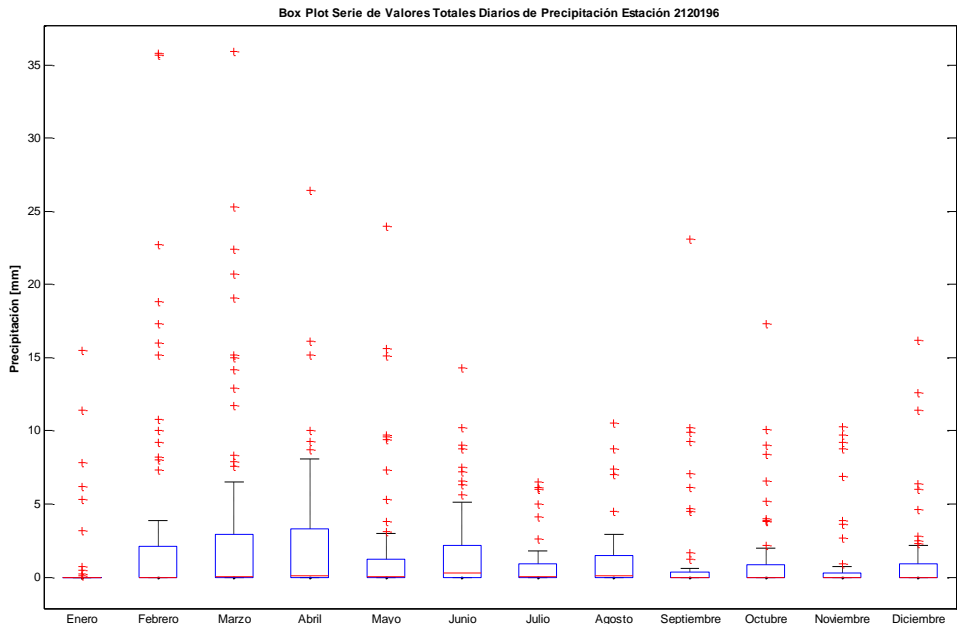


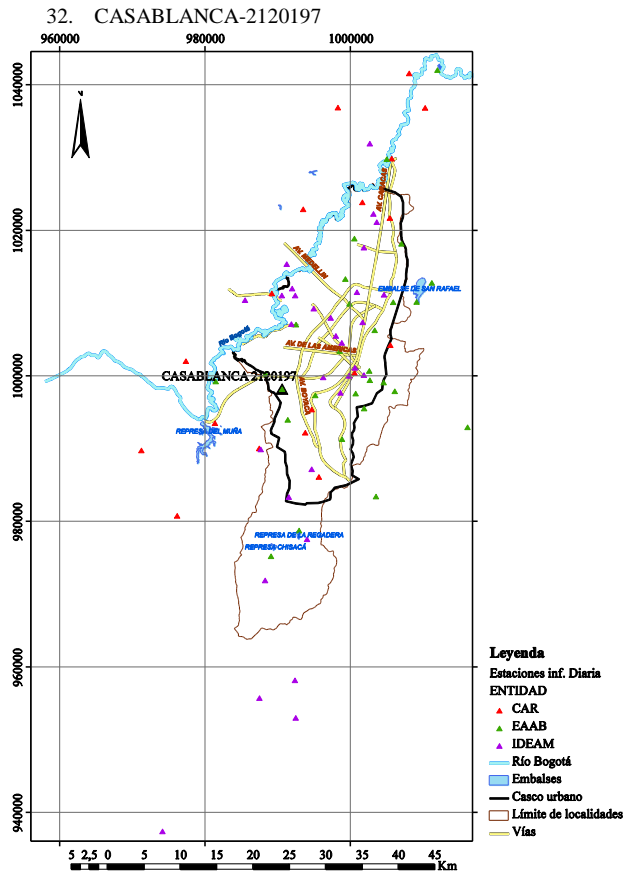




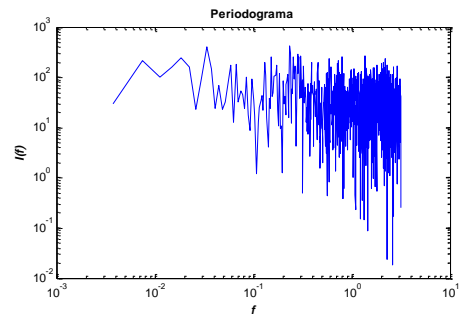
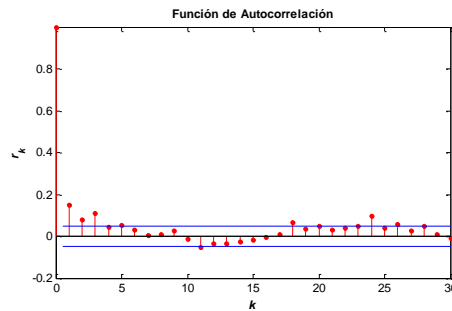
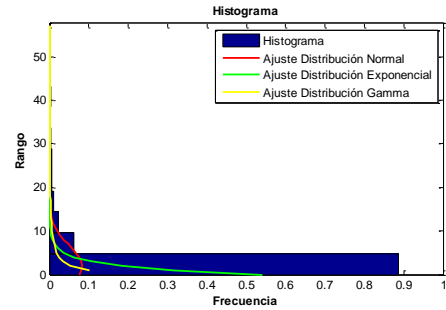
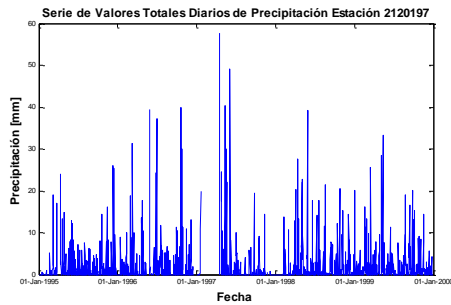
Este	999780
Norte	1010008
Cota	2580
Número de Datos	740
Número de Datos Sin Registro	1086
Media	1.9
Mediana	0.0
Máximo	35.9
Desviación Estándar	4.5
Varianza	20.0
Coefficiente de Variación	7.4E-04
Coefficiente de Asimetría	3.8
Coefficiente de Curtosis	21.7

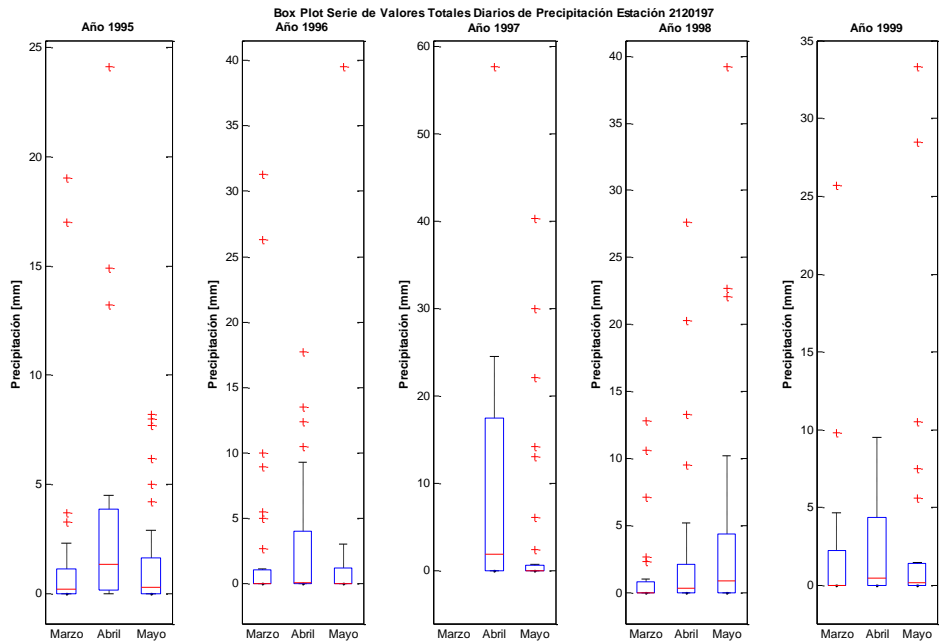
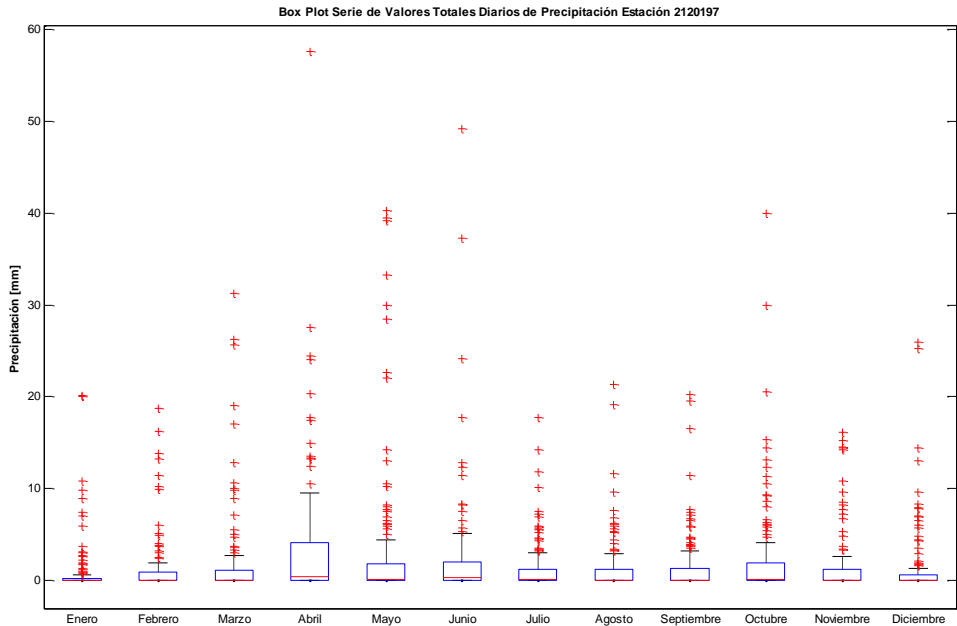




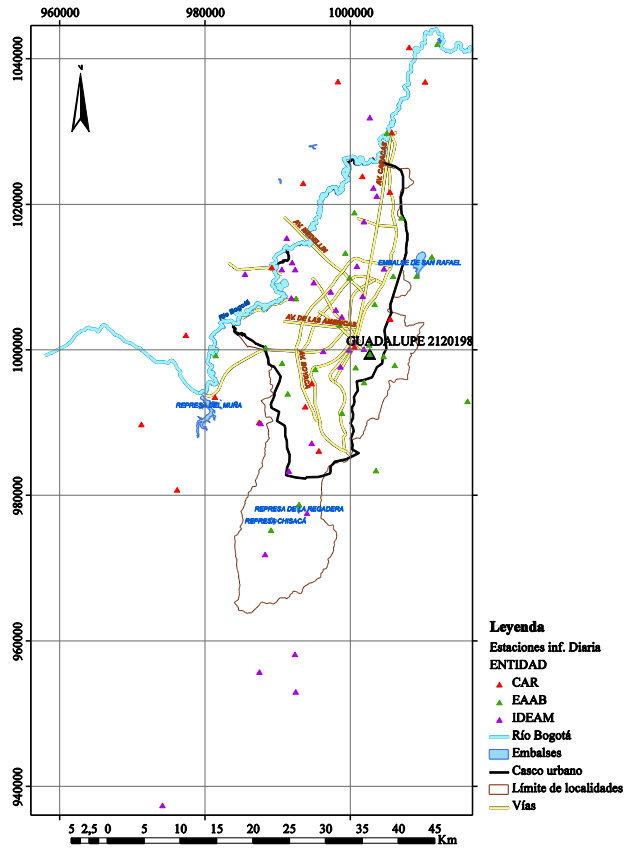


Este	990520
Norte	998273
Cota	2665
Número de Datos	1686
Número de Datos Sin Registro	140
Media	1.8
Mediana	0.0
Máximo	57.6
Desviación Estándar	4.8
Varianza	23.1
Coefficiente de Variación	6.9E-04
Coefficiente de Asimetría	5.0
Coefficiente de Curtosis	36.9

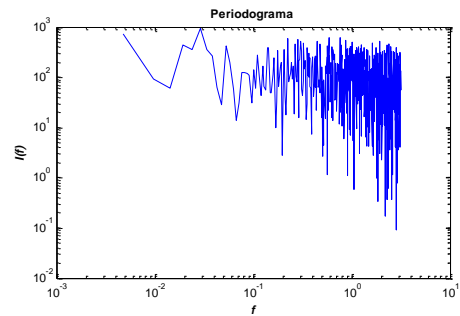
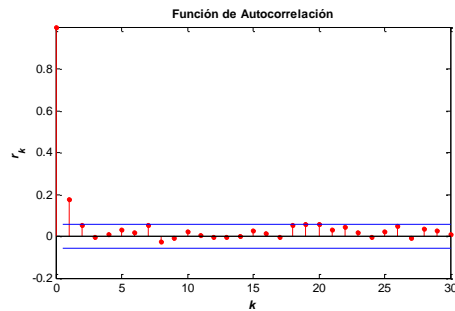
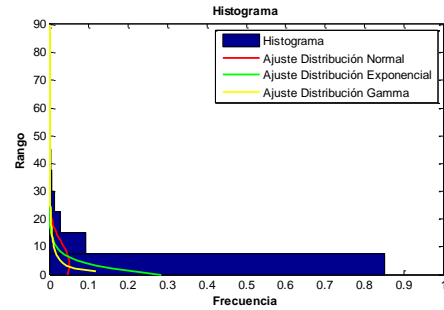
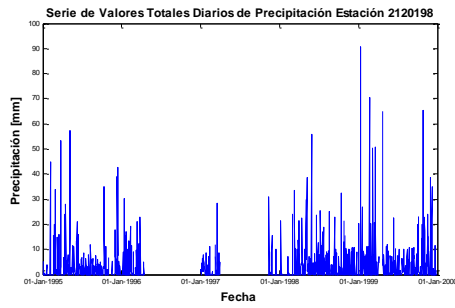


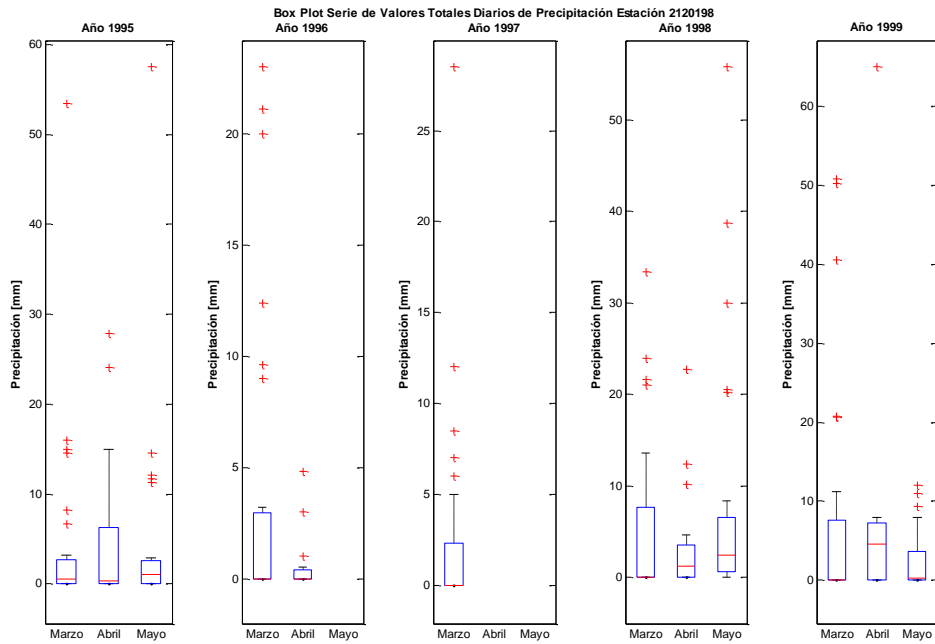
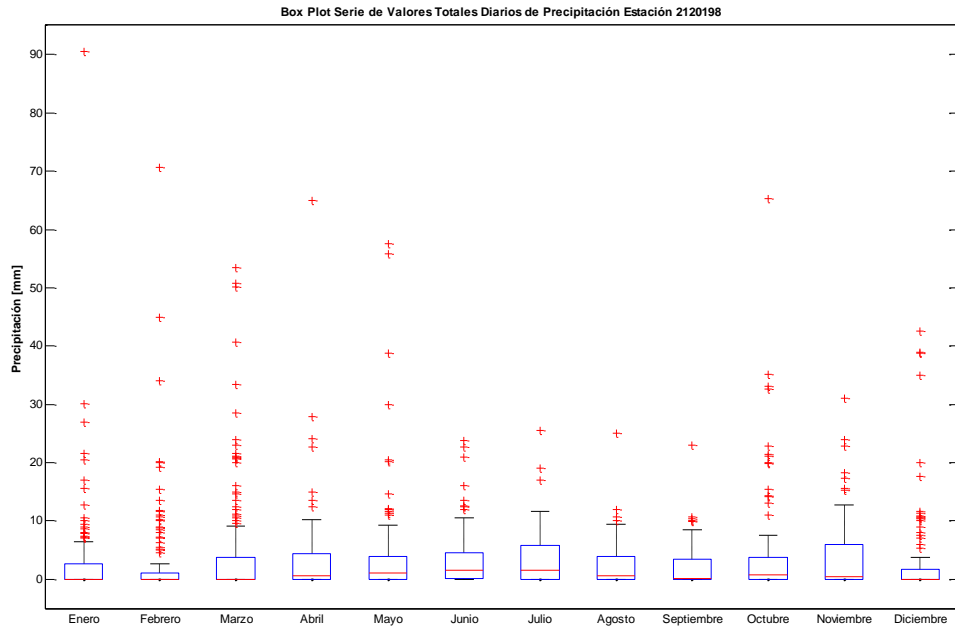


33. GUADALUPE-2120198

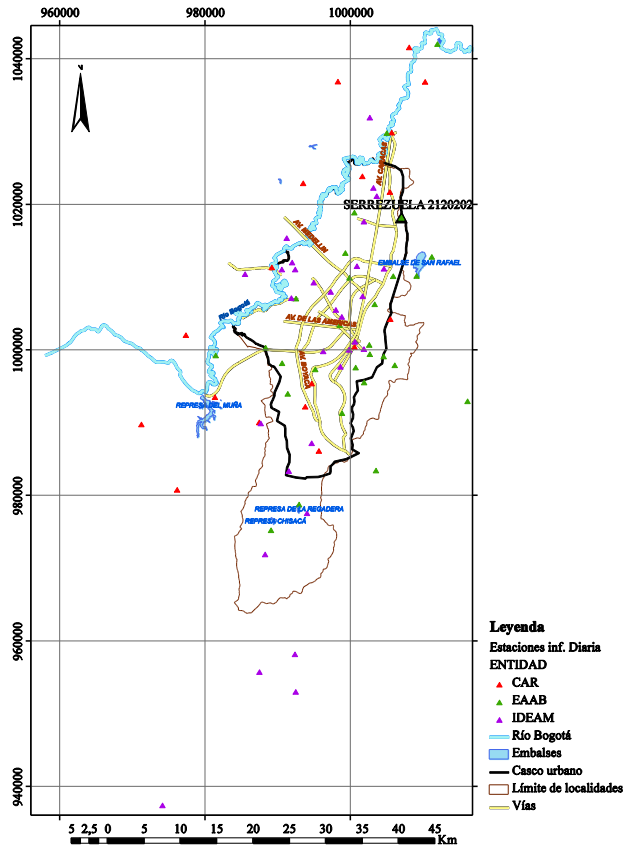


Este	1002583
Norte	999527
Cota	3316
Número de Datos	1311
Número de Datos Sin Registro	515
Media	3.5
Mediana	0.1
Máximo	90.5
Desviación Estándar	7.8
Varianza	60.6
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	4.7
Coefficiente de Curtosis	33.8

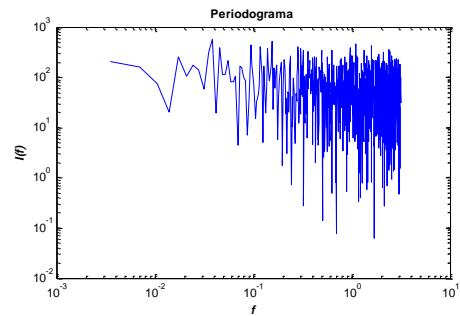
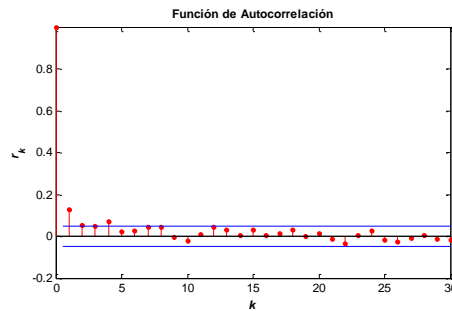
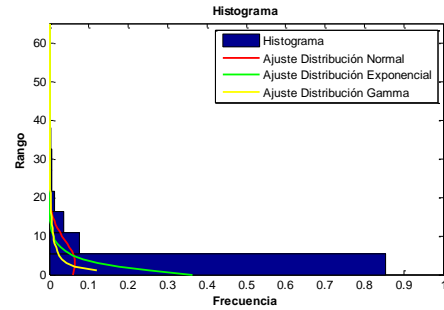
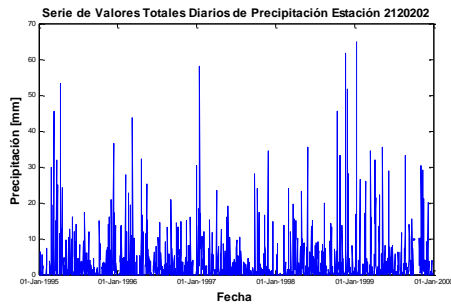


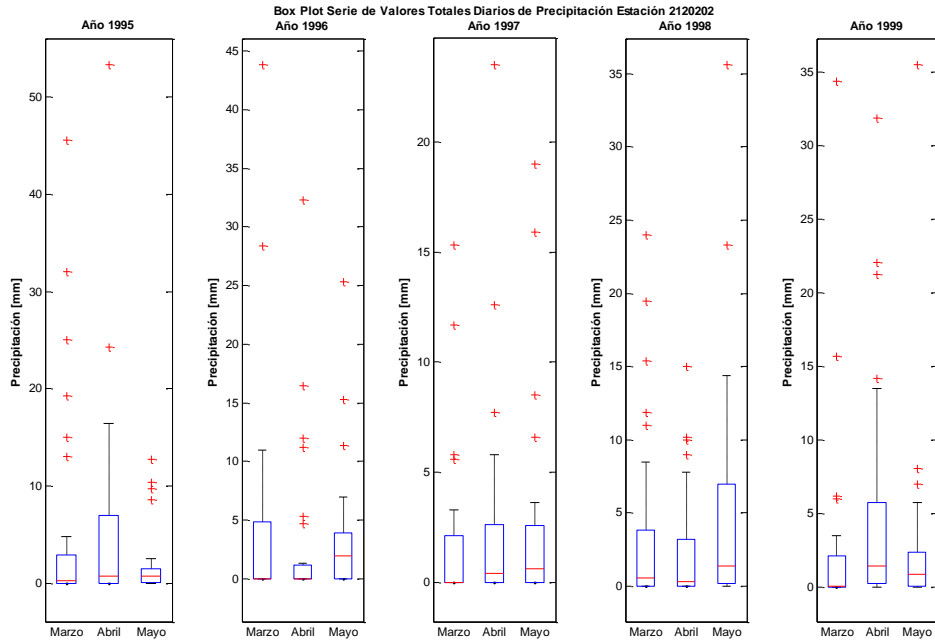
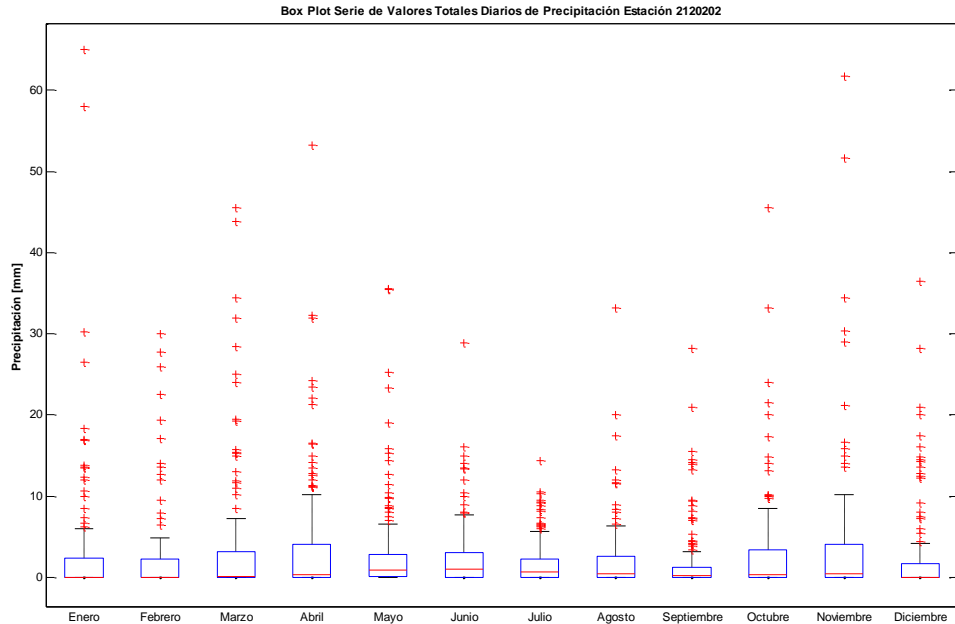


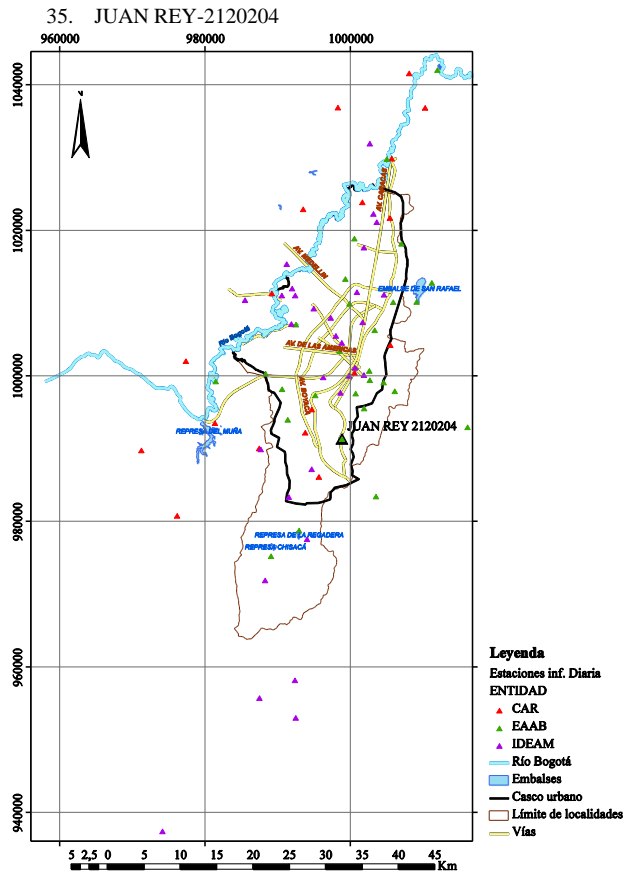
34. SERREZUELA-2120202



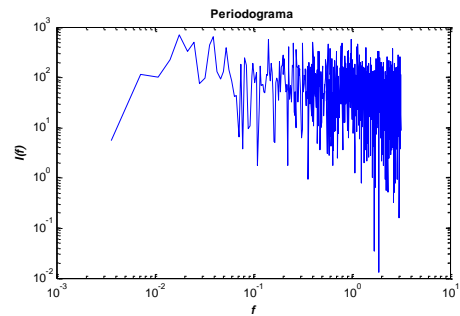
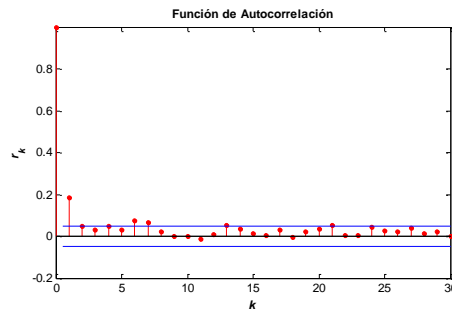
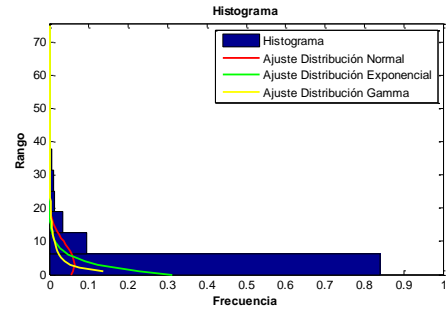
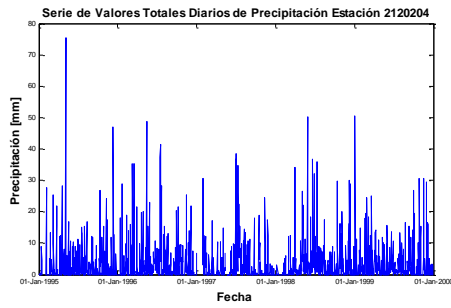
Este	1006941
Norte	1018259
Cota	2800
Número de Datos	1802
Número de Datos Sin Registro	24
Media	2.8
Mediana	0.3
Máximo	65.0
Desviación Estándar	6.1
Varianza	37.3
Coefficiente de Variación	9.8E-04
Coefficiente de Asimetría	4.5
Coefficiente de Curtosis	30.9

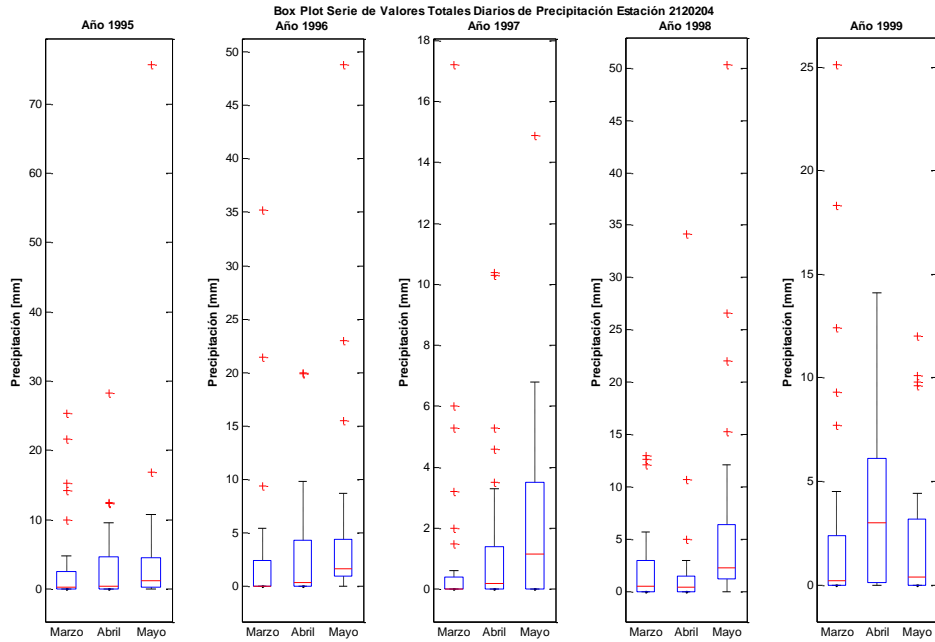
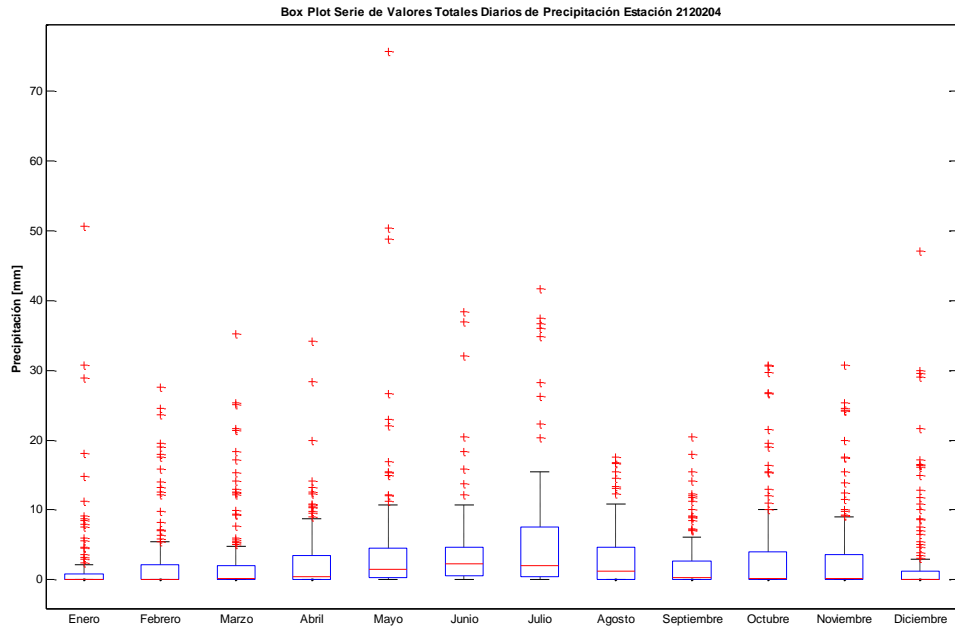


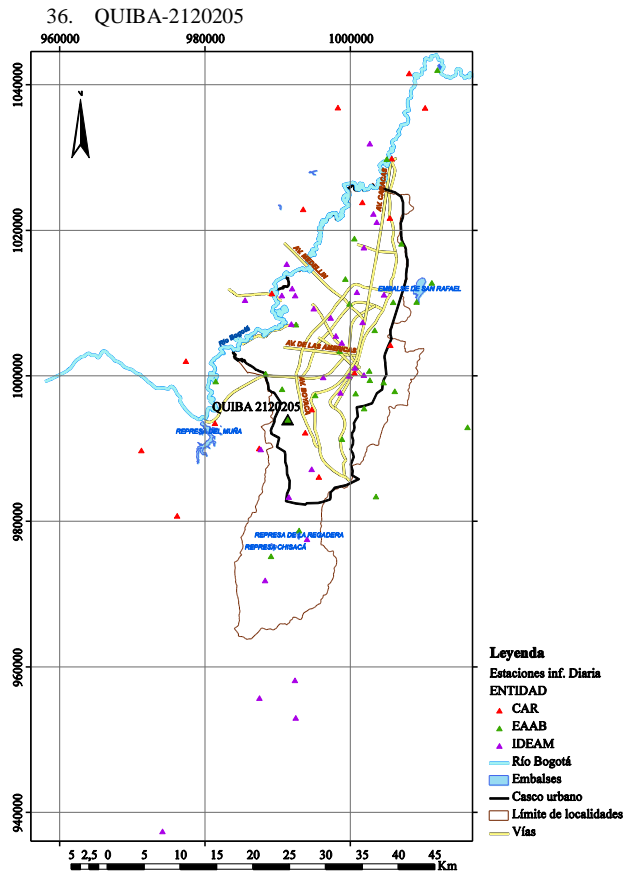




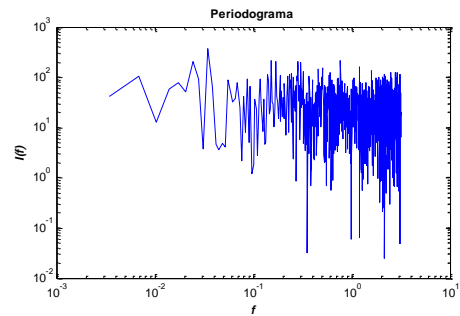
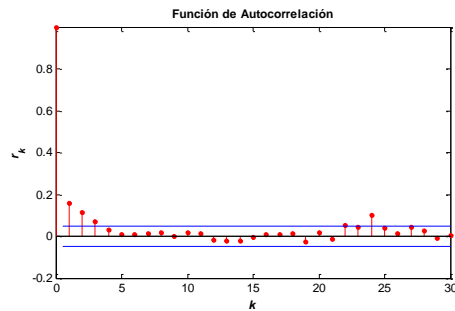
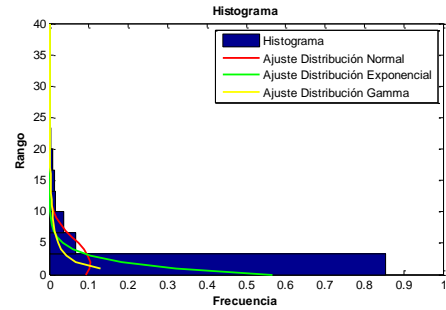
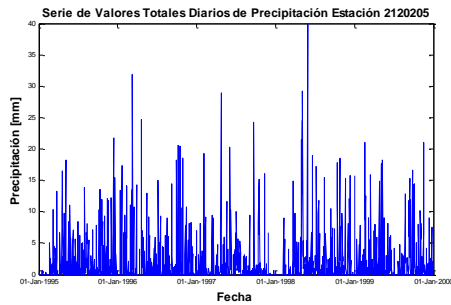
Este	998762
Norte	991429
Cota	2985
Número de Datos	1765
Número de Datos Sin Registro	61
Media	3.2
Mediana	0.5
Máximo	75.6
Desviación Estándar	6.3
Varianza	39.8
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	25.3

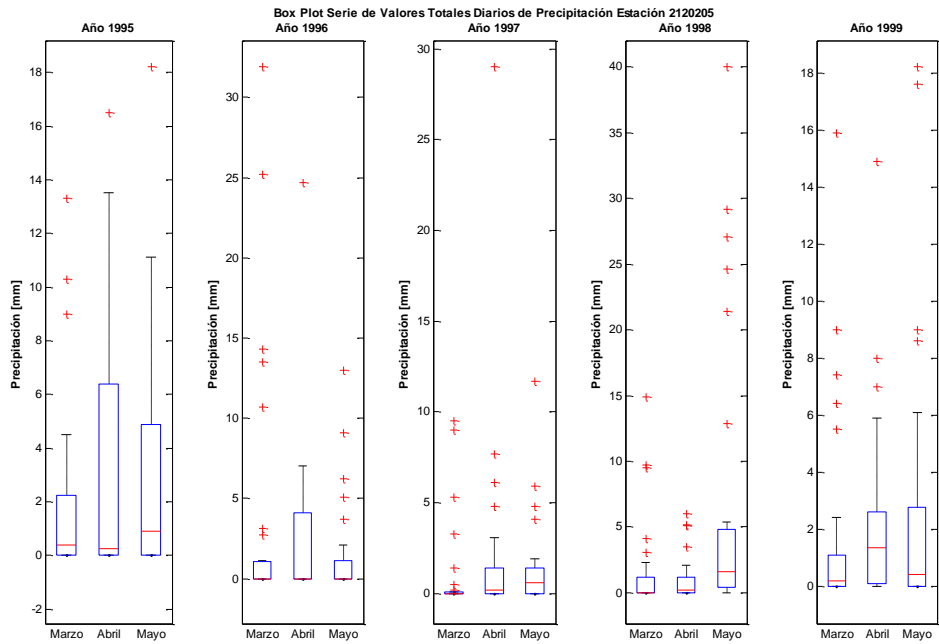
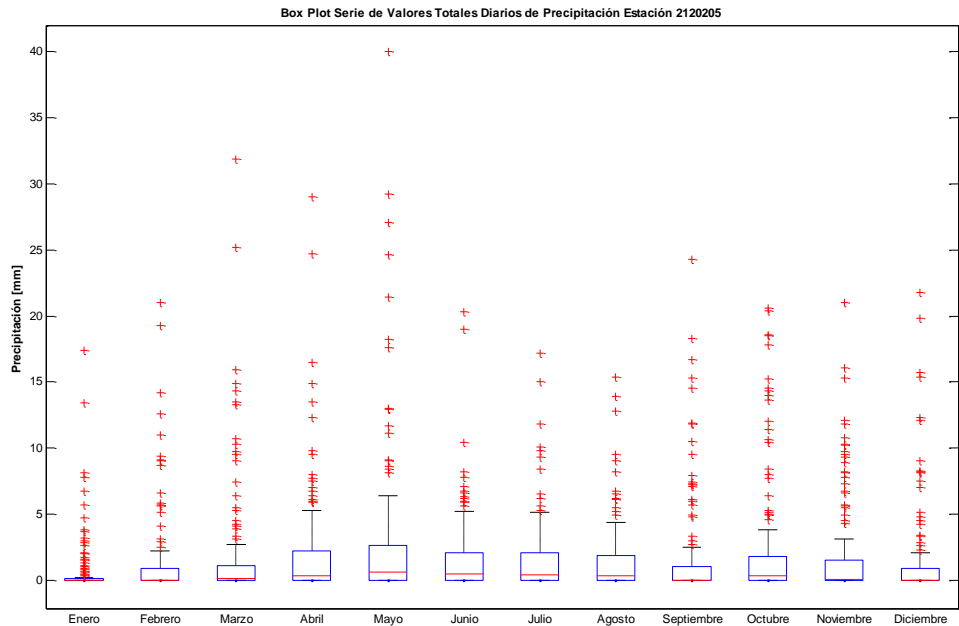




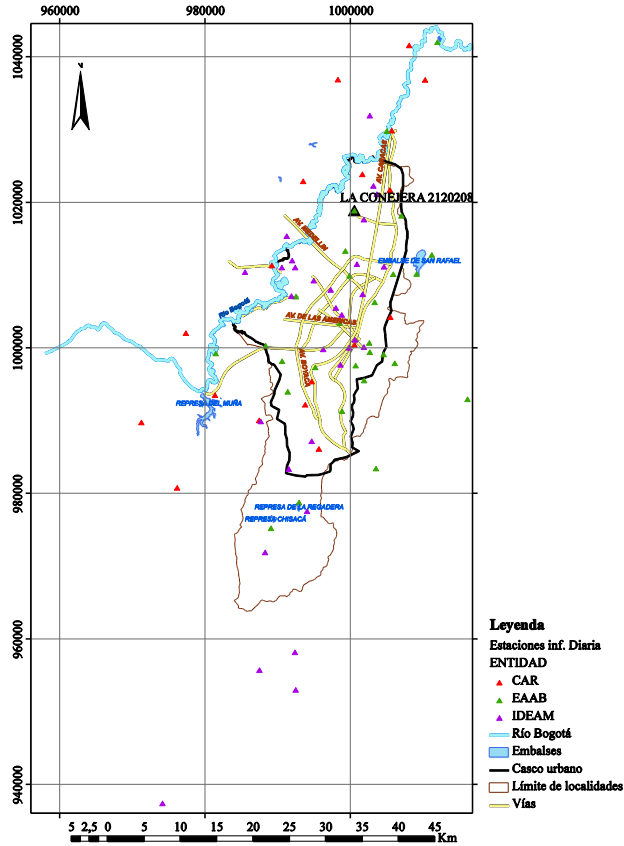


Este	991300
Norte	994082
Cota	3000
Número de Datos	1817
Número de Datos Sin Registro	9
Media	1.8
Mediana	0.1
Máximo	40.0
Desviación Estándar	3.9
Varianza	15.0
Coefficiente de Variación	5.9E-04
Coefficiente de Asimetría	3.8
Coefficiente de Curtosis	21.2

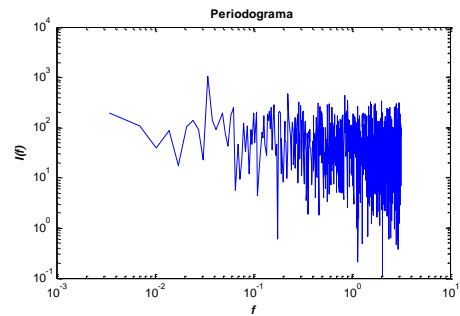
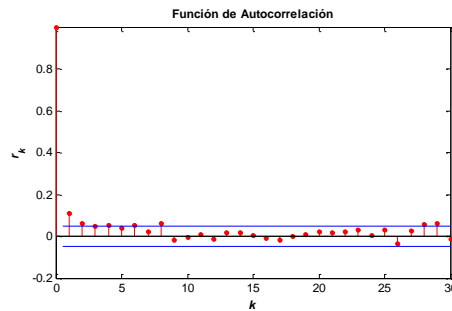
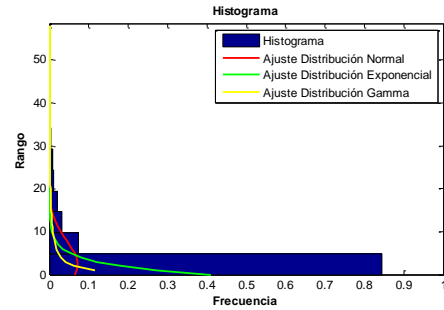
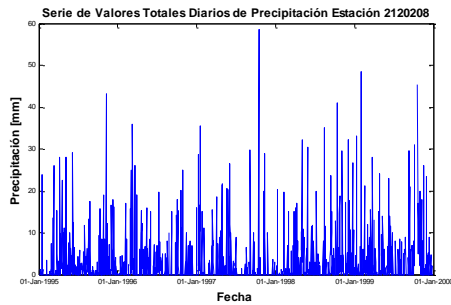


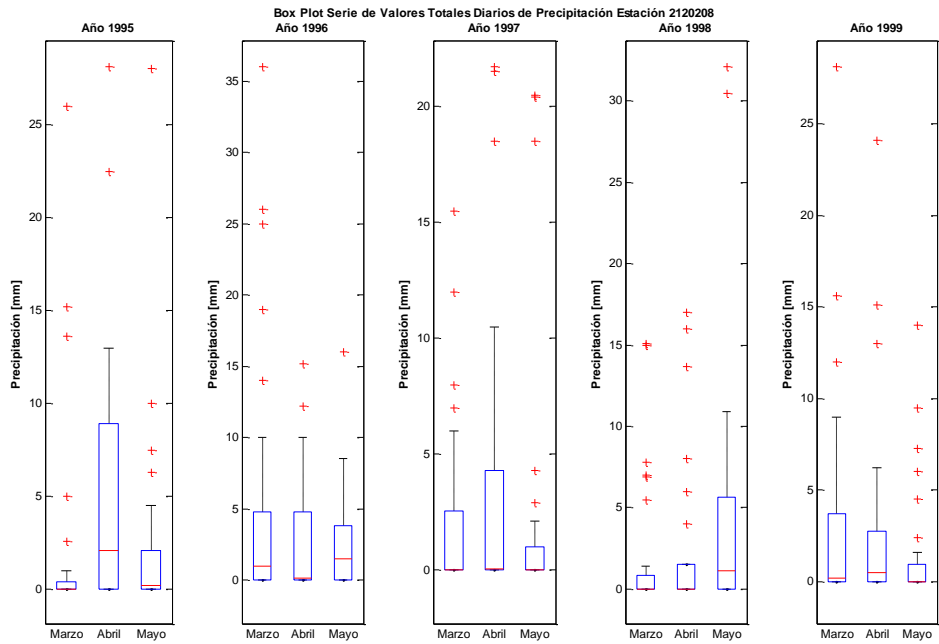
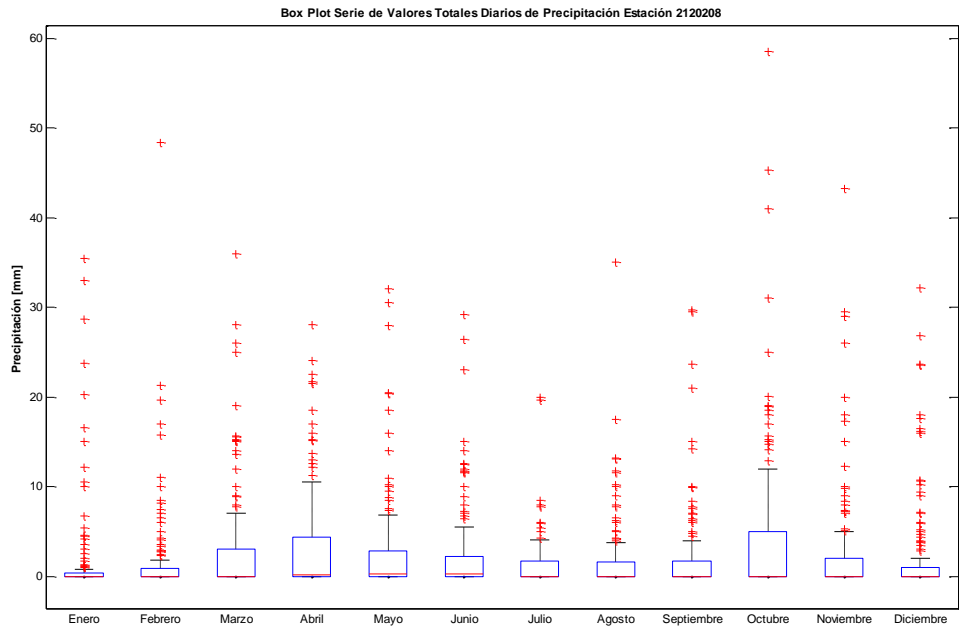


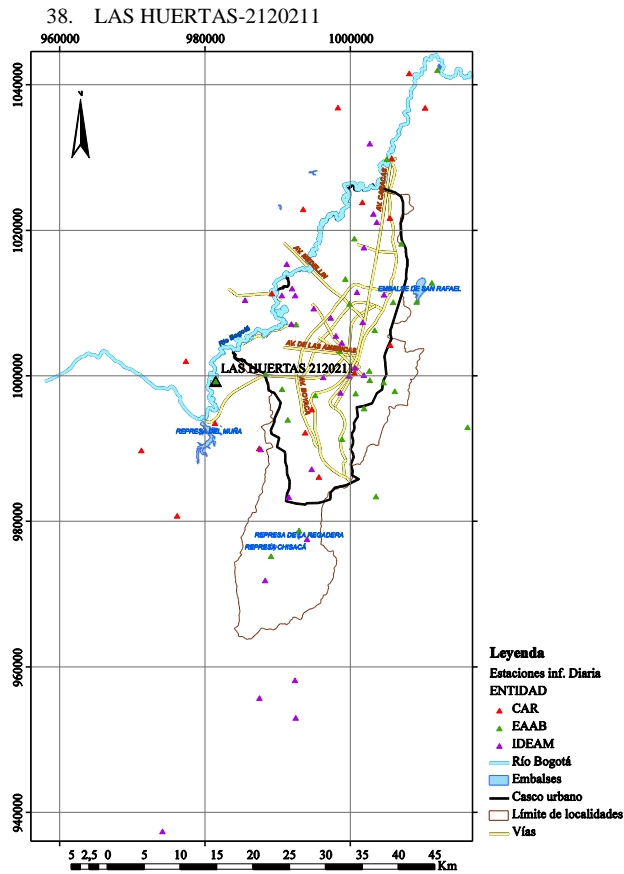
37. LA CONEJERA-2120208



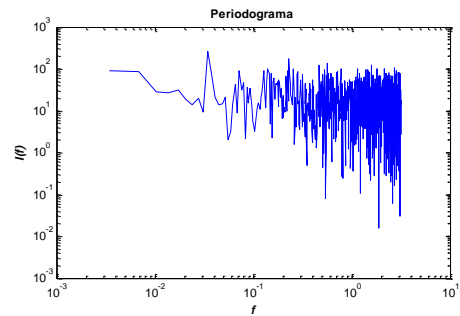
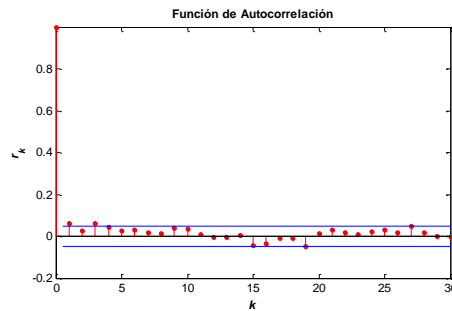
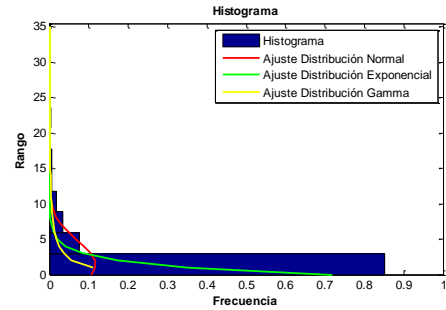
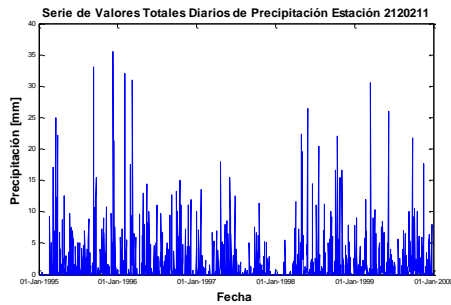
Este	1000443
Norte	1018986
Cota	2500
Número de Datos	1808
Número de Datos Sin Registro	18
Media	2.4
Mediana	0.0
Máximo	58.5
Desviación Estándar	5.7
Varianza	32.0
Coefficiente de Variación	9.8E-04
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	22.4

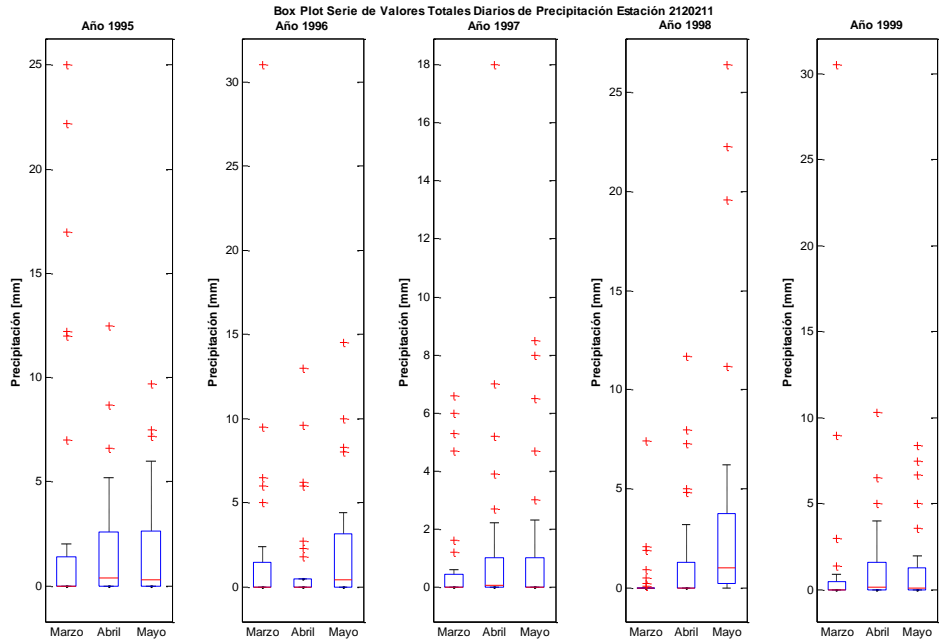
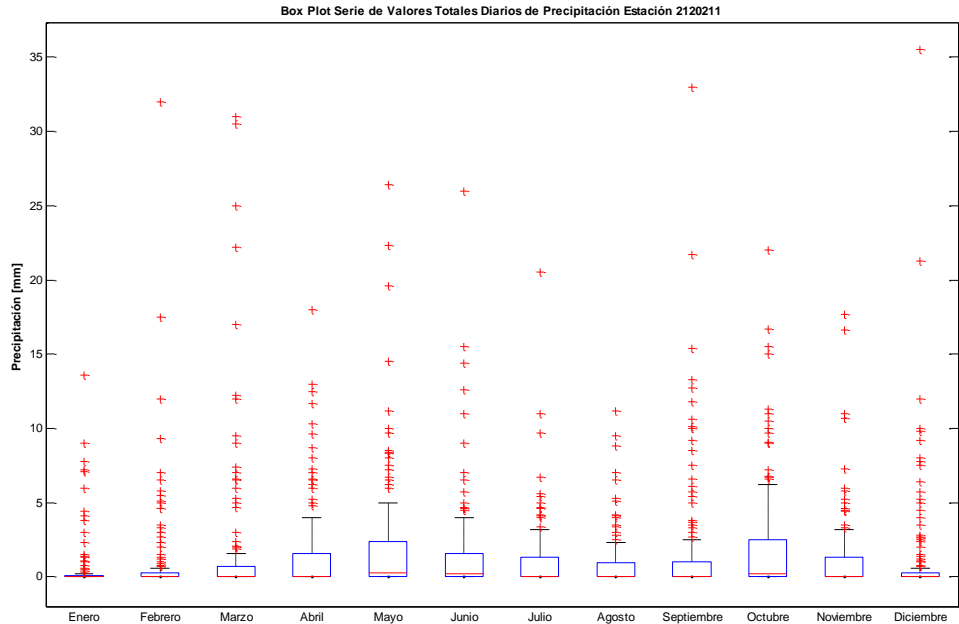


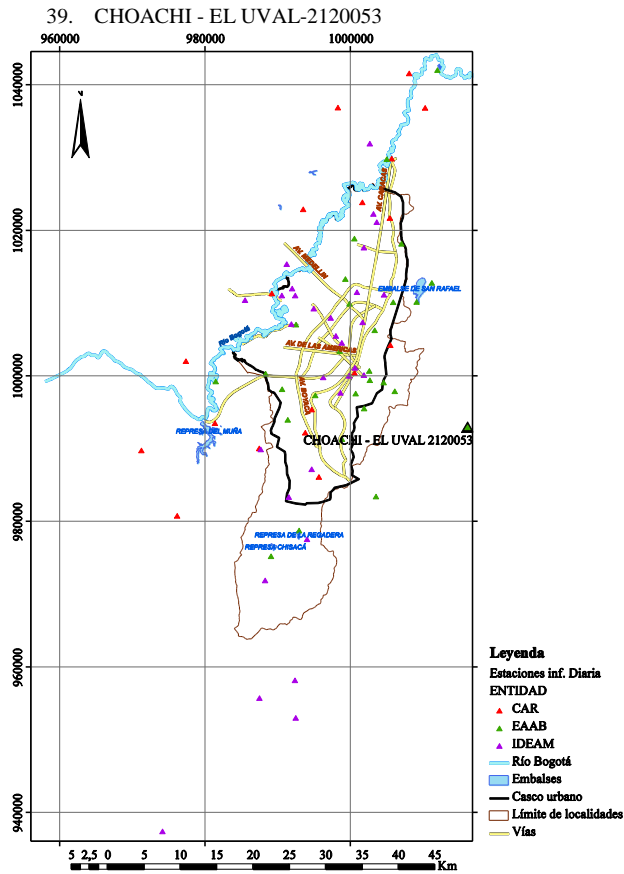




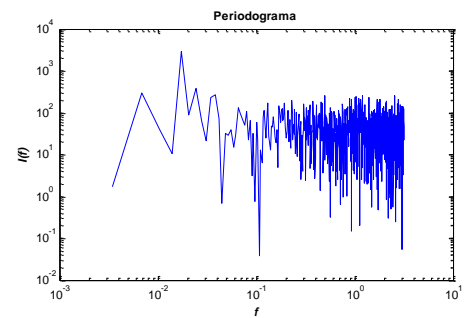
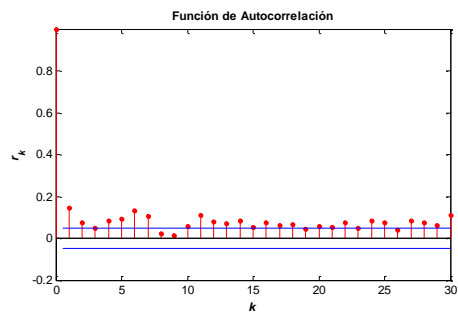
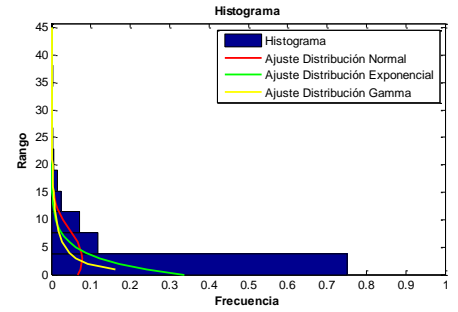
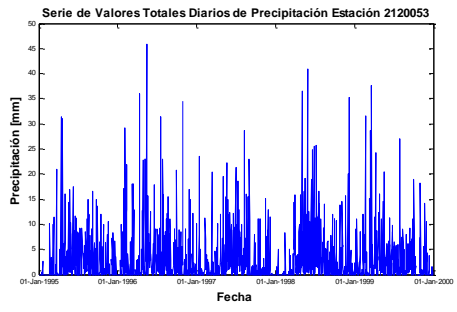
Este	981377
Norte	999344
Cota	2572
Número de Datos	1826
Número de Datos Sin Registro	0
Media	1.4
Mediana	0.0
Máximo	35.5
Desviación Estándar	3.4
Varianza	11.7
Coefficiente de Variación	5.4E-04
Coefficiente de Asimetría	4.6
Coefficiente de Curtosis	31.7

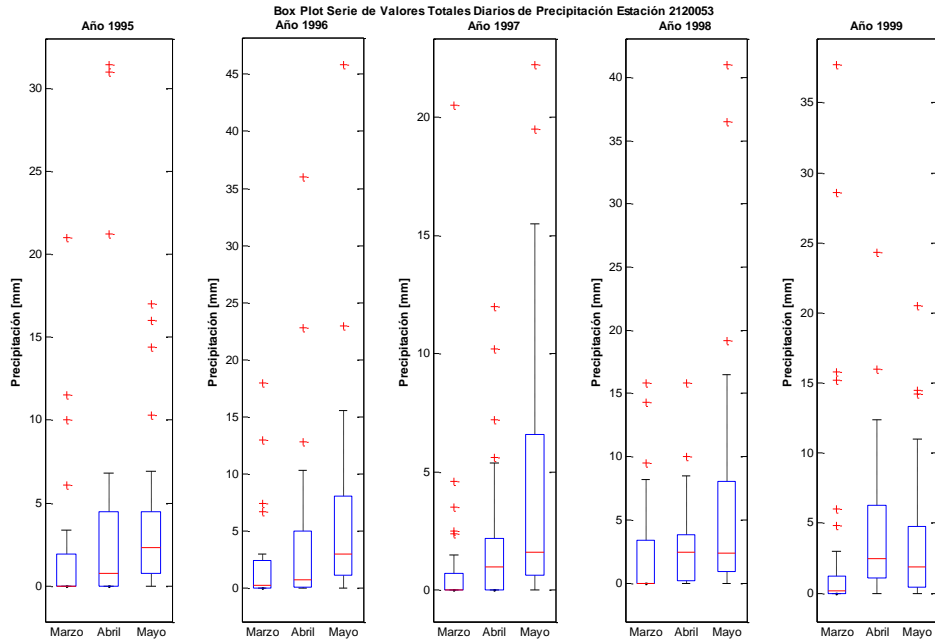
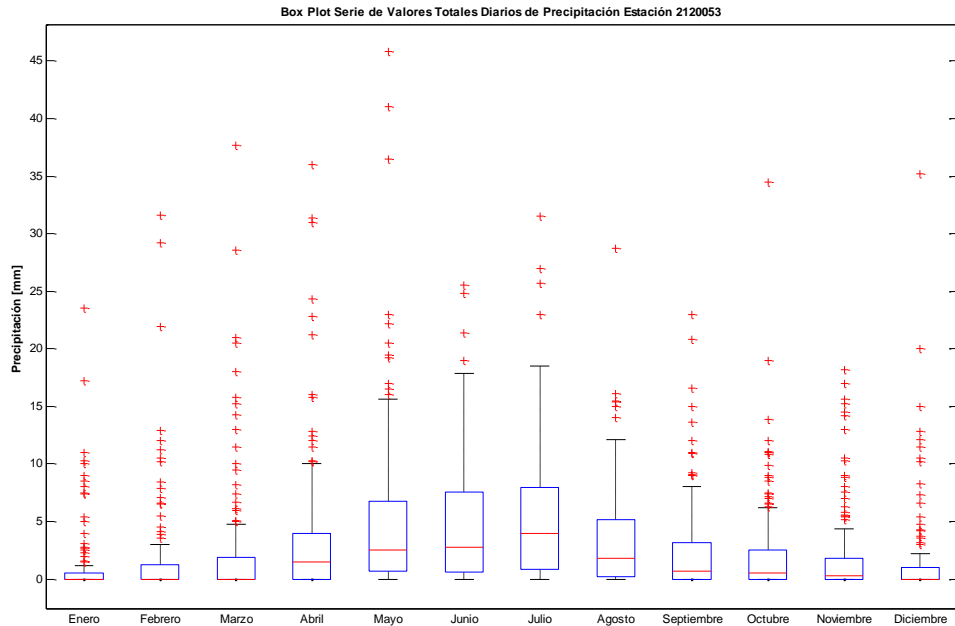


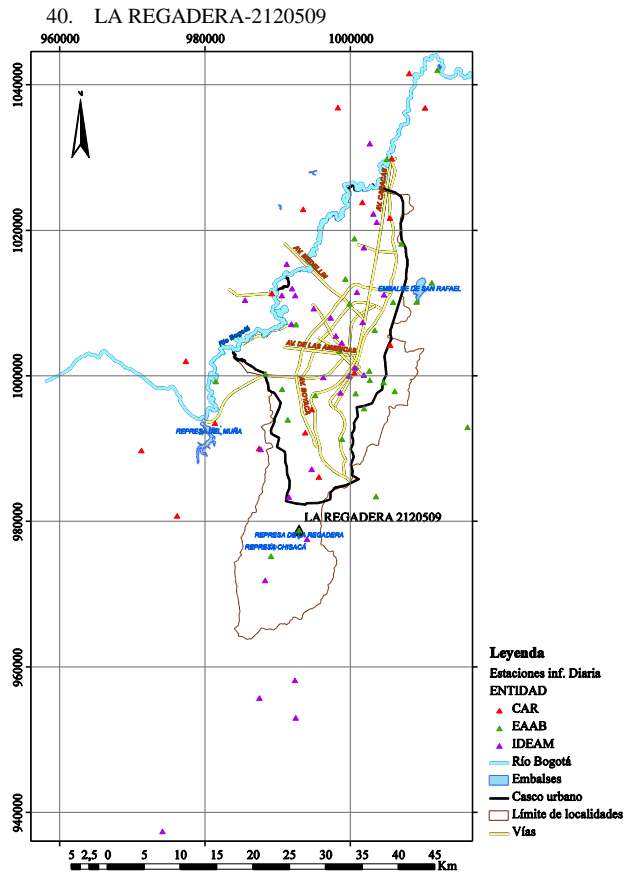




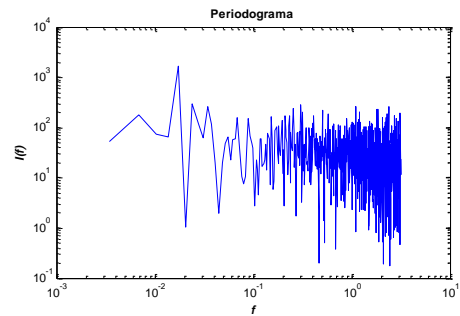
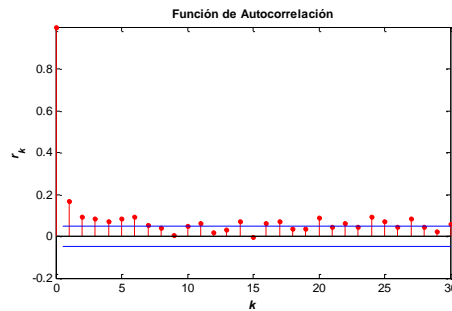
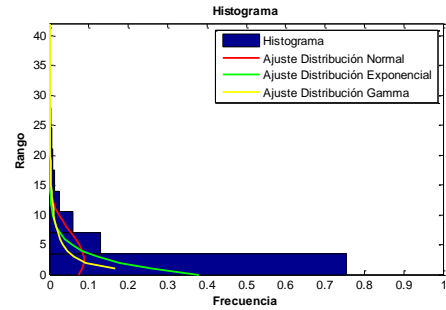
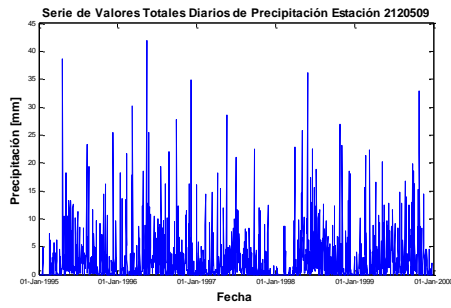
Este	1016011
Norte	993046
Cota	1950
Número de Datos	1818
Número de Datos Sin Registro	8
Media	3.0
Mediana	0.7
Máximo	45.8
Desviación Estándar	5.1
Varianza	25.9
Coefficiente de Variación	1.5E-03
Coefficiente de Asimetría	3.2
Coefficiente de Curtosis	16.8

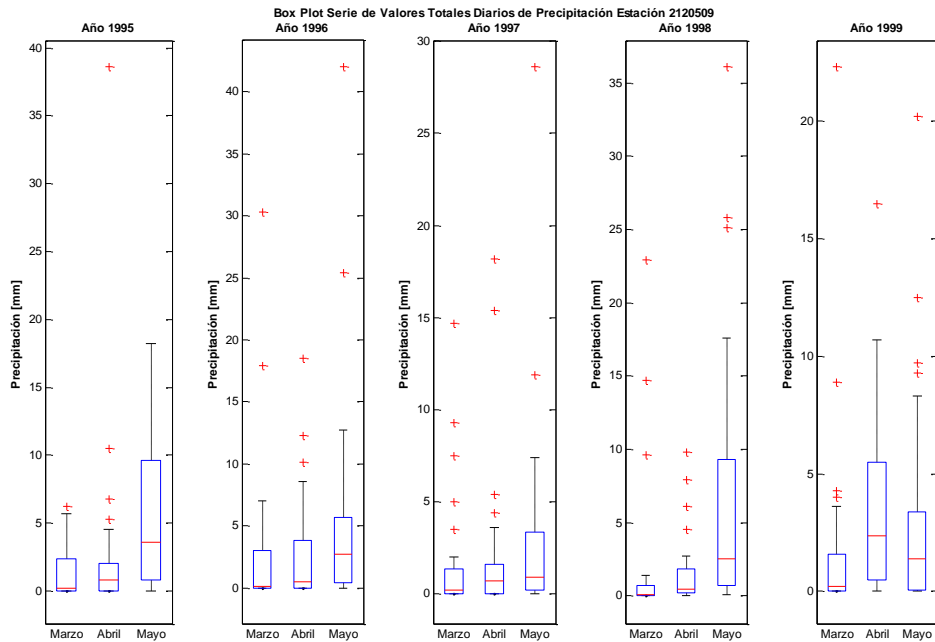
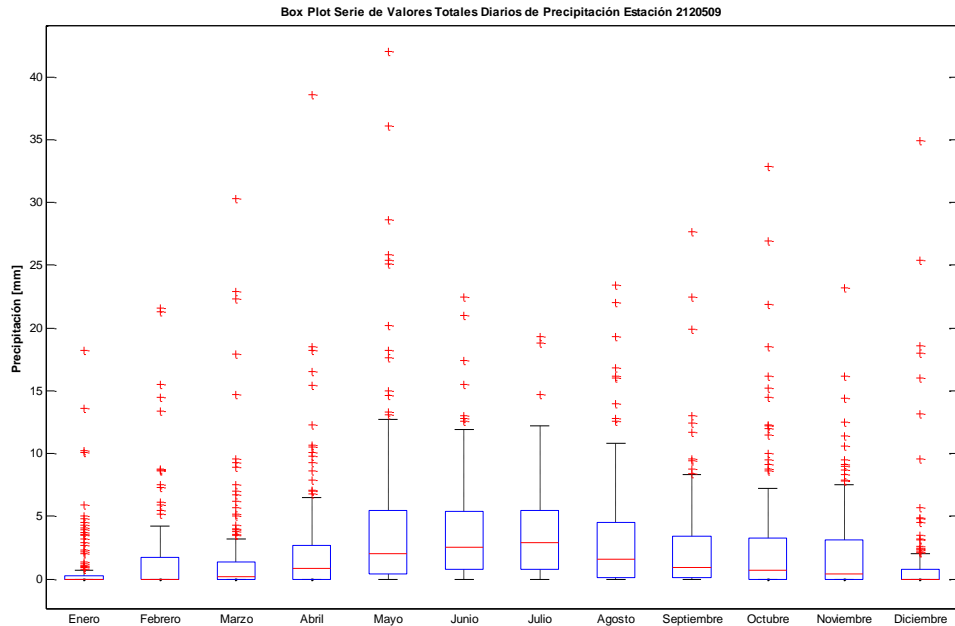


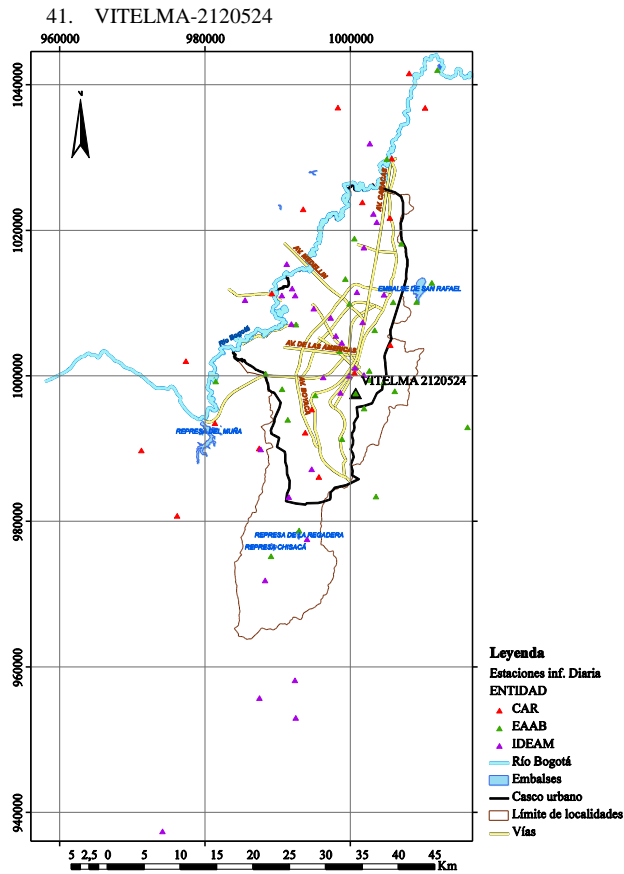




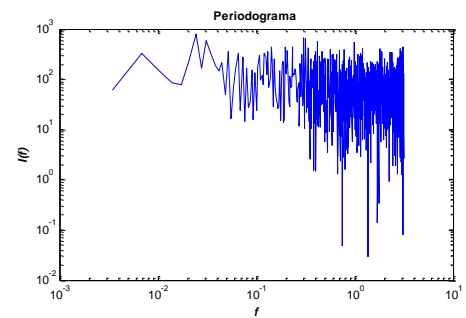
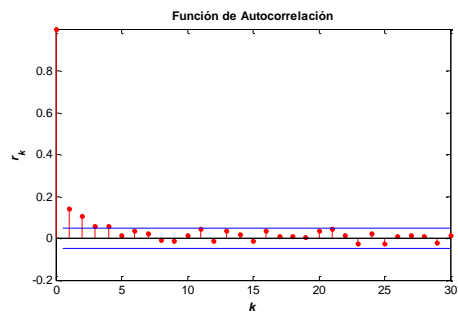
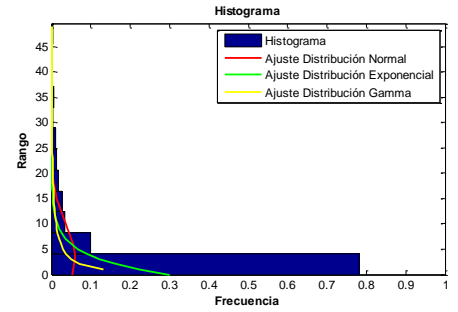
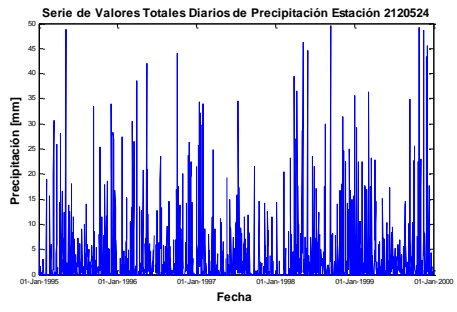
Este	992841
Norte	978850
Cota	3050
Número de Datos	1826
Número de Datos Sin Registro	0
Media	2.6
Mediana	0.7
Máximo	42.0
Desviación Estándar	4.5
Varianza	20.7
Coefficiente de Variación	8.7E-04
Coefficiente de Asimetría	3.3
Coefficiente de Curtosis	17.8

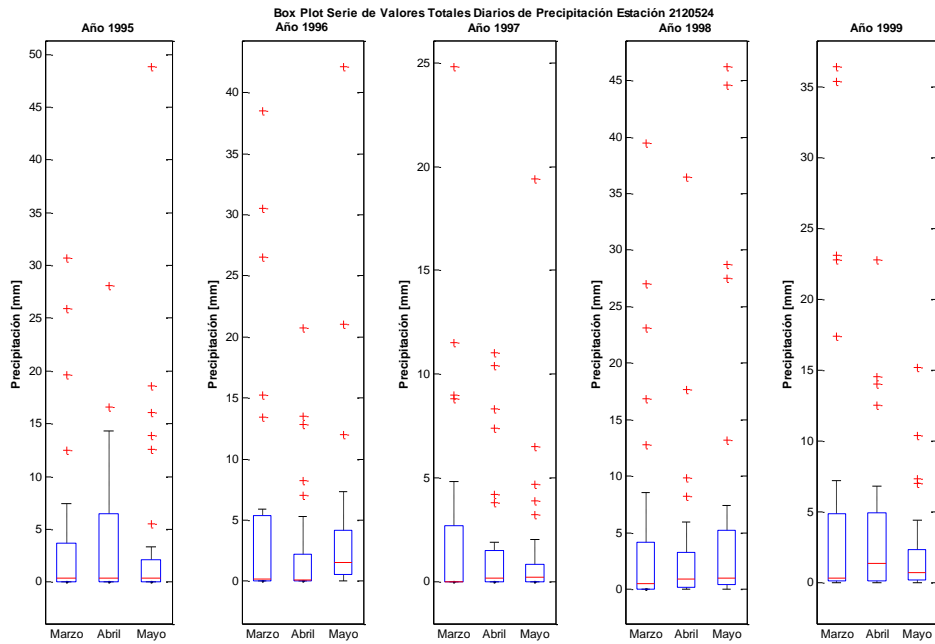
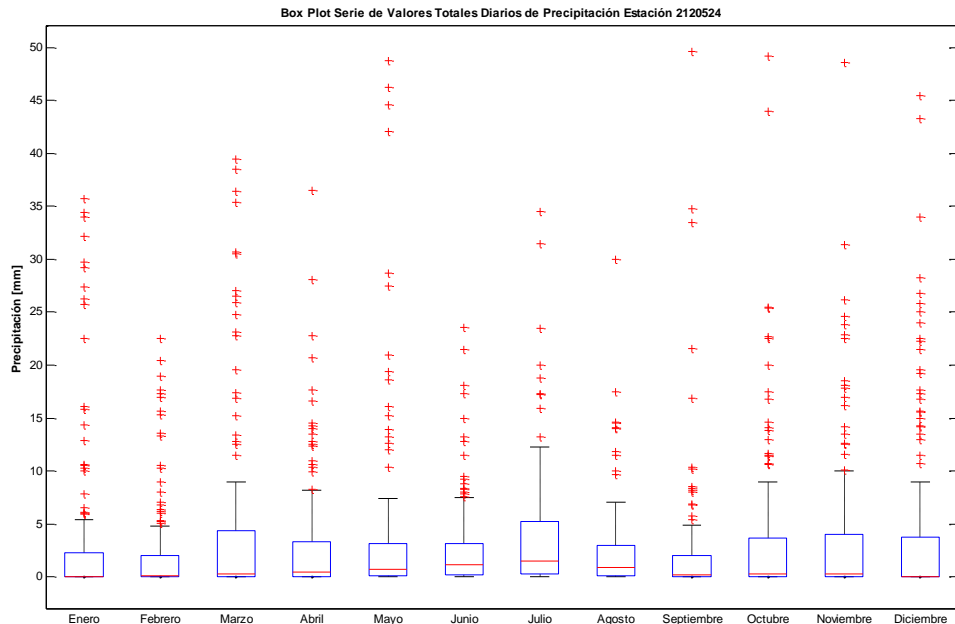




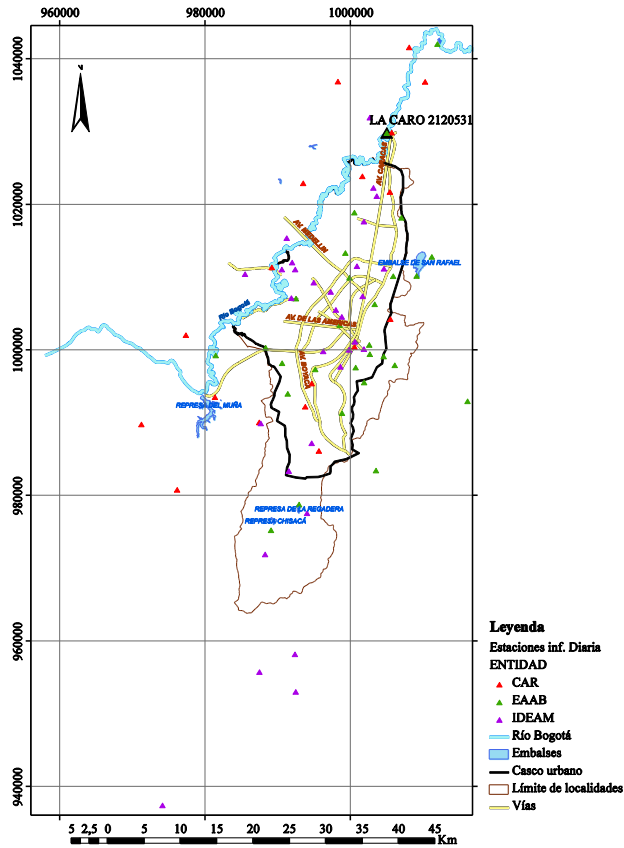


Este	1000641
Norte	997681
Cota	2800
Número de Datos	1817
Número de Datos Sin Registro	9
Media	3.3
Mediana	0.4
Máximo	49.6
Desviación Estándar	6.8
Varianza	45.9
Coefficiente de Variación	1.2E-03
Coefficiente de Asimetría	3.4
Coefficiente de Curtosis	16.4

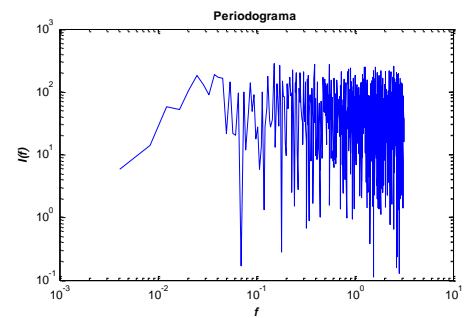
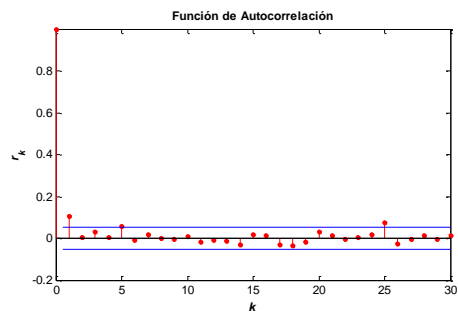
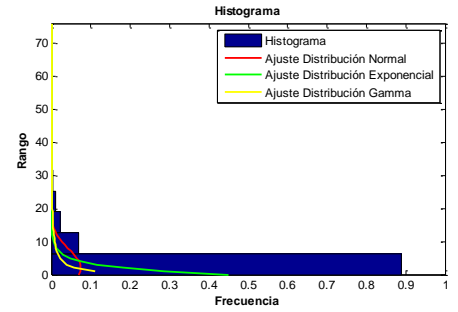
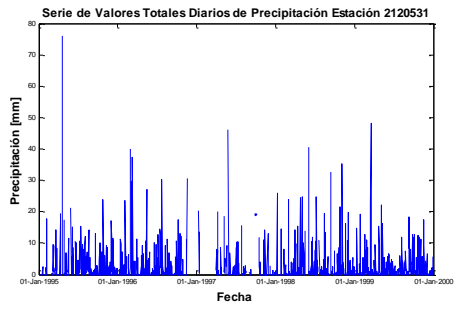


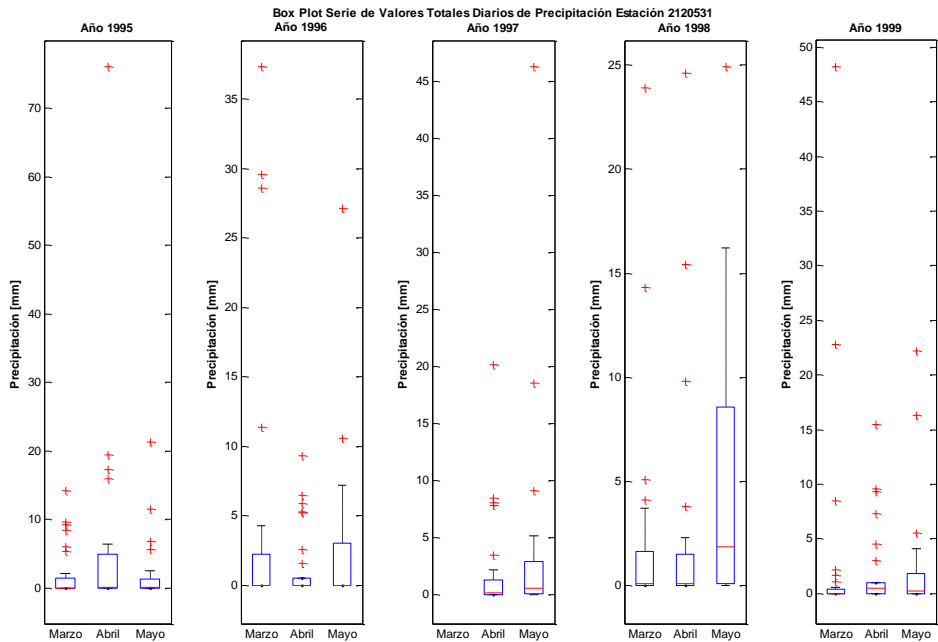
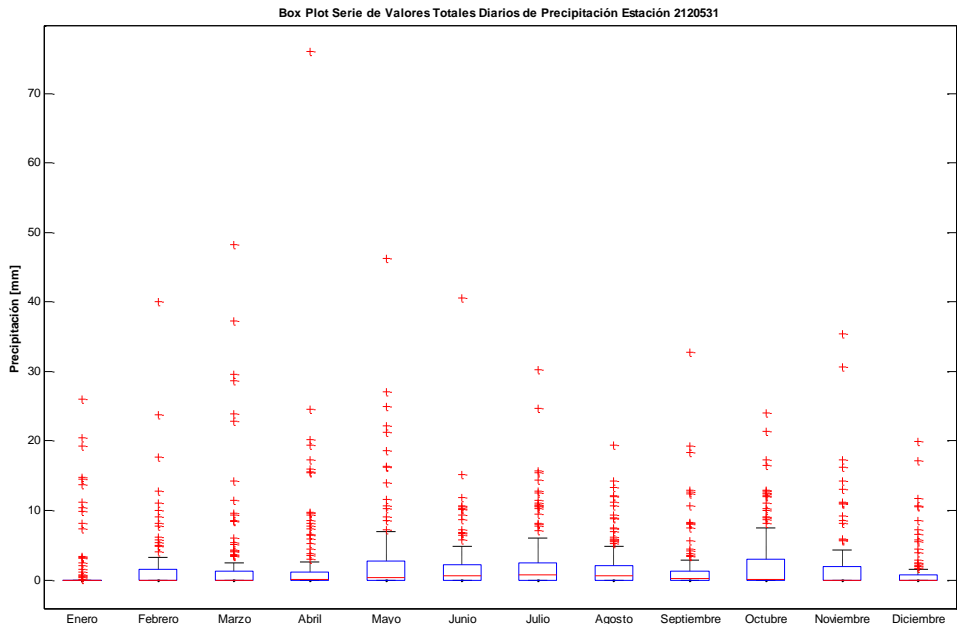


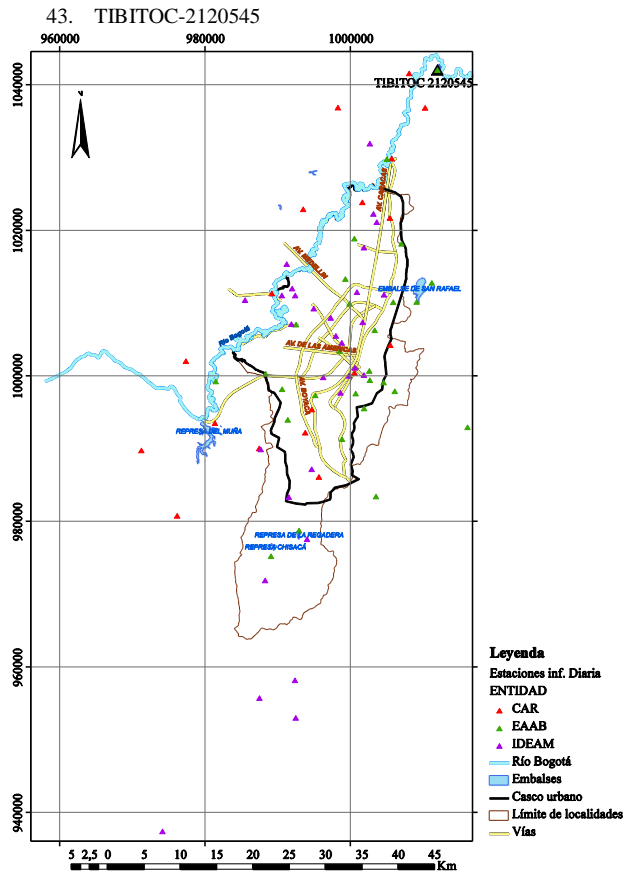
42. LA CARO-2120531



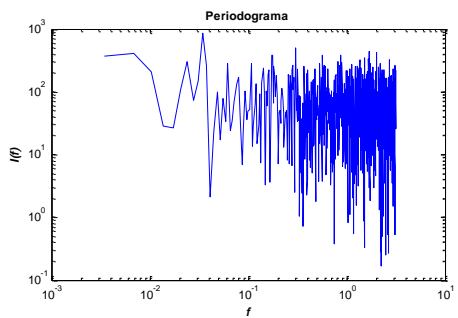
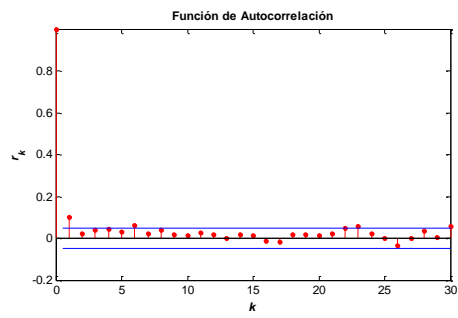
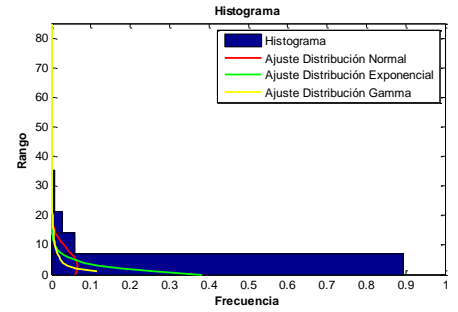
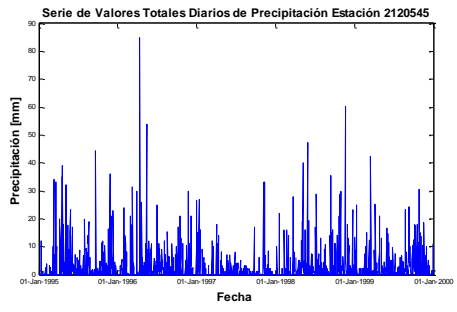
Este	1004903
Norte	1029918
Cota	2560
Número de Datos	1521
Número de Datos Sin Registro	305
Media	2.2
Mediana	0.1
Máximo	76.0
Desviación Estándar	5.3
Varianza	28.5
Coefficiente de Variación	8.7E-04
Coefficiente de Asimetría	5.0
Coefficiente de Curtosis	42.9

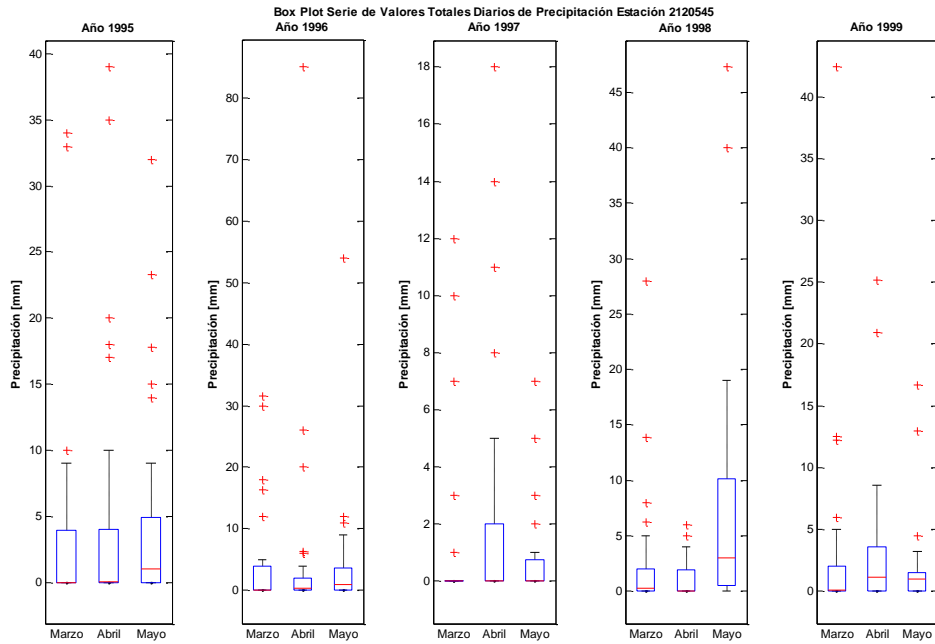
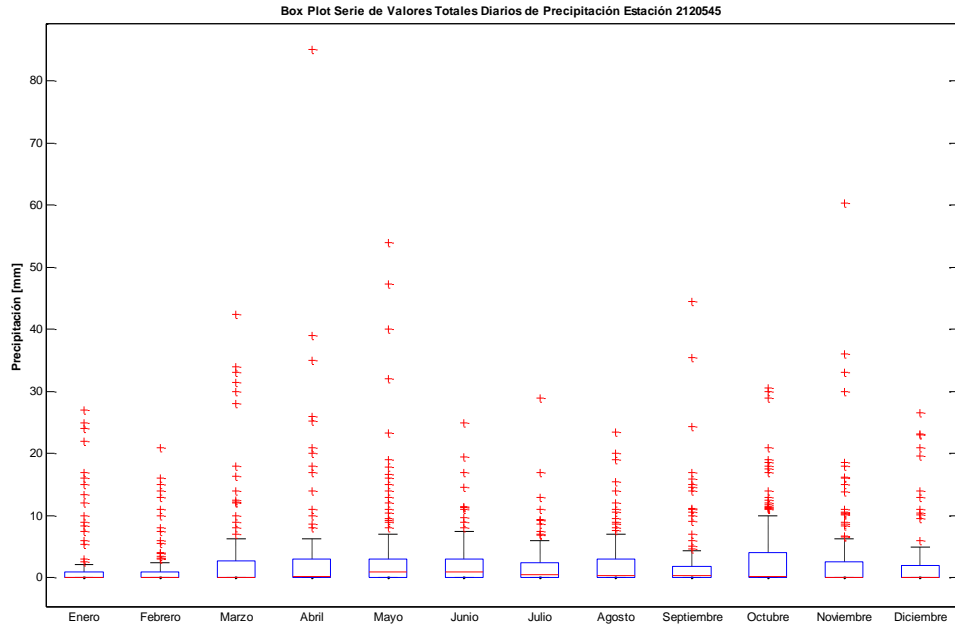


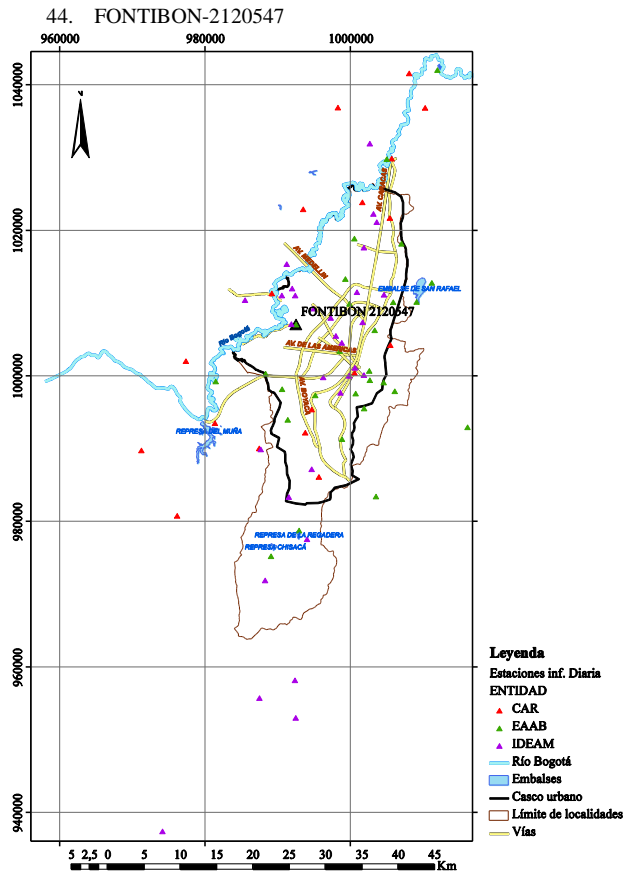




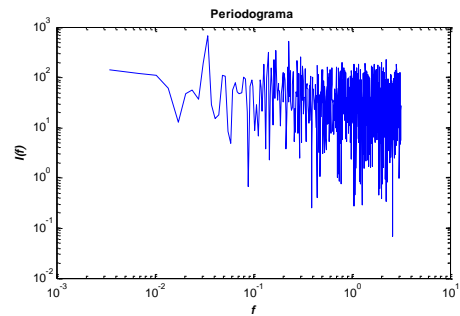
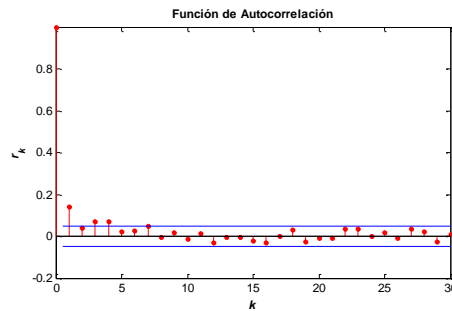
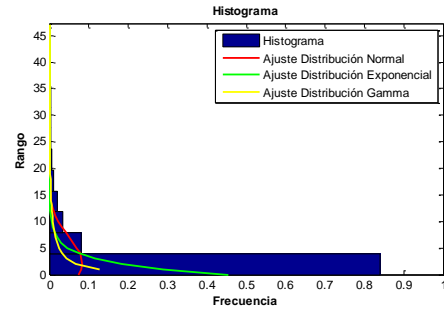
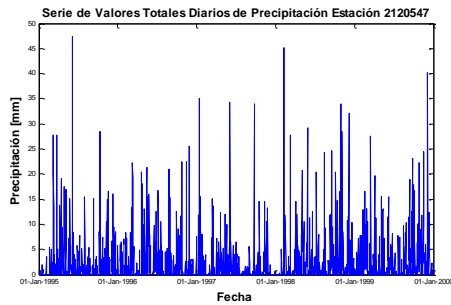
Este	1011894
Norte	1042150
Cota	2698
Número de Datos	1825
Número de Datos Sin Registro	1
Media	2.6
Mediana	0.2
Máximo	85.0
Desviación Estándar	6.1
Varianza	36.9
Coefficiente de Variación	9.7E-04
Coefficiente de Asimetría	4.8
Coefficiente de Curtosis	38.5

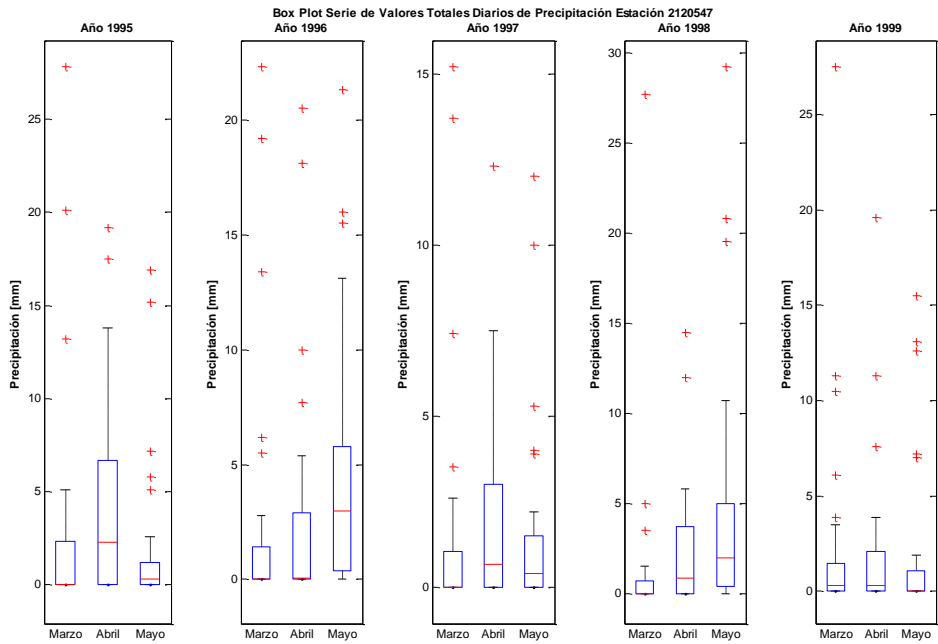
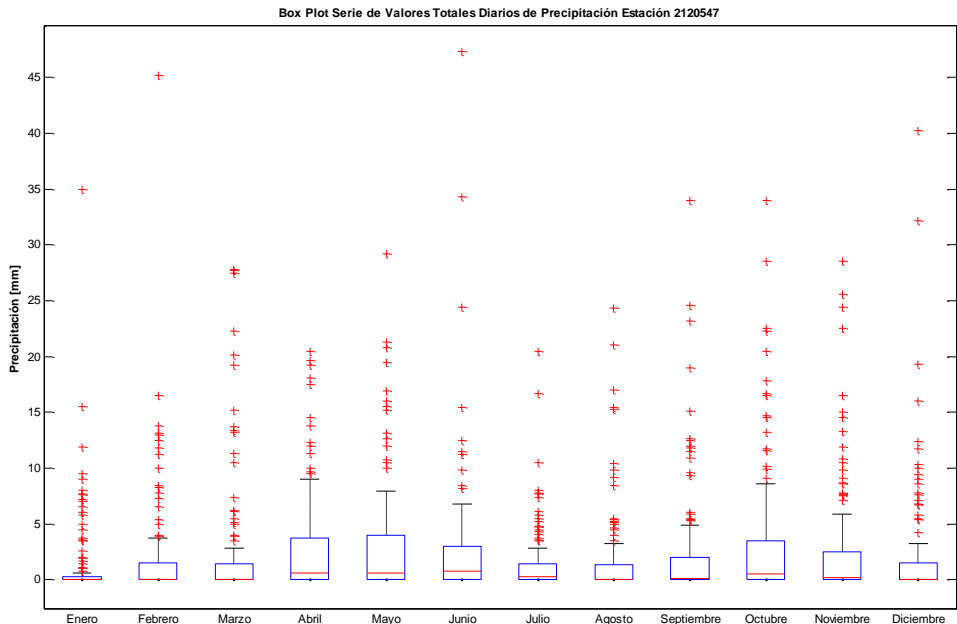


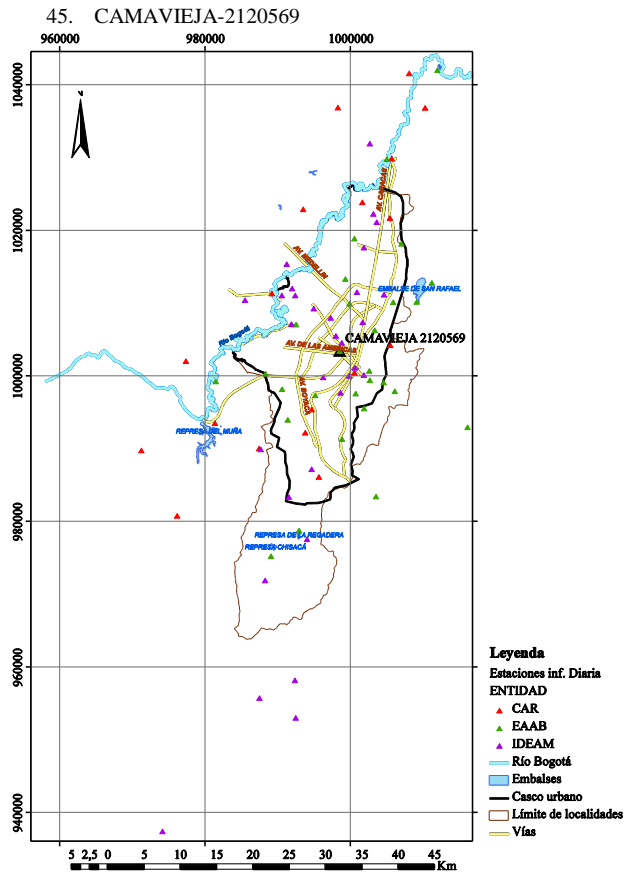




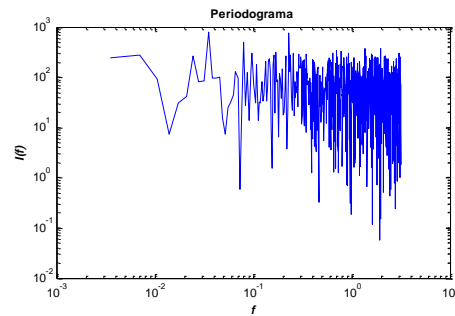
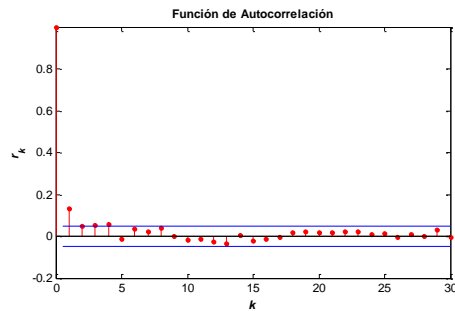
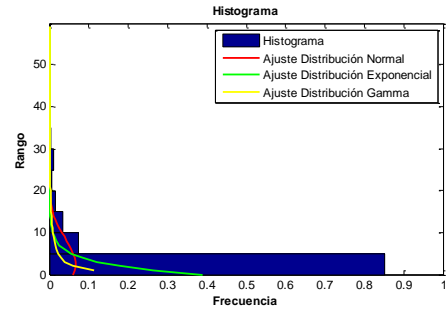
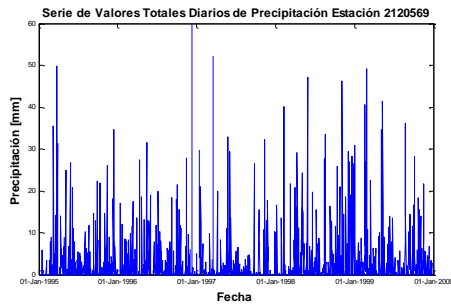
Este	992432
Norte	1007176
Cota	2518
Número de Datos	1825
Número de Datos Sin Registro	1
Media	2.2
Mediana	0.2
Máximo	47.3
Desviación Estándar	4.8
Varianza	23.2
Coefficiente de Variación	8.7E-04
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	23.3

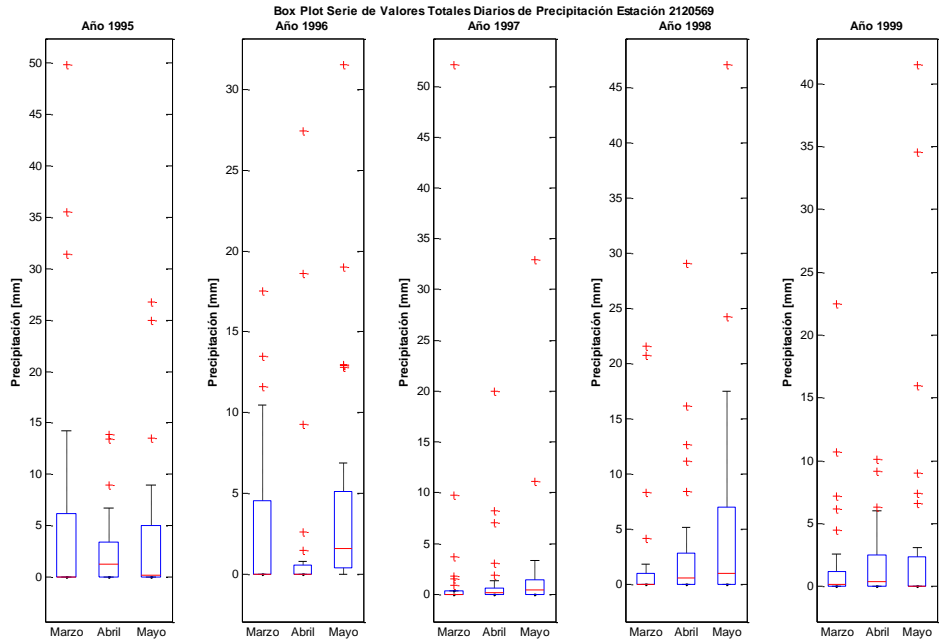
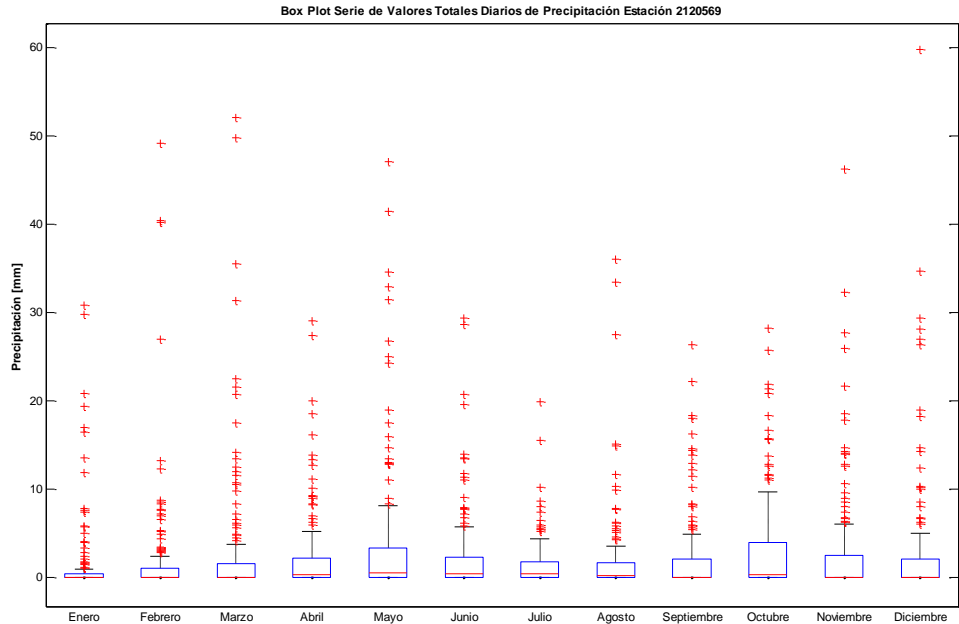


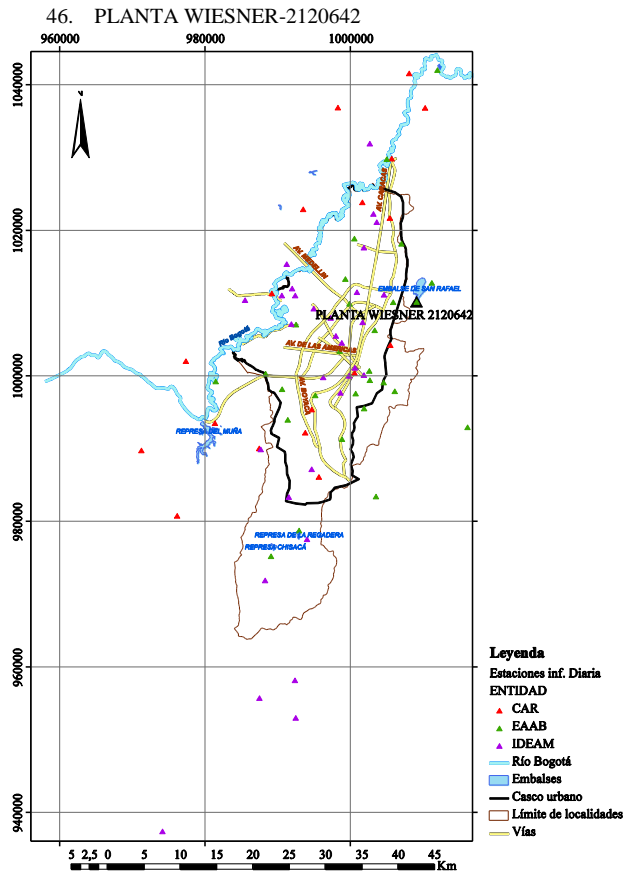




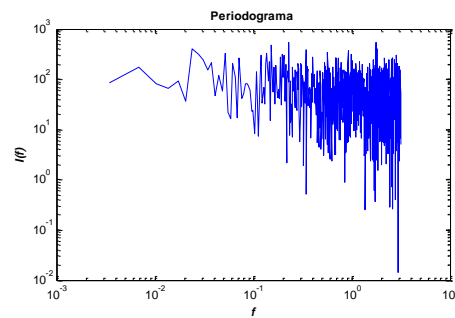
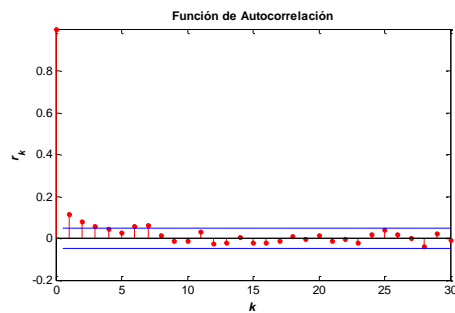
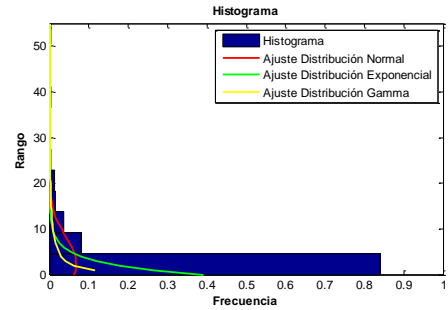
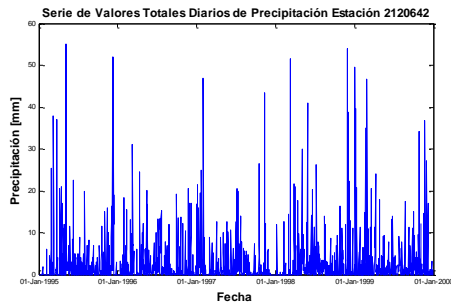
Este	998447
Norte	1003528
Cota	2680
Número de Datos	1801
Número de Datos Sin Registro	25
Media	2.6
Mediana	0.1
Máximo	59.8
Desviación Estándar	6.1
Varianza	36.8
Coefficiente de Variación	9.6E-04
Coefficiente de Asimetría	4.1
Coefficiente de Curtosis	24.8

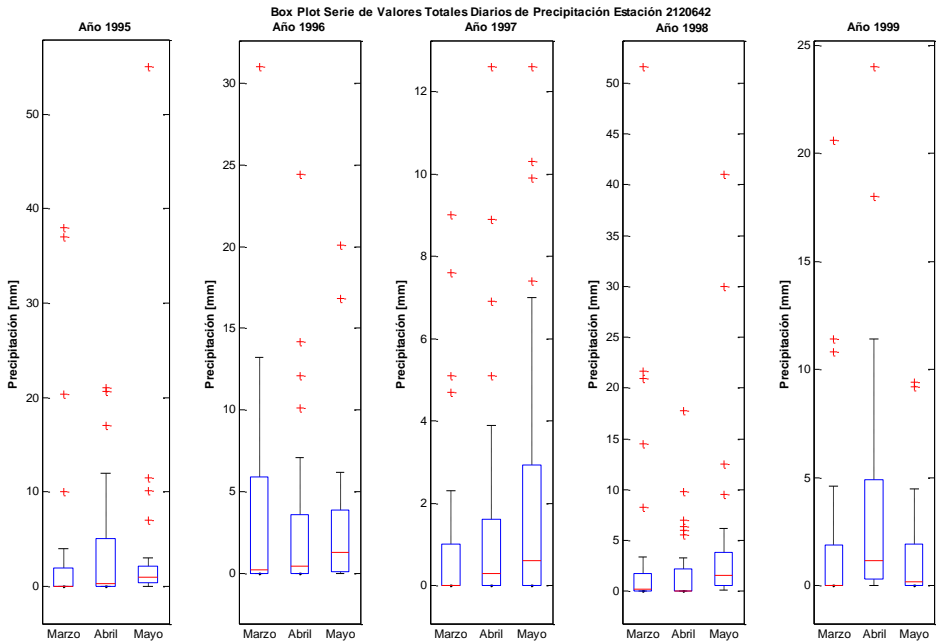
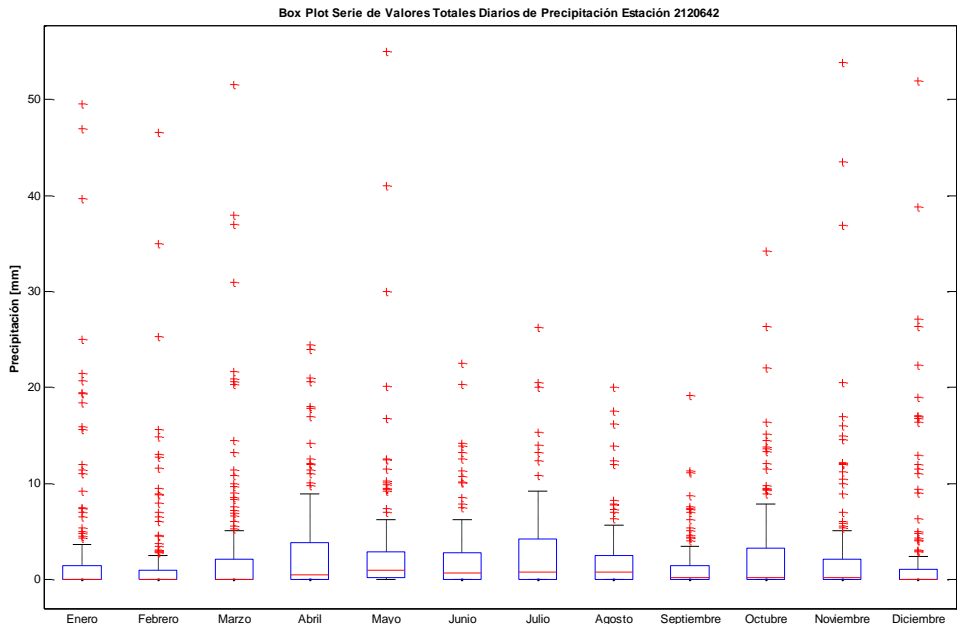


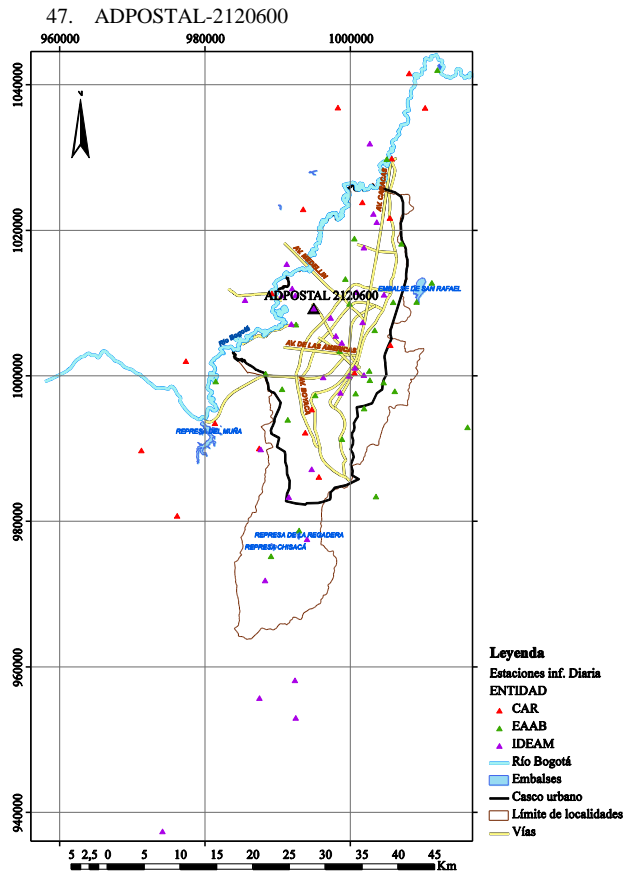




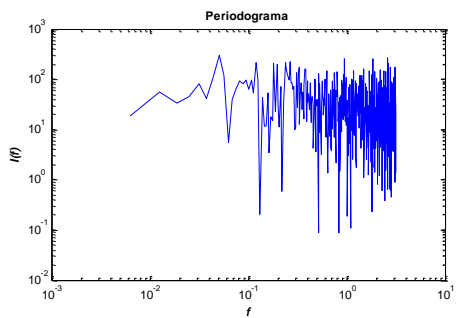
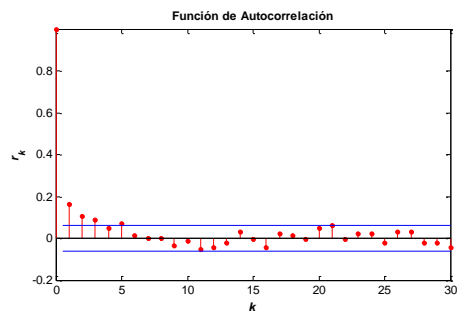
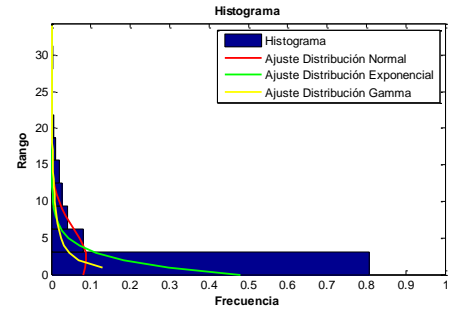
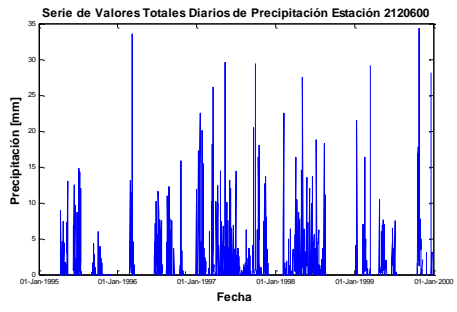
Este	1009006
Norte	1010268
Cota	2795
Número de Datos	1824
Número de Datos Sin Registro	2
Media	2.6
Mediana	0.3
Máximo	55.0
Desviación Estándar	5.9
Varianza	34.6
Coefficiente de Variación	9.1E-04
Coefficiente de Asimetría	4.6
Coefficiente de Curtosis	30.9

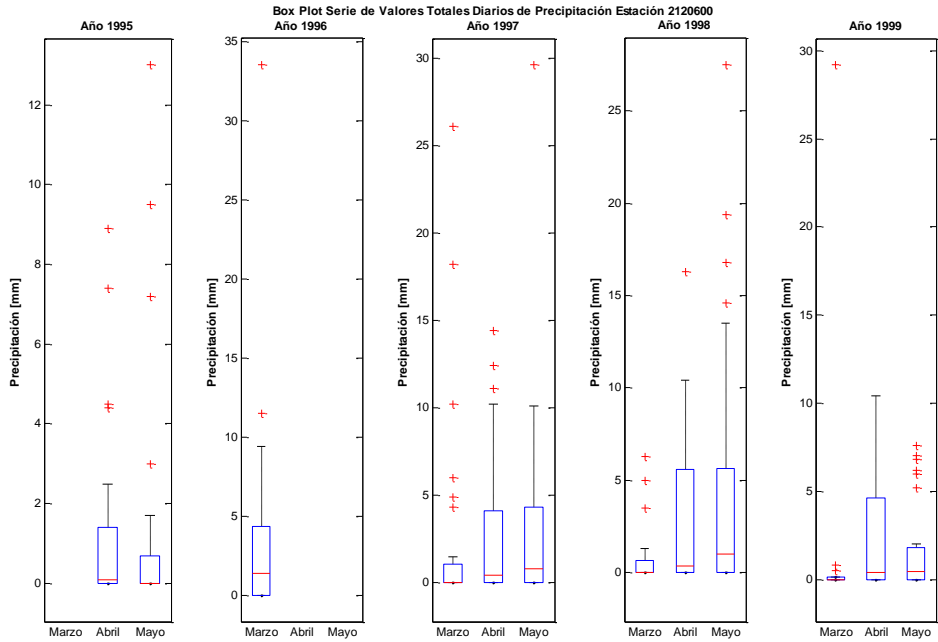
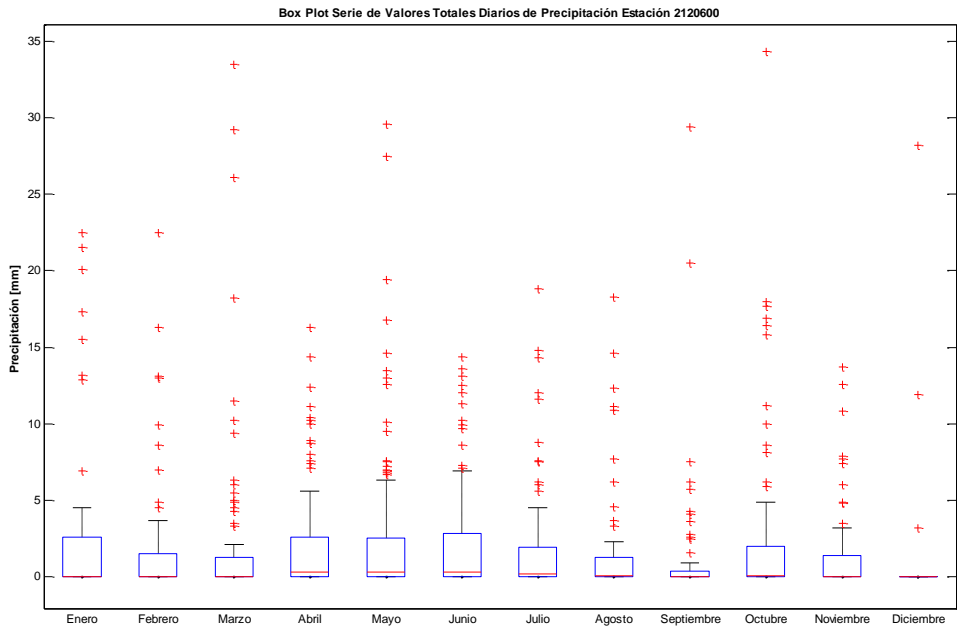


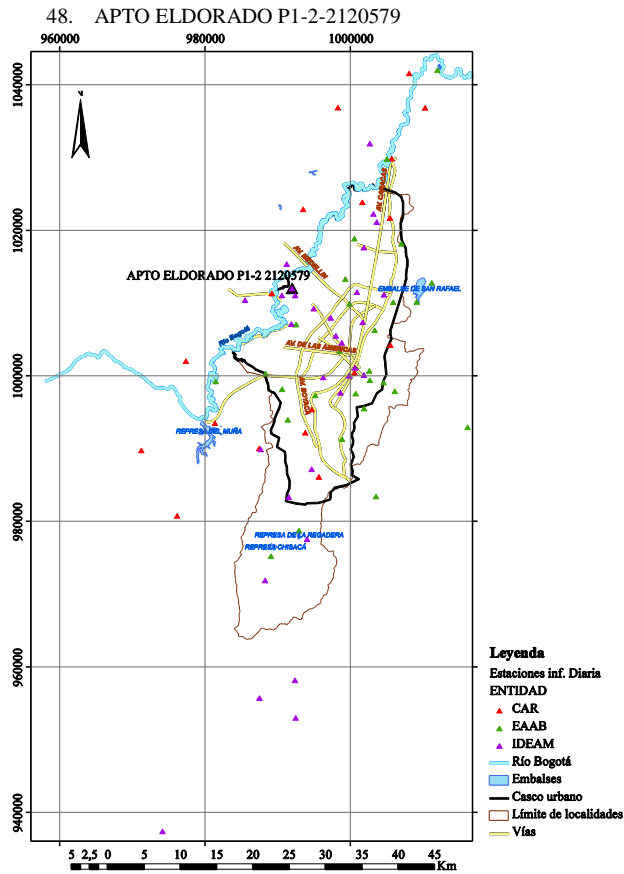




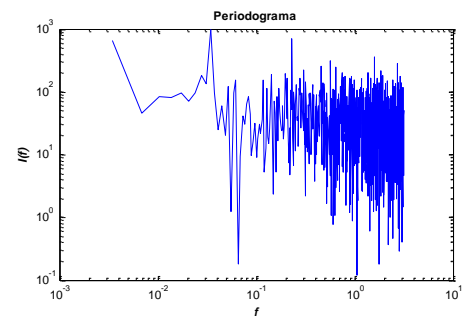
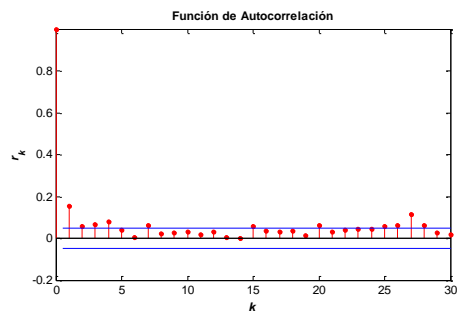
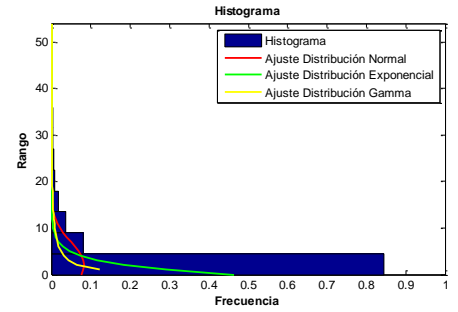
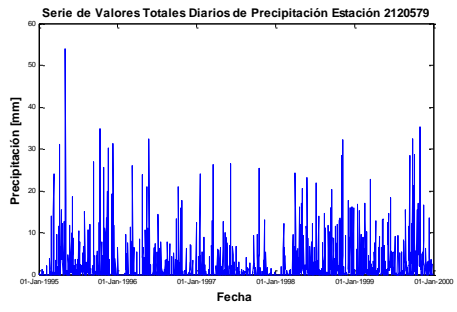
Este	994882
Norte	1009350
Cota	2550
Número de Datos	998
Número de Datos Sin Registro	828
Media	2.1
Mediana	0.1
Máximo	34.3
Desviación Estándar	4.5
Varianza	20.2
Coefficiente de Variación	8.2E-04
Coefficiente de Asimetría	3.4
Coefficiente de Curtosis	17.0

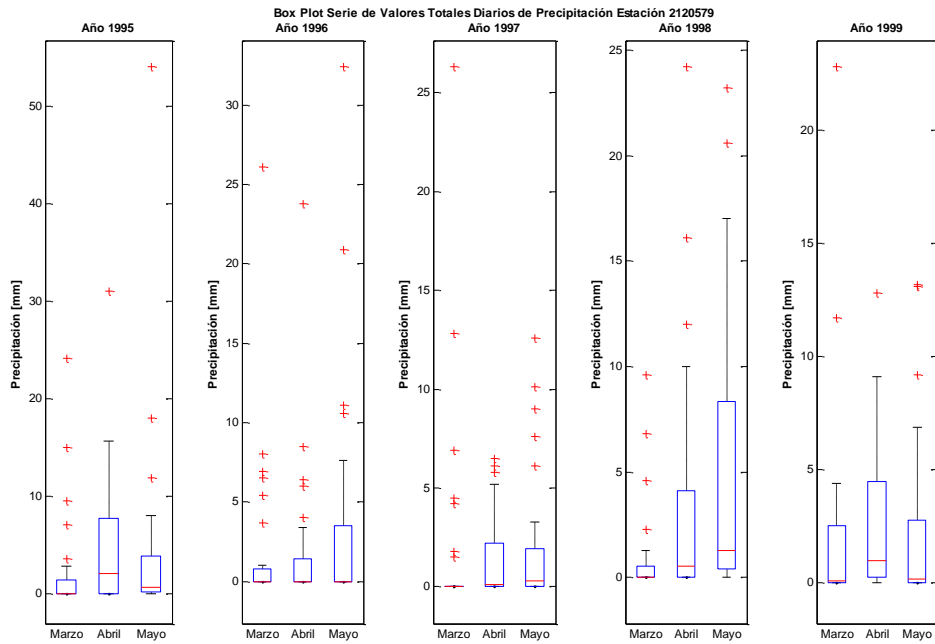
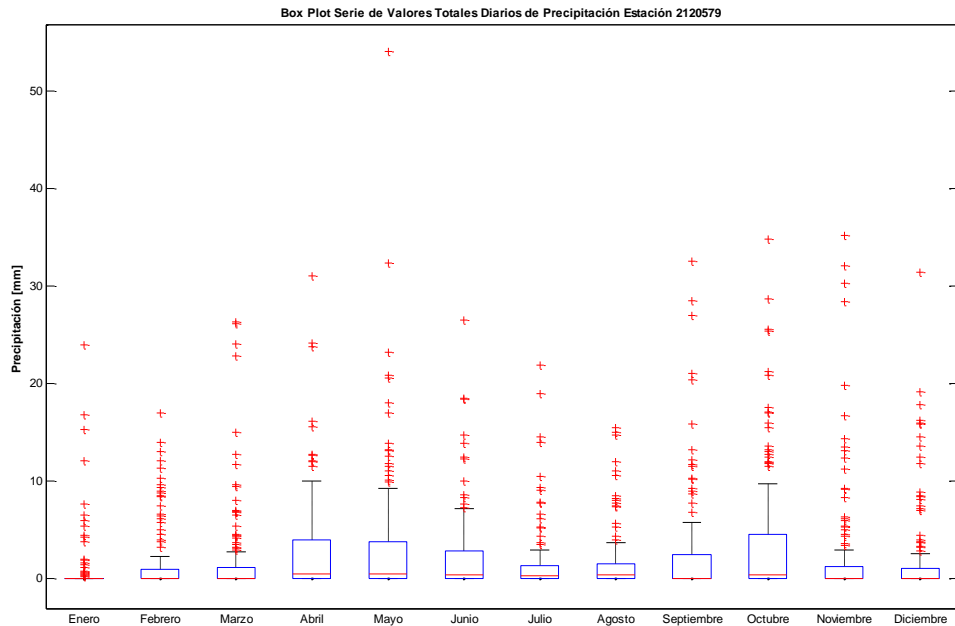


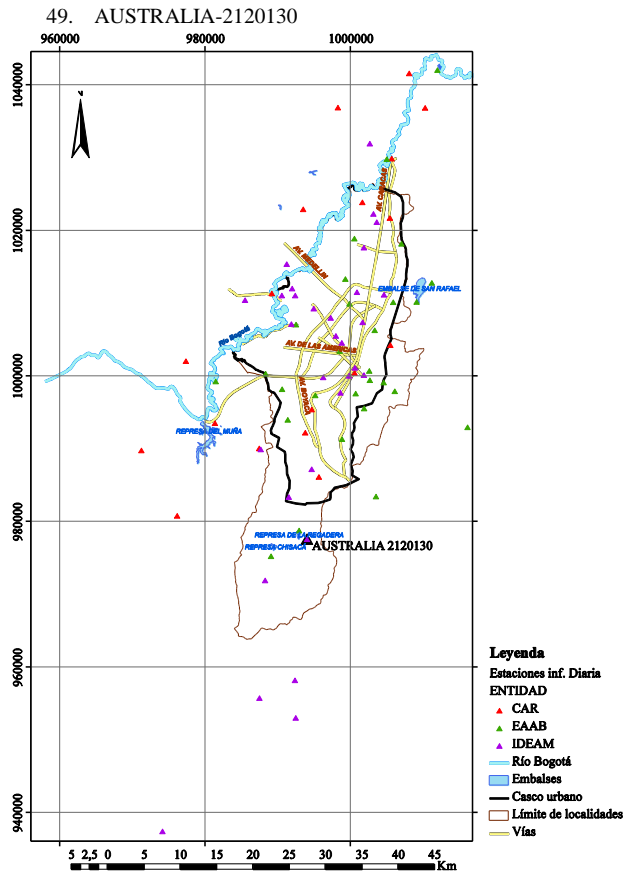




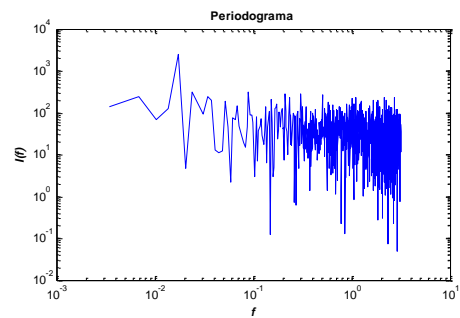
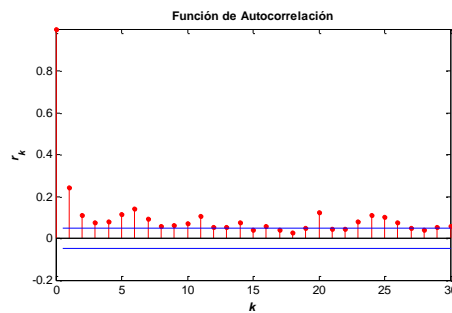
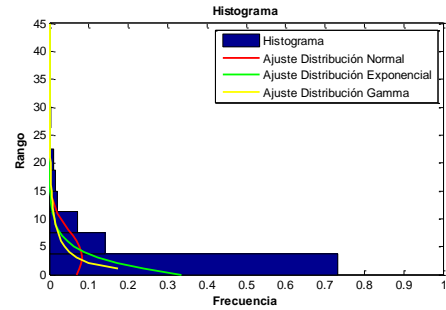
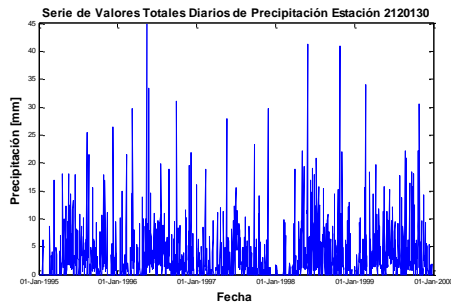
Este	991883
Norte	1012096
Cota	2547
Número de Datos	1822
Número de Datos Sin Registro	4
Media	2.2
Mediana	0.0
Máximo	54.0
Desviación Estándar	4.8
Varianza	23.2
Coefficiente de Variación	8.5E-04
Coefficiente de Asimetría	3.8
Coefficiente de Curtosis	22.3

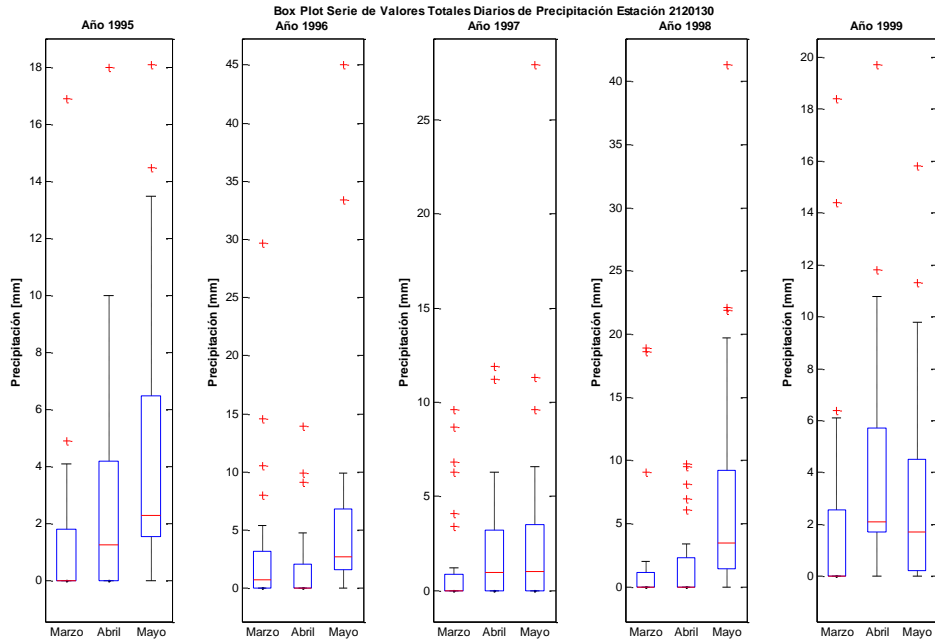
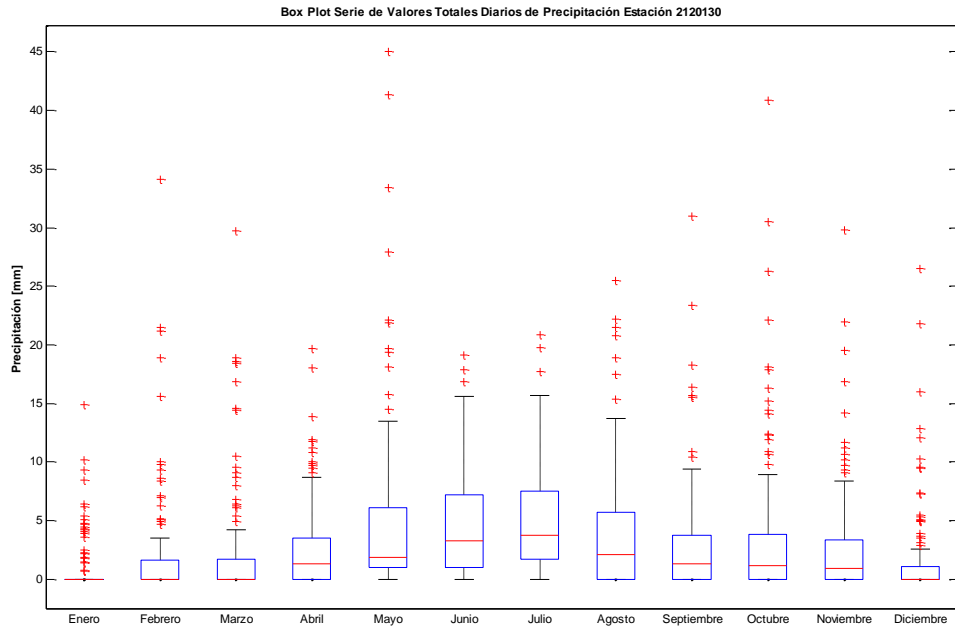


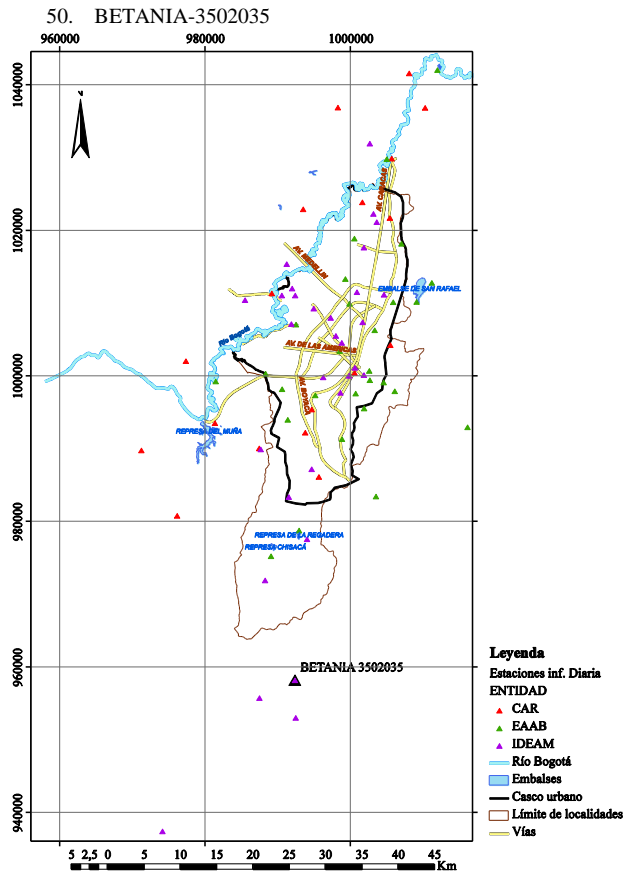




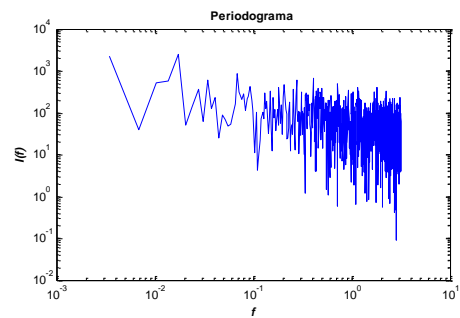
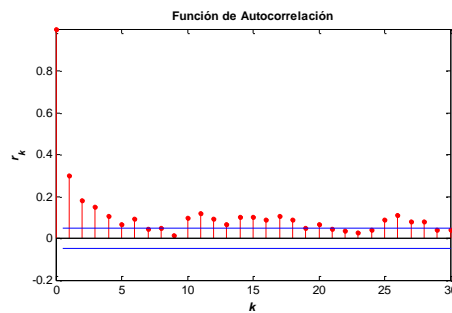
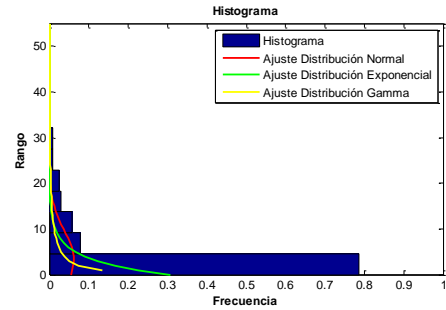
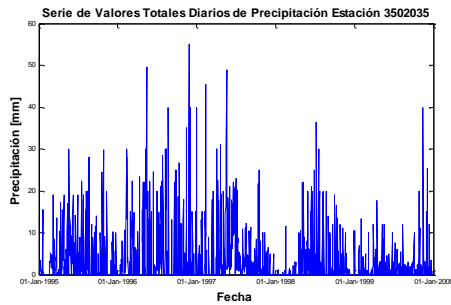
Este	993952
Norte	977668
Cota	3050
Número de Datos	1826
Número de Datos Sin Registro	0
Media	3.0
Mediana	1.2
Máximo	45.0
Desviación Estándar	4.8
Varianza	23.0
Coefficiente de Variación	9.8E-04
Coefficiente de Asimetría	3.1
Coefficiente de Curtosis	17.0

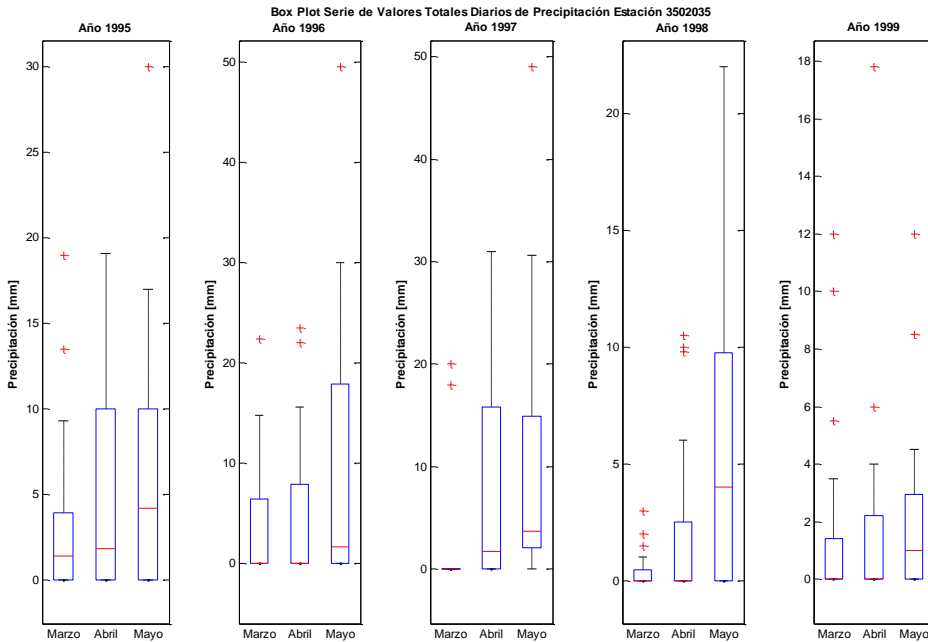
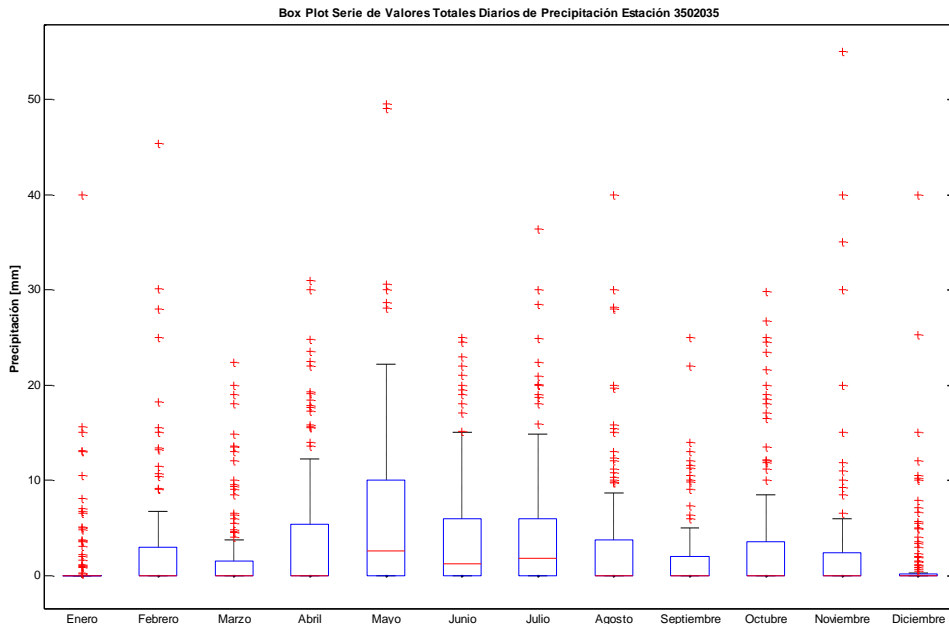


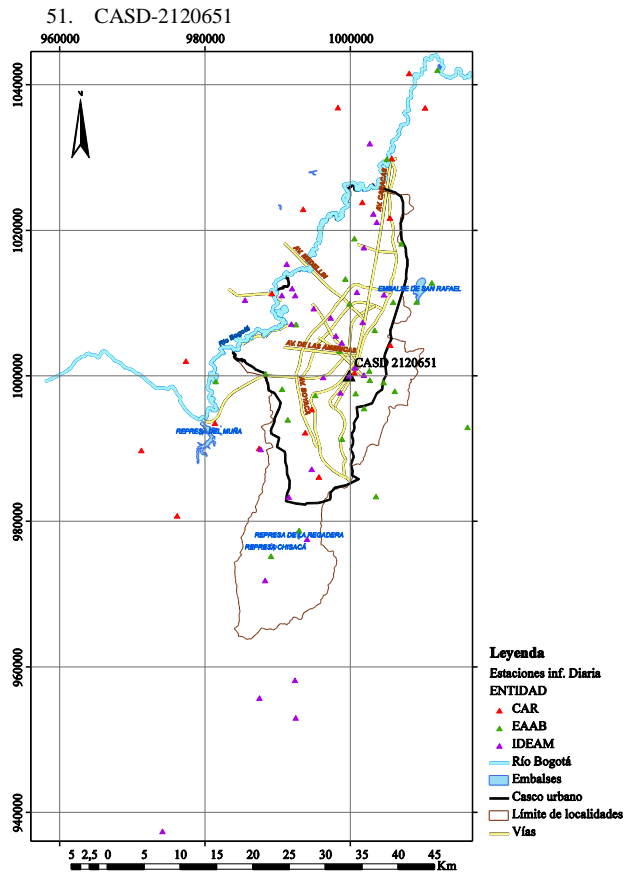




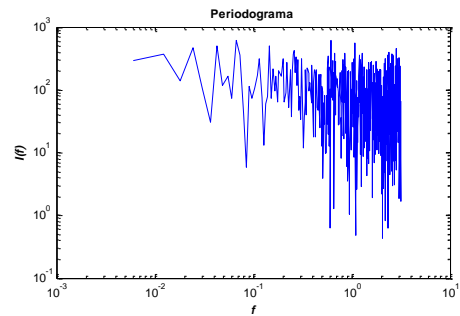
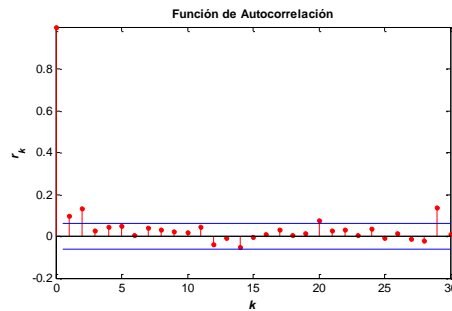
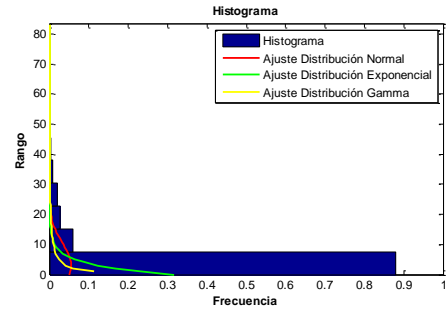
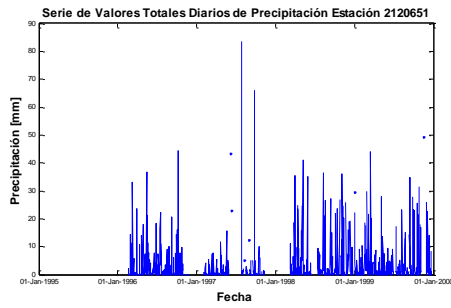
Este	992300
Norte	958277
Cota	3150
Número de Datos	1823
Número de Datos Sin Registro	3
Media	3.2
Mediana	0.0
Máximo	55.0
Desviación Estándar	6.5
Varianza	41.9
Coefficiente de Variación	1.0E-03
Coefficiente de Asimetría	3.0
Coefficiente de Curtosis	14.4

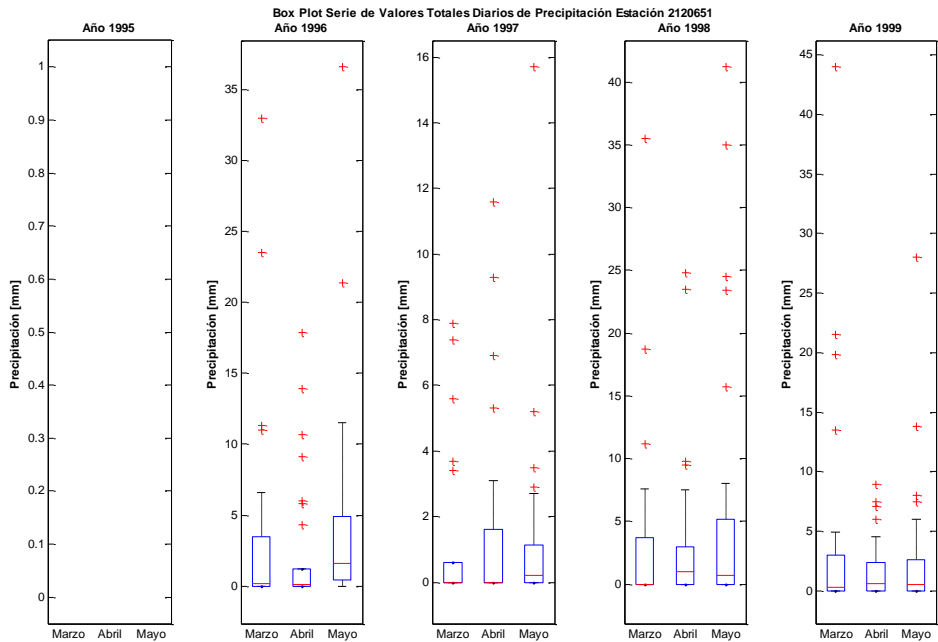
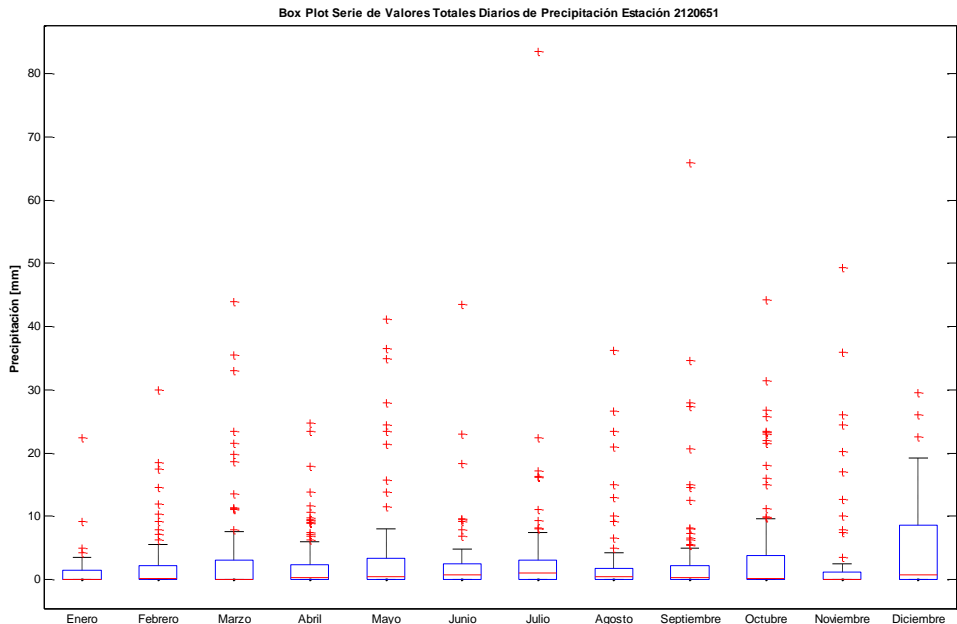


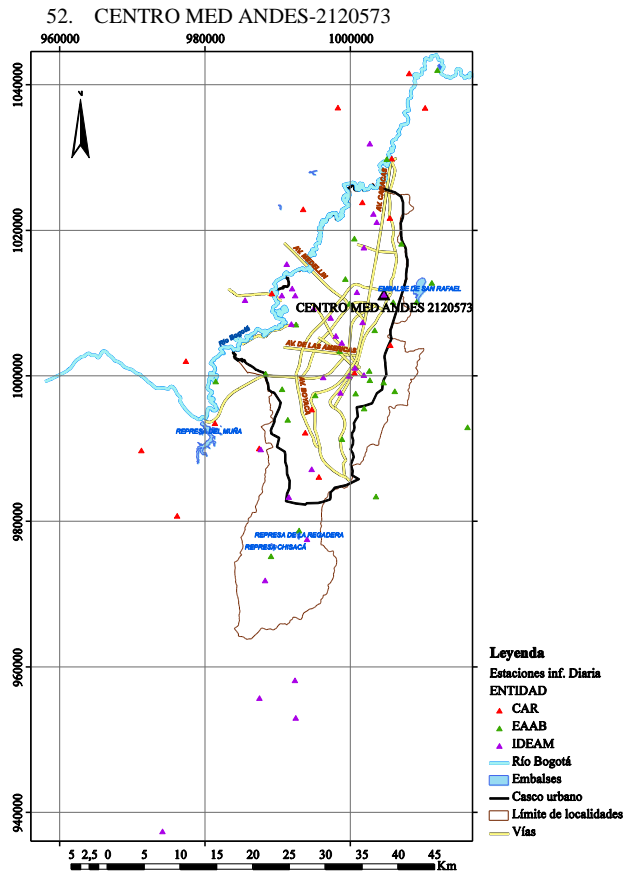




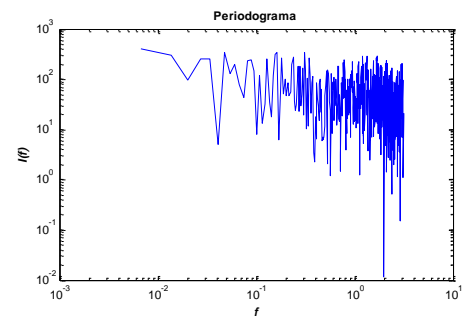
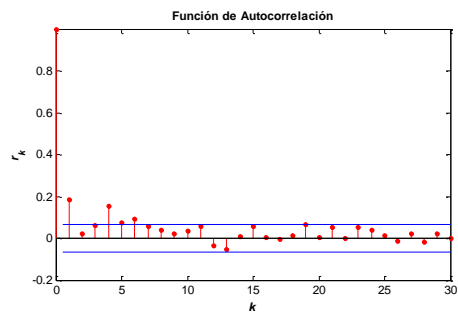
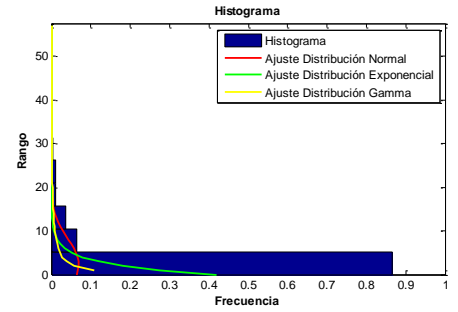
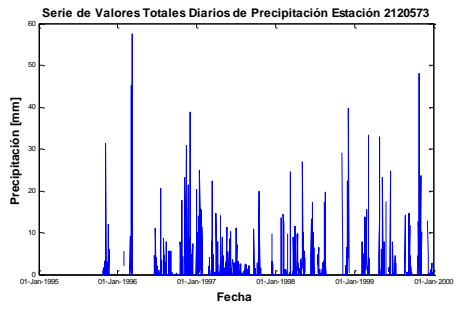
Este	999733
Norte	1000106
Cota	2600
Número de Datos	1033
Número de Datos Sin Registro	793
Media	3.2
Mediana	0.3
Máximo	83.5
Desviación Estándar	7.3
Varianza	53.1
Coefficiente de Variación	1.2E-03
Coefficiente de Asimetría	4.4
Coefficiente de Curtosis	30.3

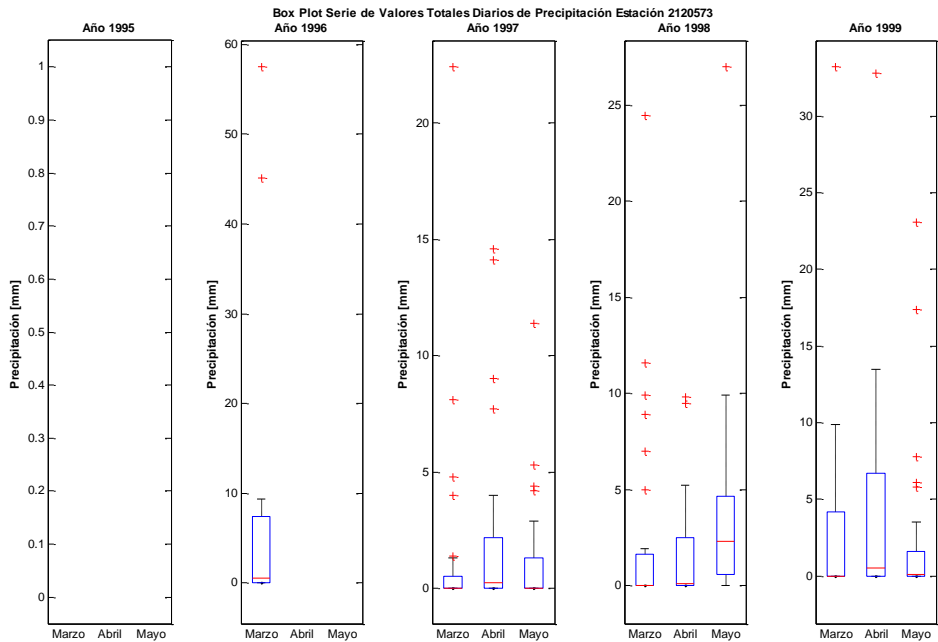
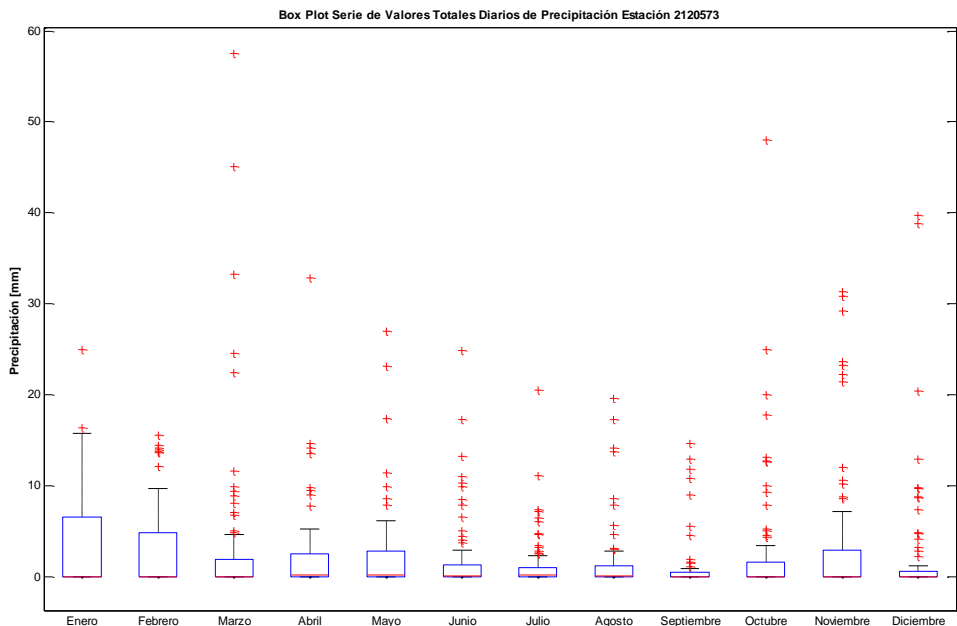


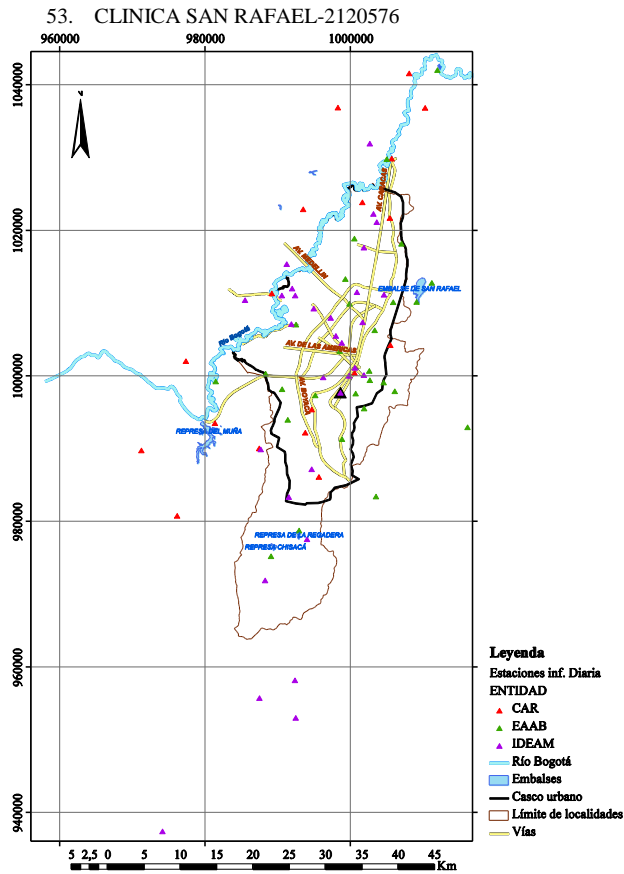




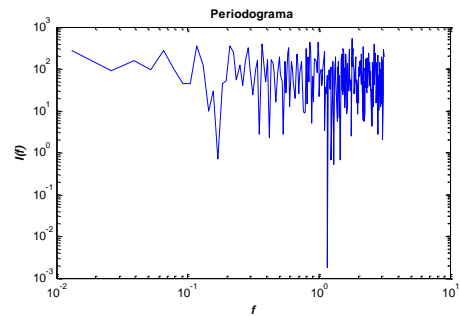
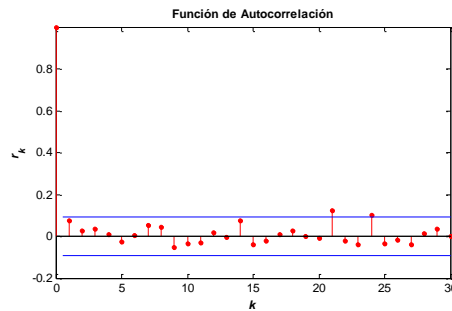
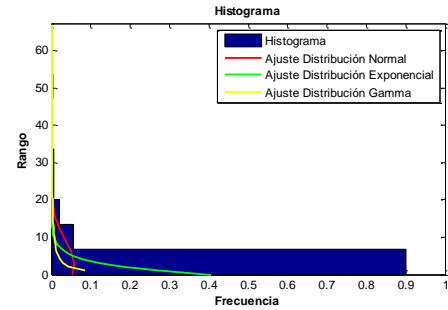
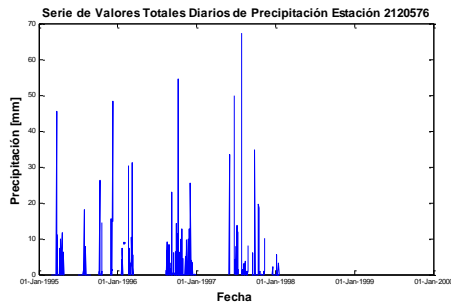
Este	1004513
Norte	1011276
Cota	2575
Número de Datos	929
Número de Datos Sin Registro	897
Media	2.4
Mediana	0.0
Máximo	57.5
Desviación Estándar	5.8
Varianza	33.6
Coefficiente de Variación	9.3E-04
Coefficiente de Asimetría	4.2
Coefficiente de Curtosis	26.7

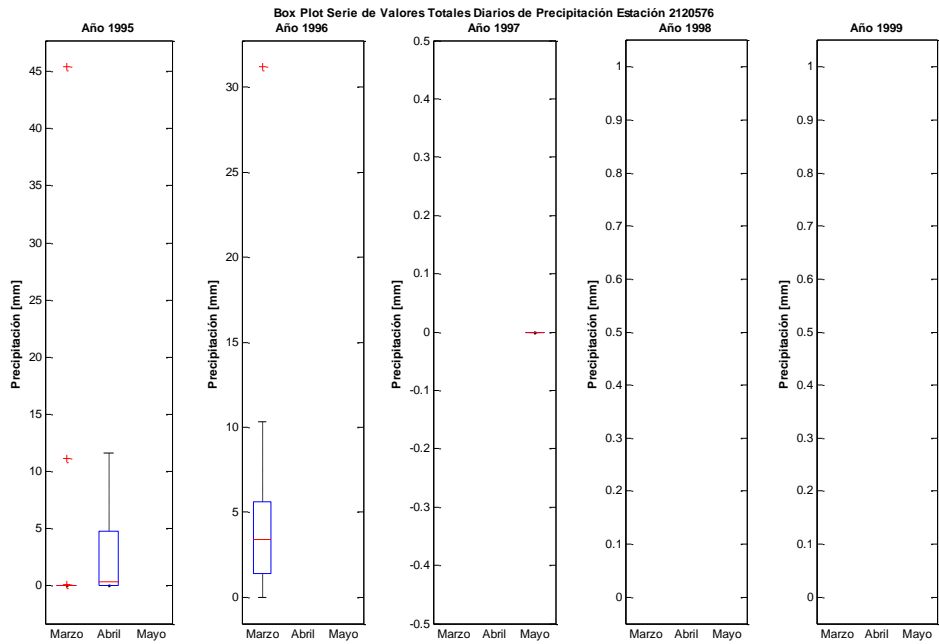
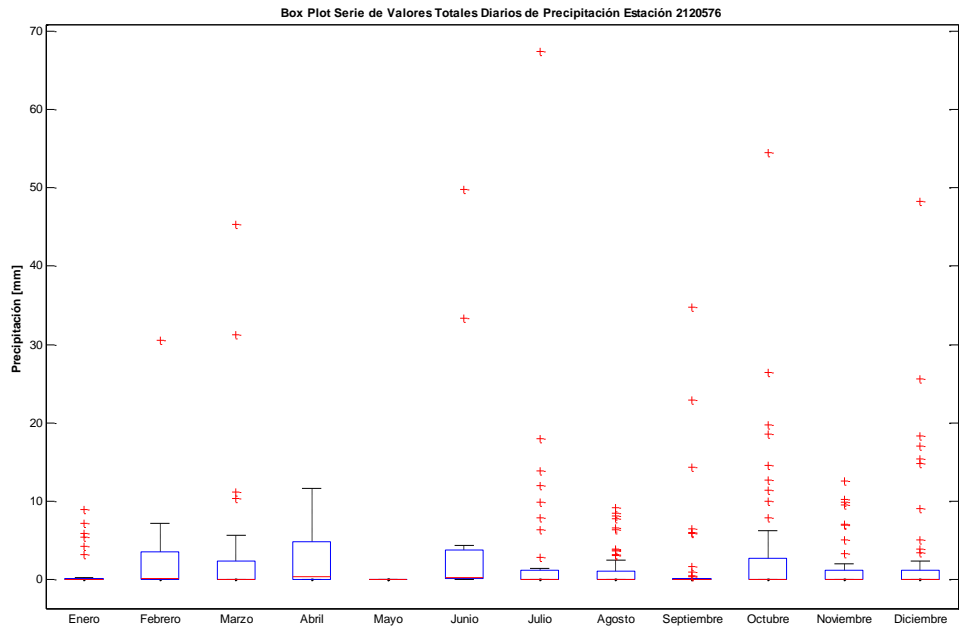


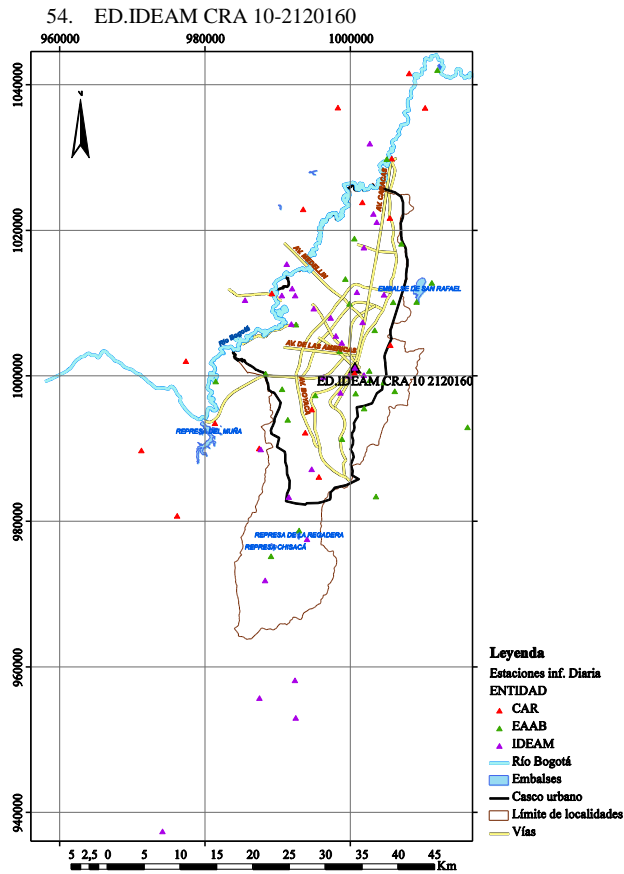




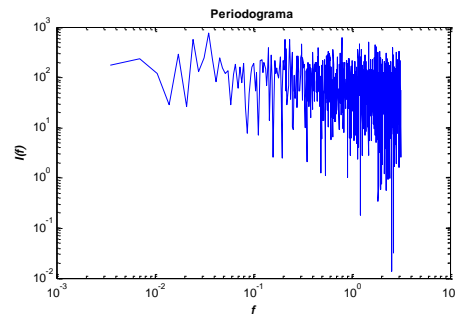
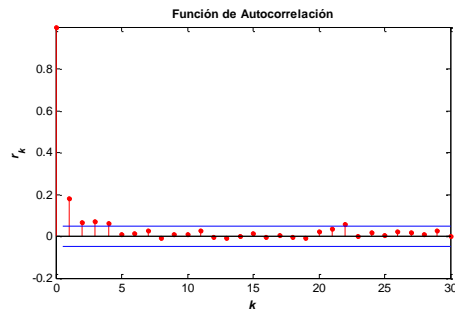
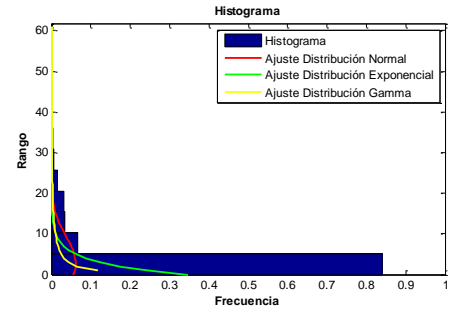
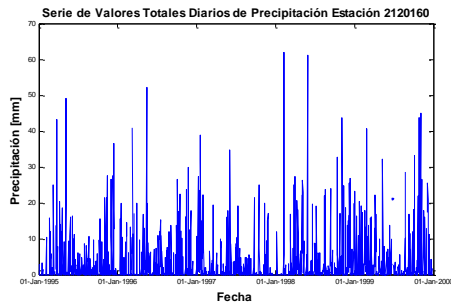
Este	998554
Norte	997786
Cota	2600
Número de Datos	480
Número de Datos Sin Registro	1346
Media	2.5
Mediana	0.0
Máximo	67.4
Desviación Estándar	7.1
Varianza	49.9
Coefficiente de Variación	9.5E-04
Coefficiente de Asimetría	5.2
Coefficiente de Curtosis	35.8

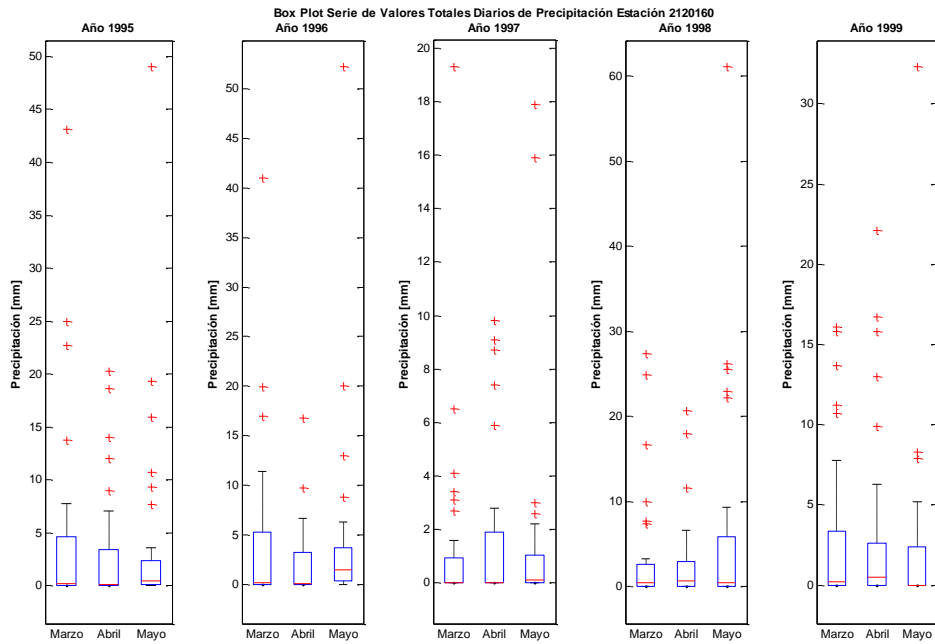
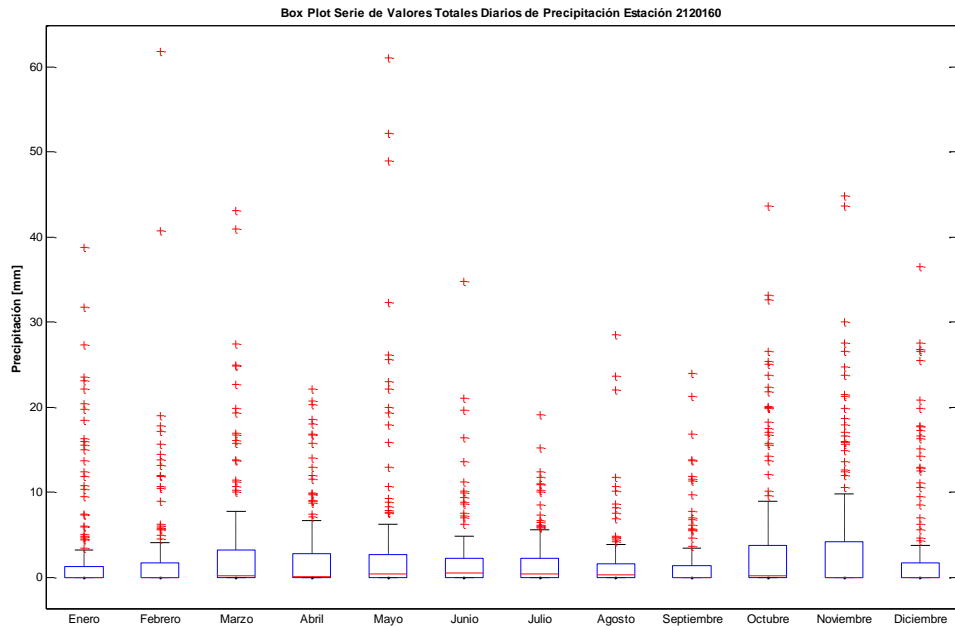


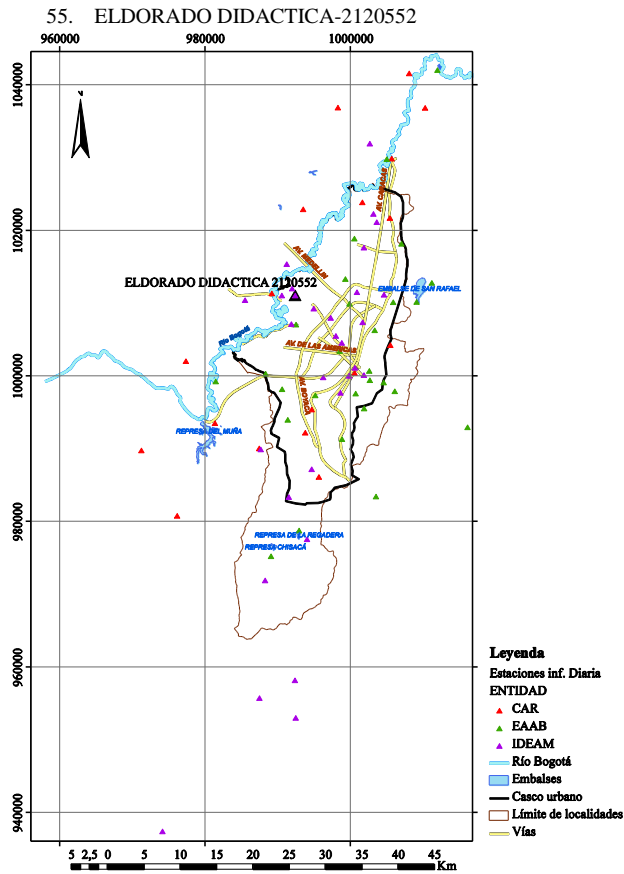




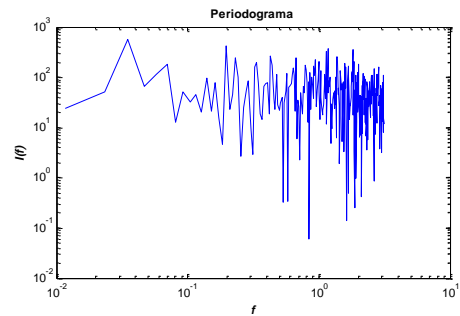
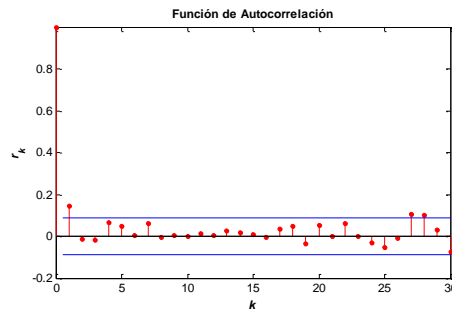
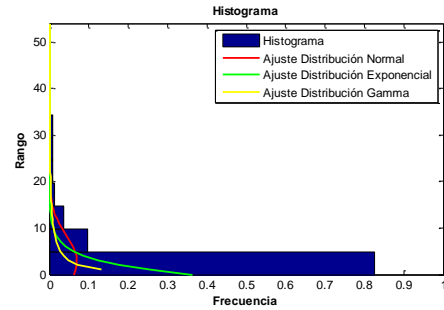
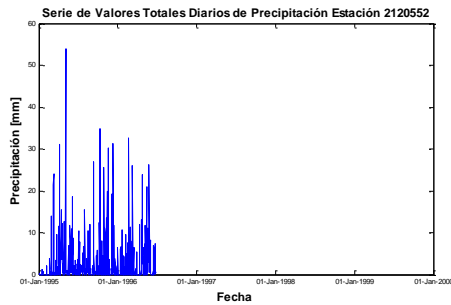
Este	1000513
Norte	1001207
Cota	2685
Número de Datos	1801
Número de Datos Sin Registro	25
Media	2.9
Mediana	0.1
Máximo	61.8
Desviación Estándar	6.5
Varianza	42.3
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	3.7
Coefficiente de Curtosis	21.4

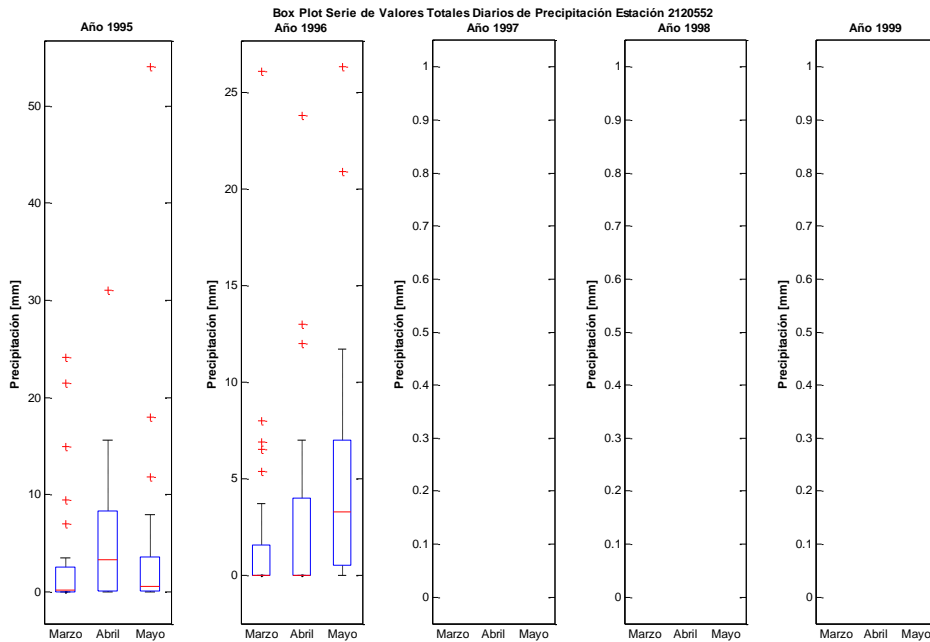
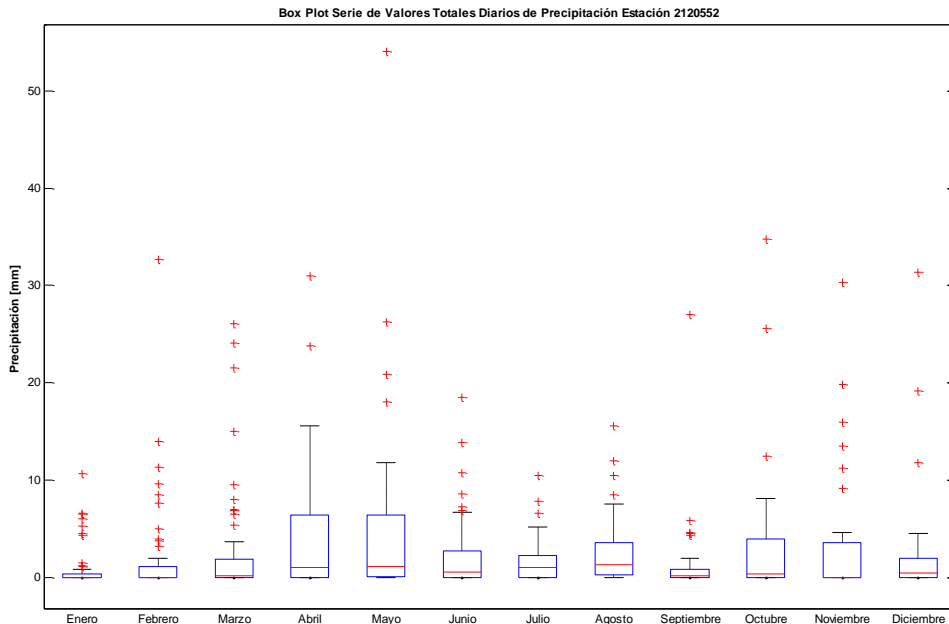


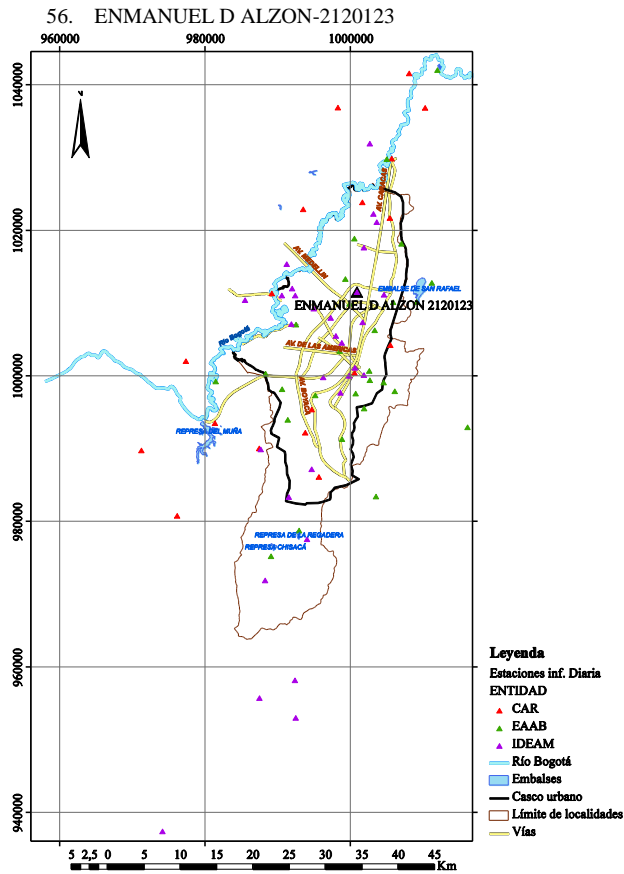




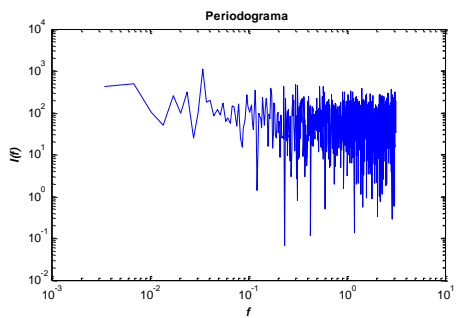
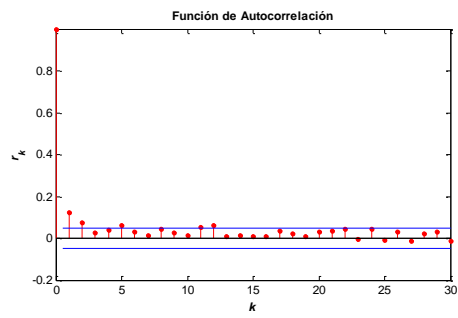
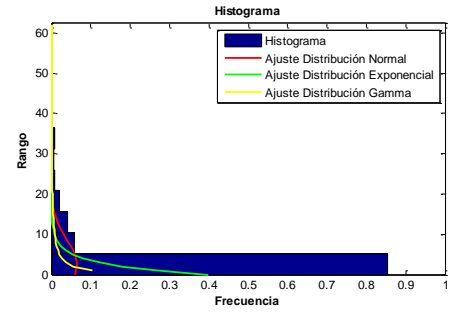
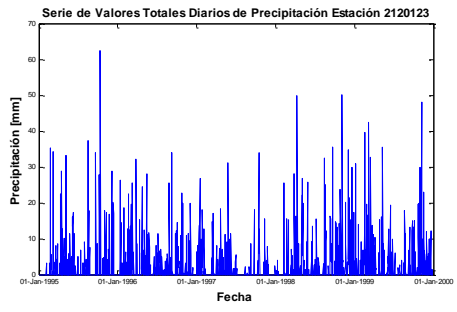
Este	992335
Norte	1011164
Cota	2546
Número de Datos	542
Número de Datos Sin Registro	1284
Media	2.7
Mediana	0.4
Máximo	54.0
Desviación Estándar	5.7
Varianza	32.6
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	3.8
Coefficiente de Curtosis	22.8

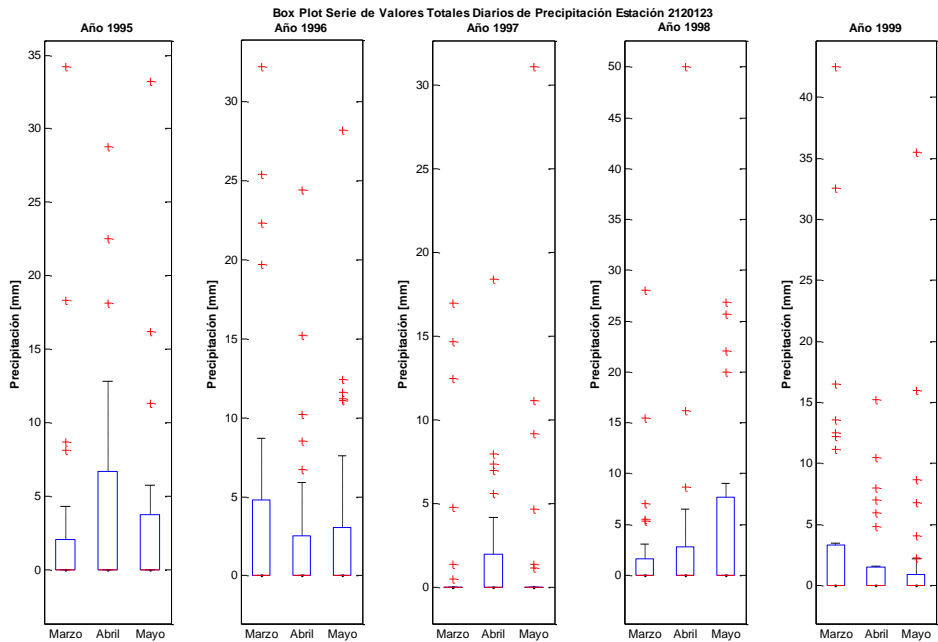
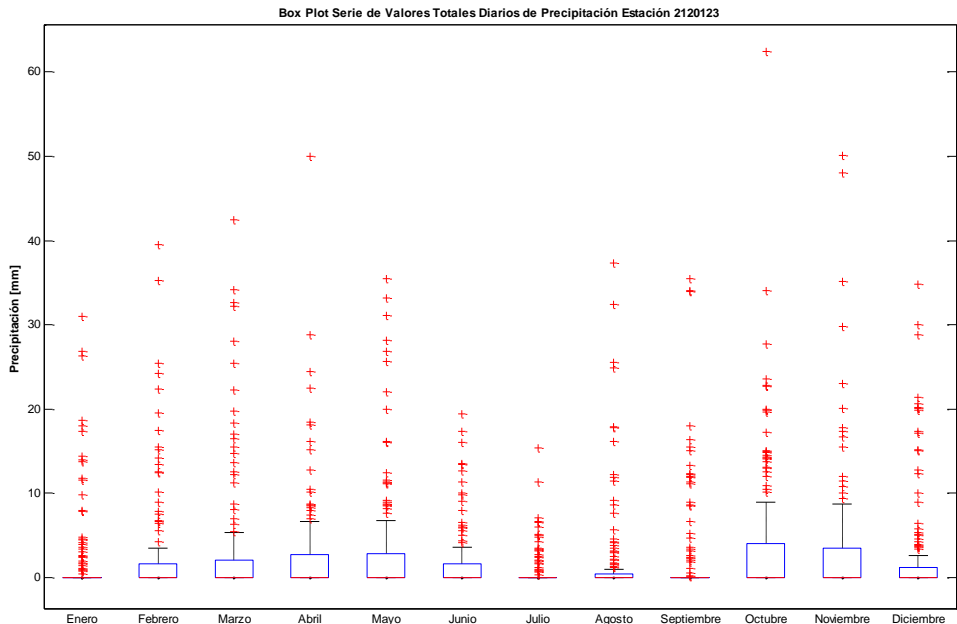


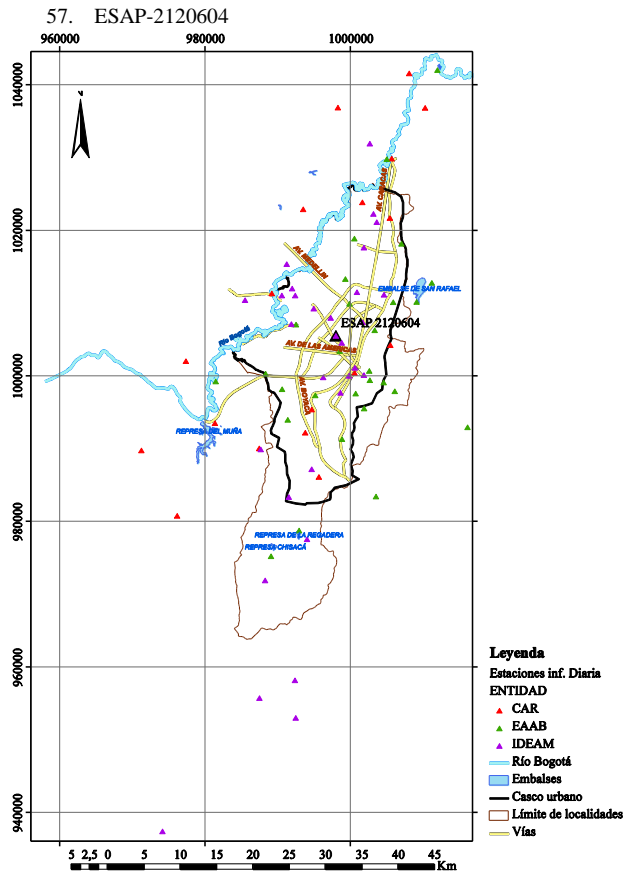




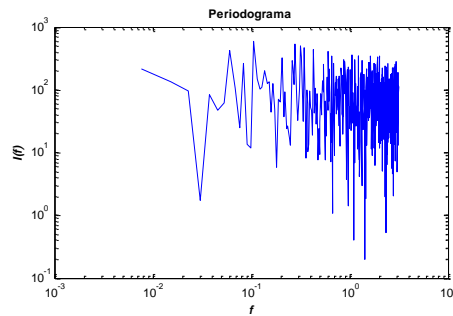
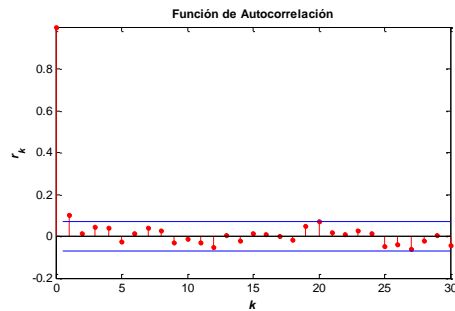
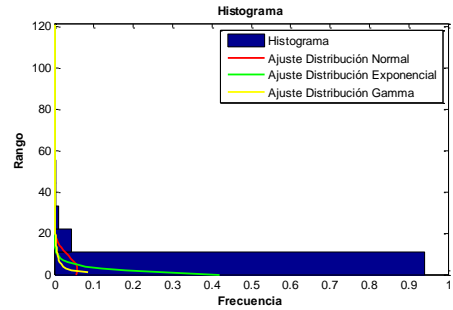
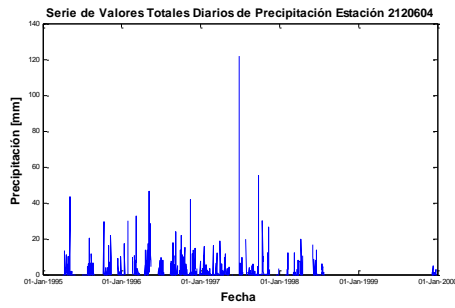
Este	1000799
Norte	1011603
Cota	2520
Número de Datos	1826
Número de Datos Sin Registro	0
Media	2.5
Mediana	0.0
Máximo	62.4
Desviación Estándar	6.3
Varianza	39.1
Coefficiente de Variación	1.0E-03
Coefficiente de Asimetría	3.8
Coefficiente de Curtosis	21.0

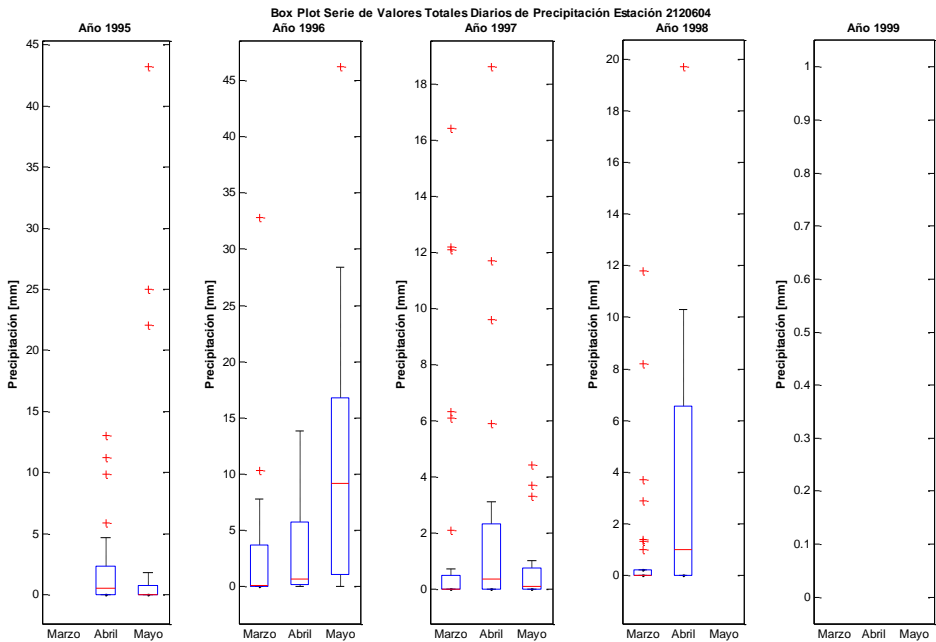
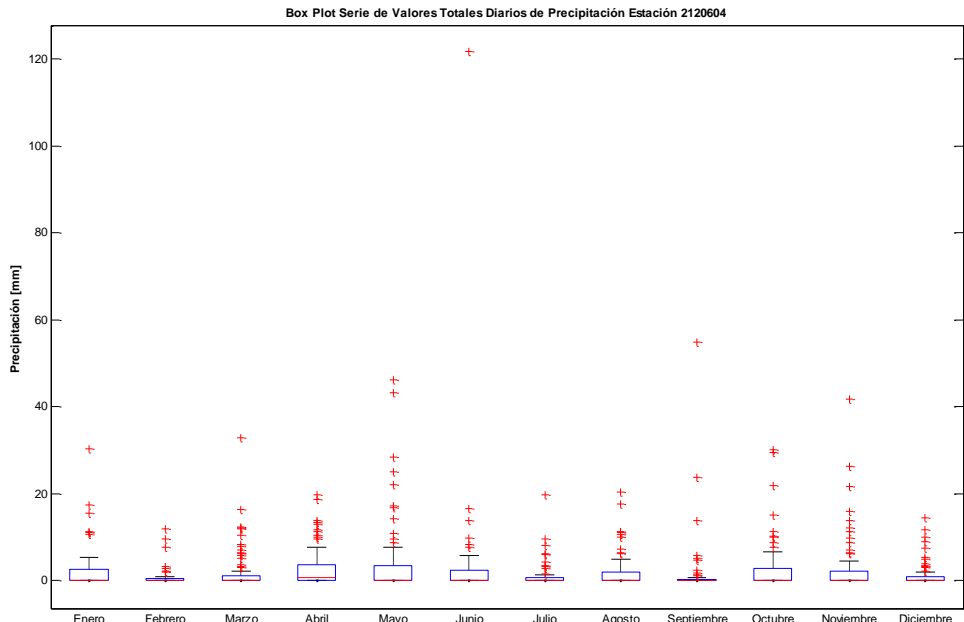


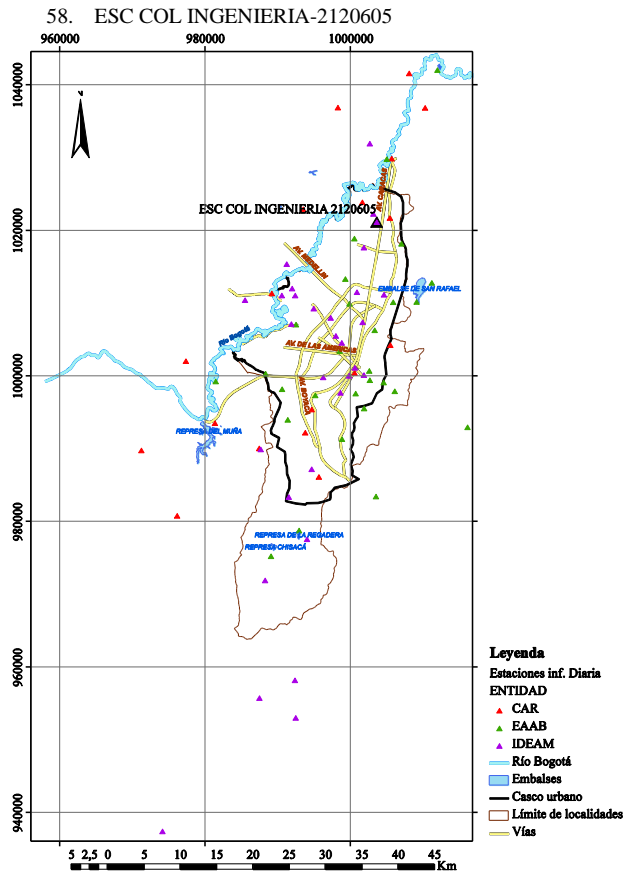




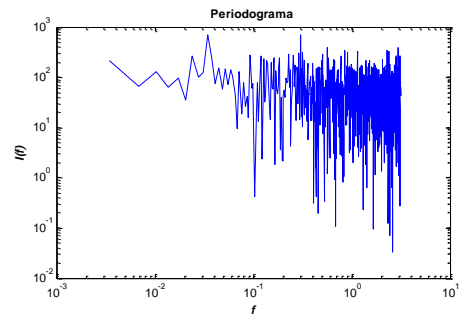
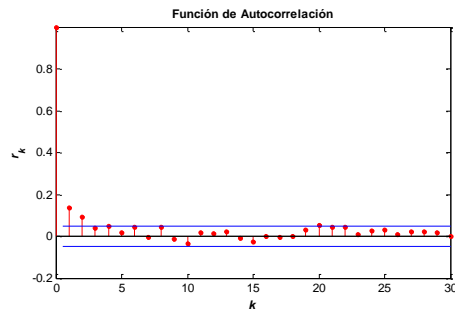
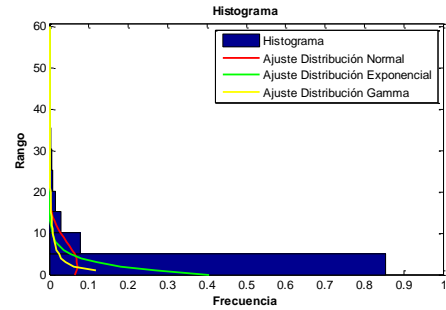
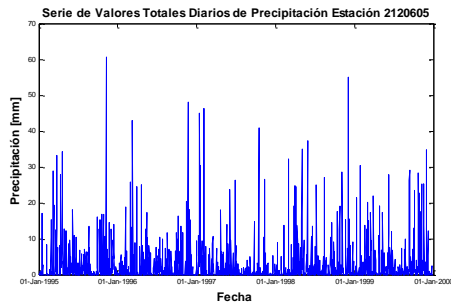
Este	997908
Norte	1005593
Cota	2553
Número de Datos	828
Número de Datos Sin Registro	998
Media	2.4
Mediana	0.0
Máximo	121.7
Desviación Estándar	6.9
Varianza	47.3
Coefficiente de Variación	9.4E-04
Coefficiente de Asimetría	8.5
Coefficiente de Curtosis	121.2

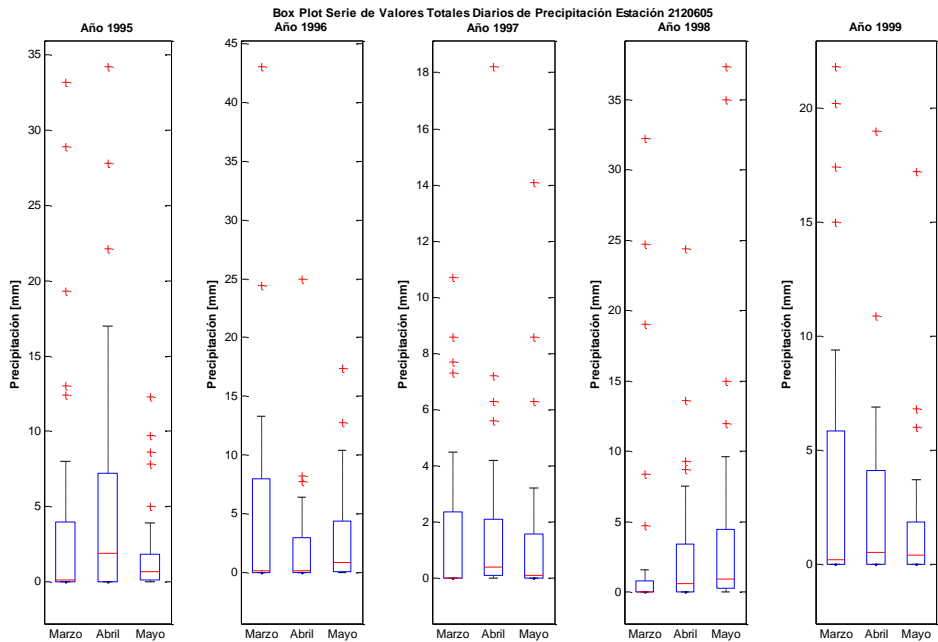
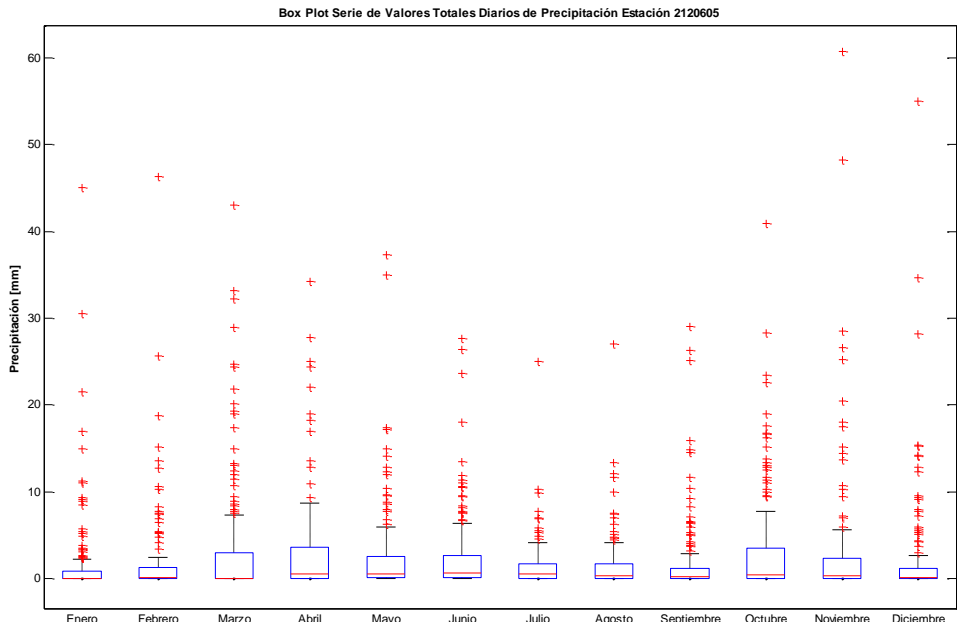


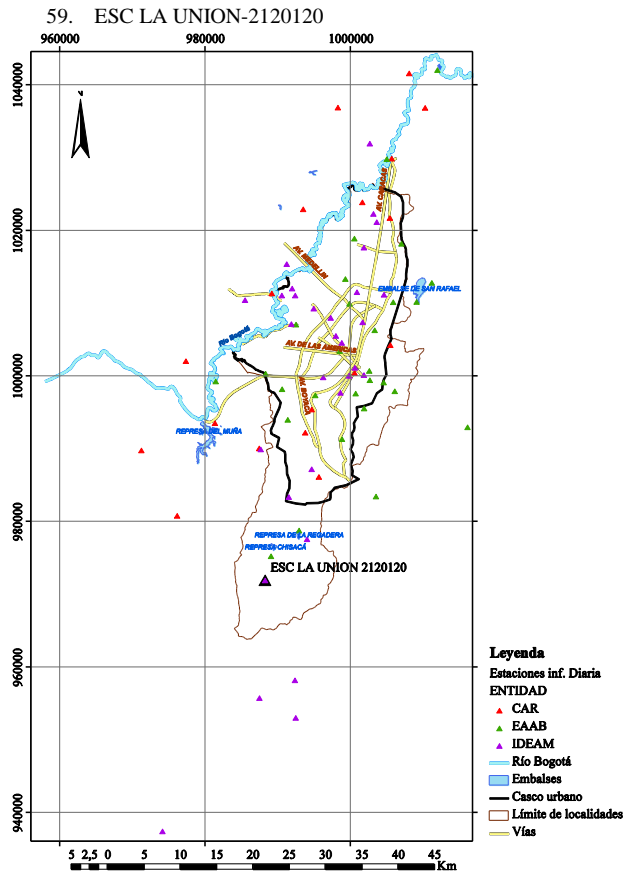




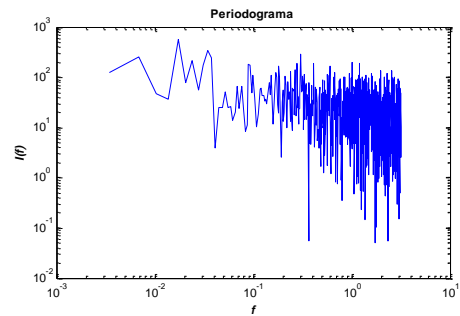
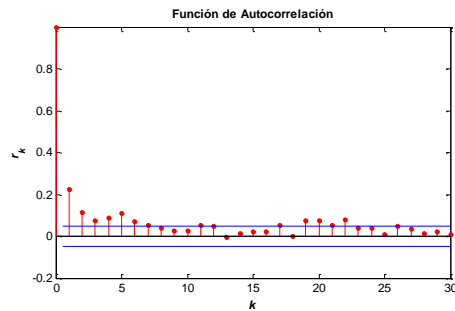
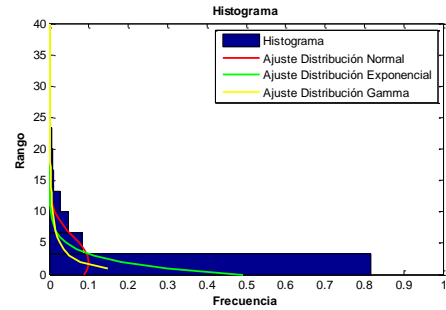
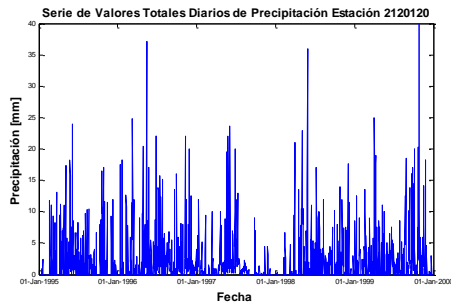
Este	1003545
Norte	1021246
Cota	2650
Número de Datos	1826
Número de Datos Sin Registro	0
Media	2.5
Mediana	0.2
Máximo	60.7
Desviación Estándar	5.7
Varianza	32.2
Coefficiente de Variación	9.3E-04
Coefficiente de Asimetría	4.3
Coefficiente de Curtosis	28.0

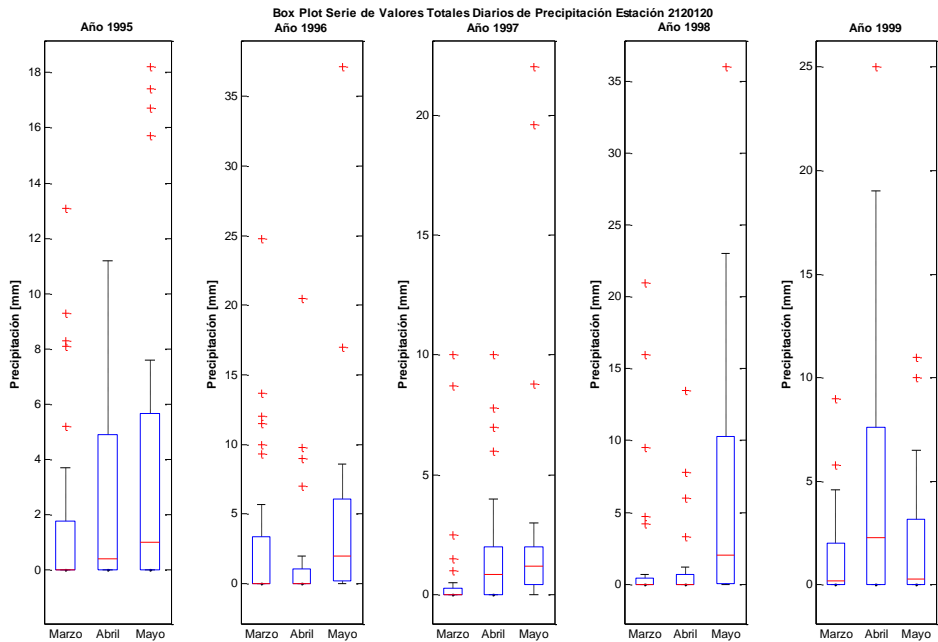
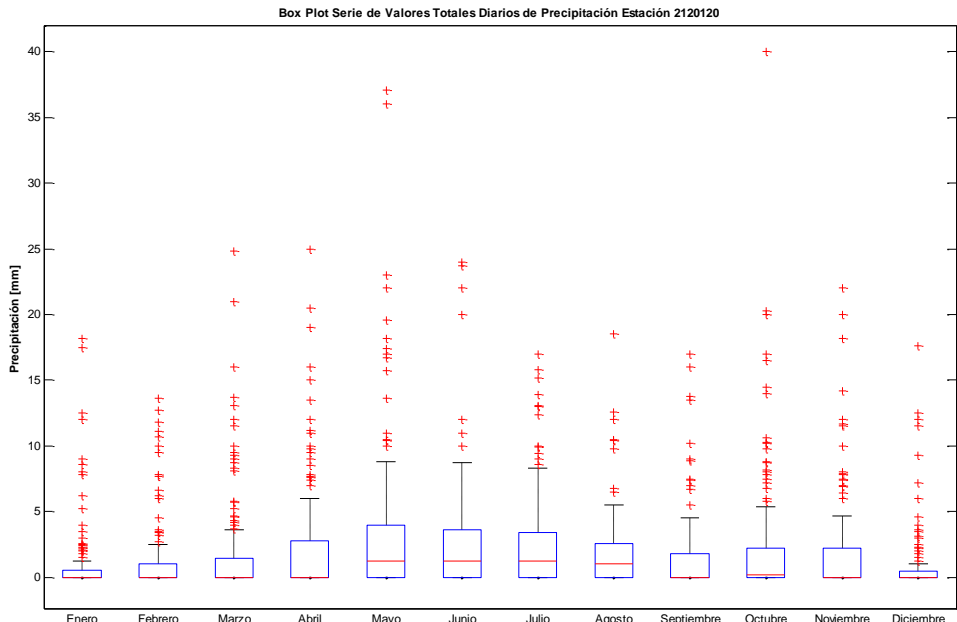


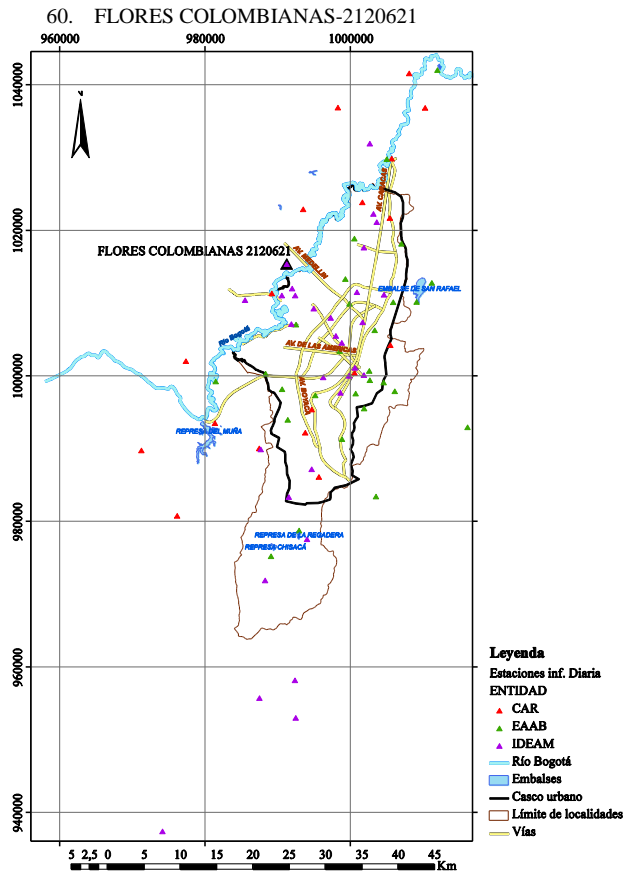




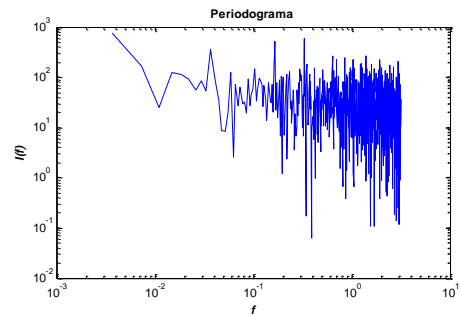
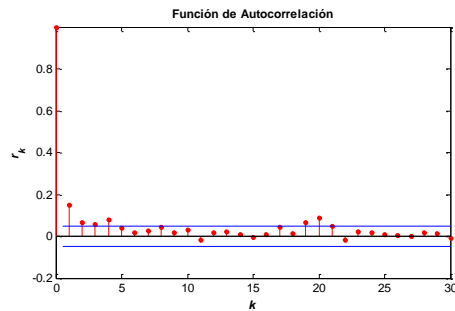
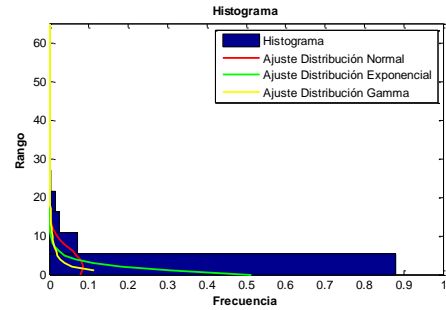
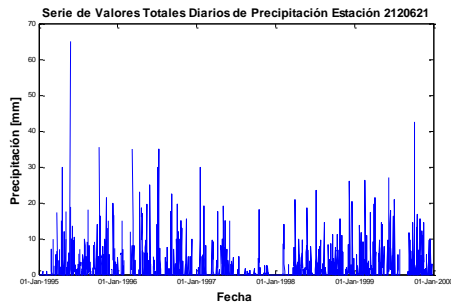
Este	988191
Norte	971996
Cota	3320
Número de Datos	1825
Número de Datos Sin Registro	1
Media	2.0
Mediana	0.1
Máximo	40.0
Desviación Estándar	4.0
Varianza	15.9
Coefficiente de Variación	6.1E-04
Coefficiente de Asimetría	3.5
Coefficiente de Curtosis	19.9

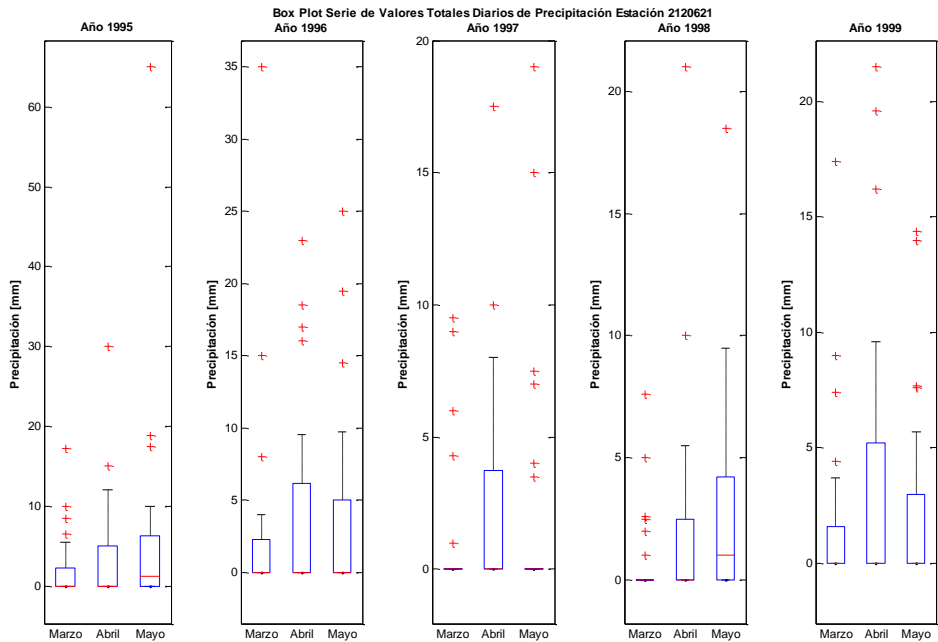
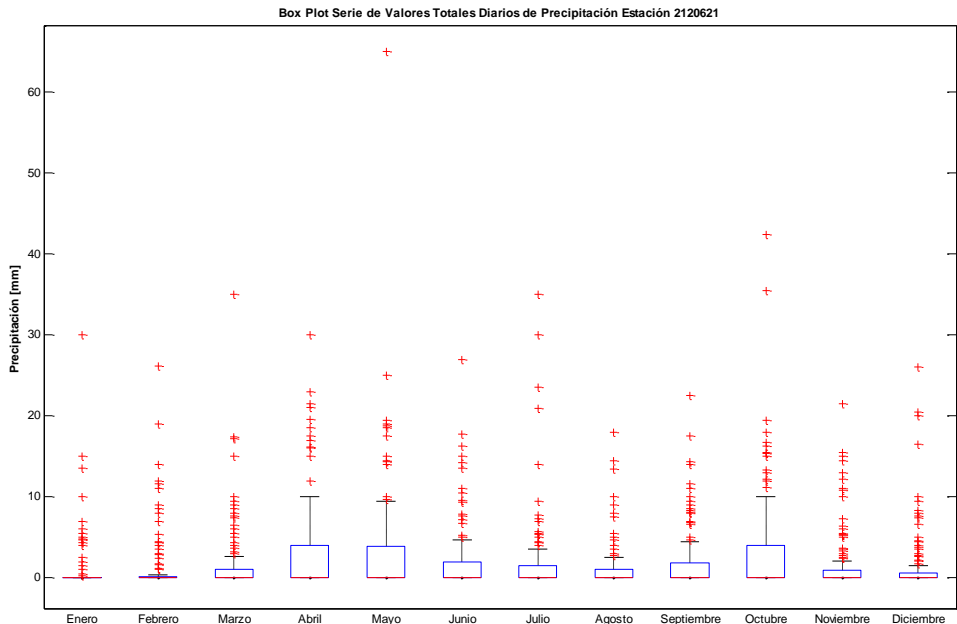


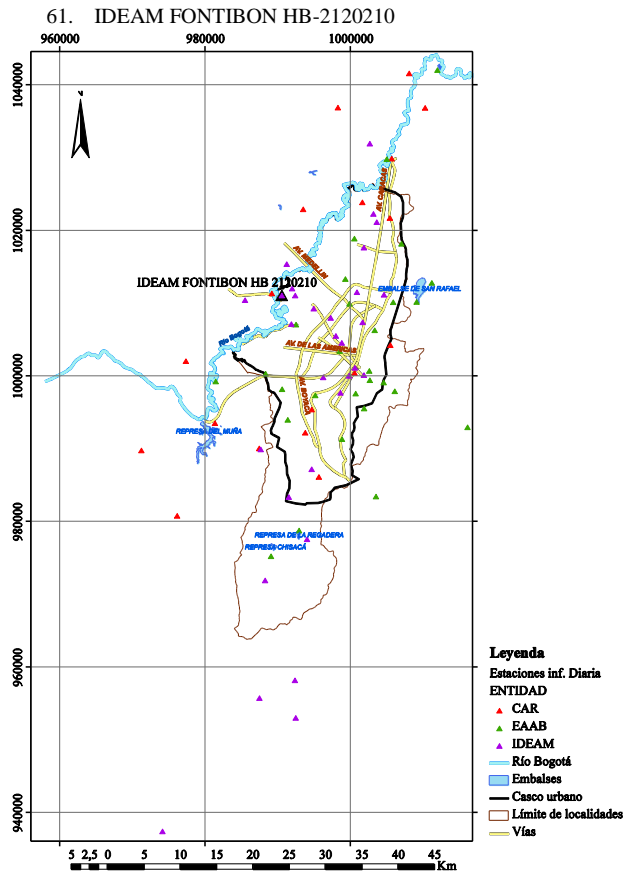




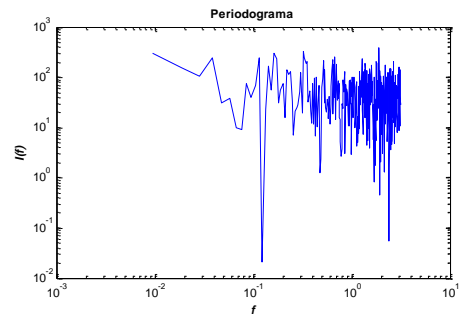
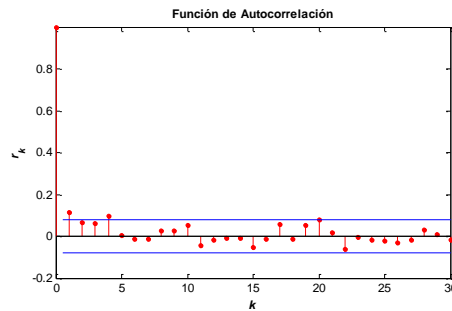
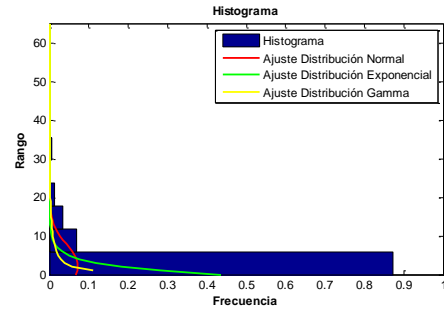
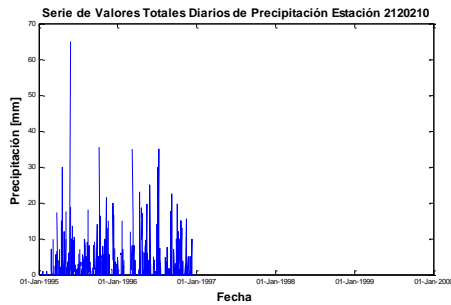
Este	991144
Norte	1015487
Cota	2560
Número de Datos	1706
Número de Datos Sin Registro	120
Media	1.9
Mediana	0.0
Máximo	65.0
Desviación Estándar	4.7
Varianza	21.9
Coefficiente de Variación	7.6E-04
Coefficiente de Asimetría	4.5
Coefficiente de Curtosis	35.6

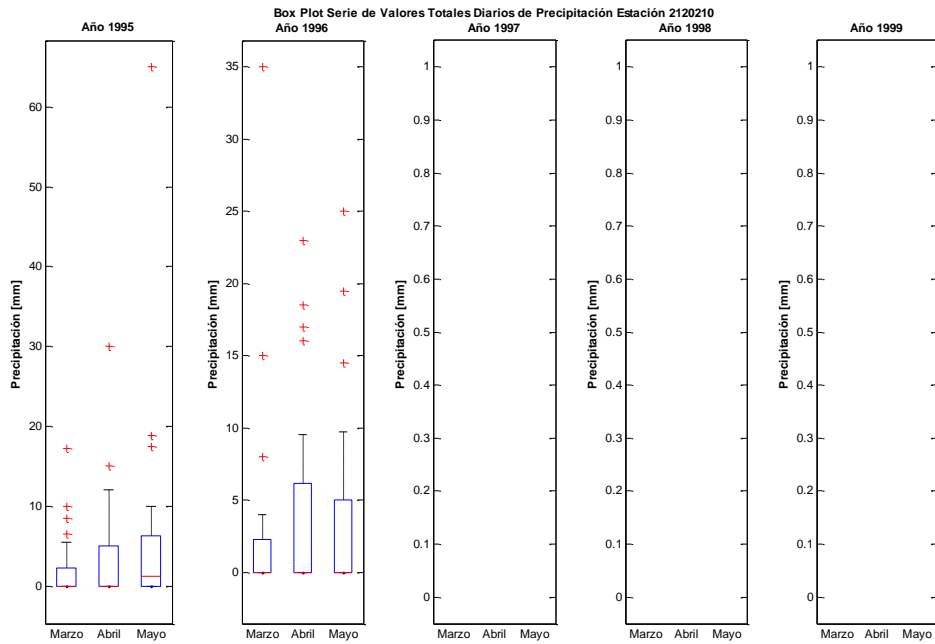
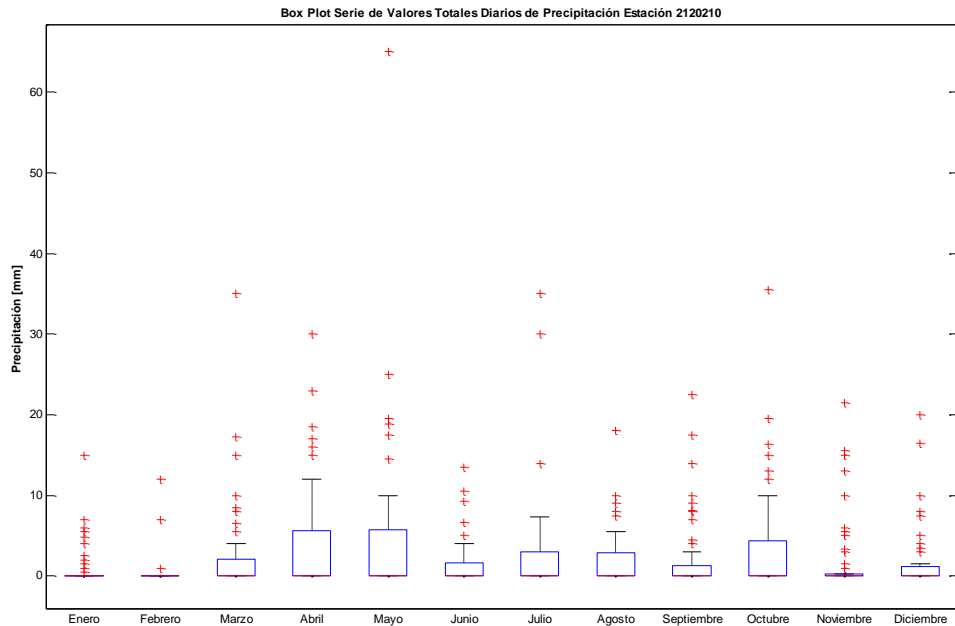


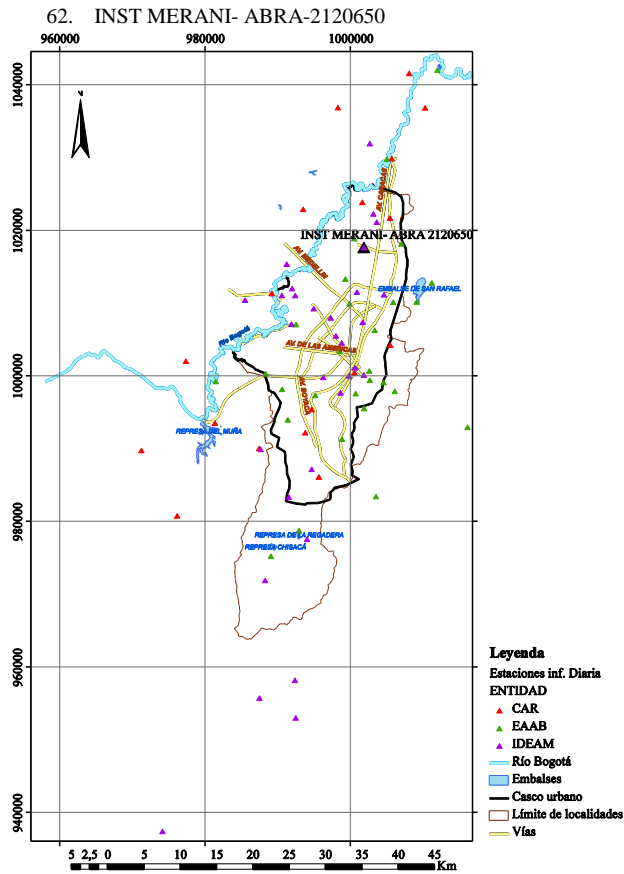




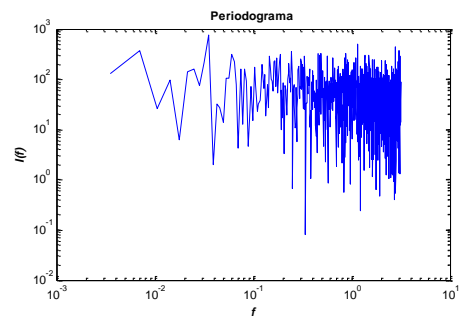
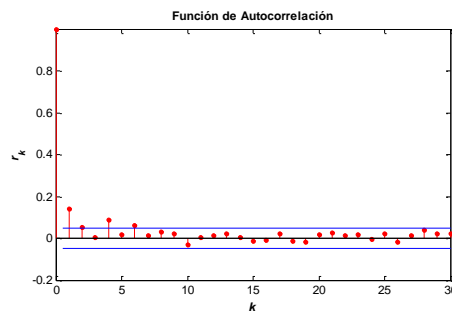
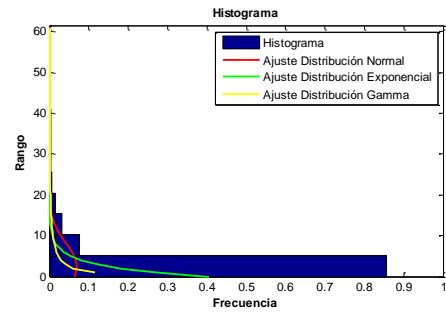
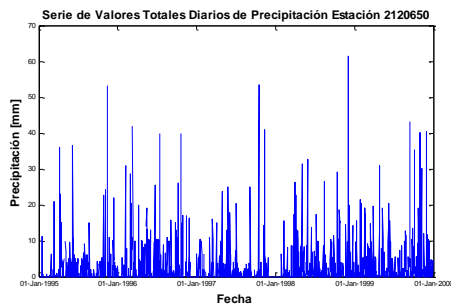
Este	990486
Norte	1011164
Cota	2511
Número de Datos	659
Número de Datos Sin Registro	1167
Media	2.3
Mediana	0.0
Máximo	65.0
Desviación Estándar	5.5
Varianza	30.3
Coefficiente de Variación	9.1E-04
Coefficiente de Asimetría	4.7
Coefficiente de Curtosis	36.7

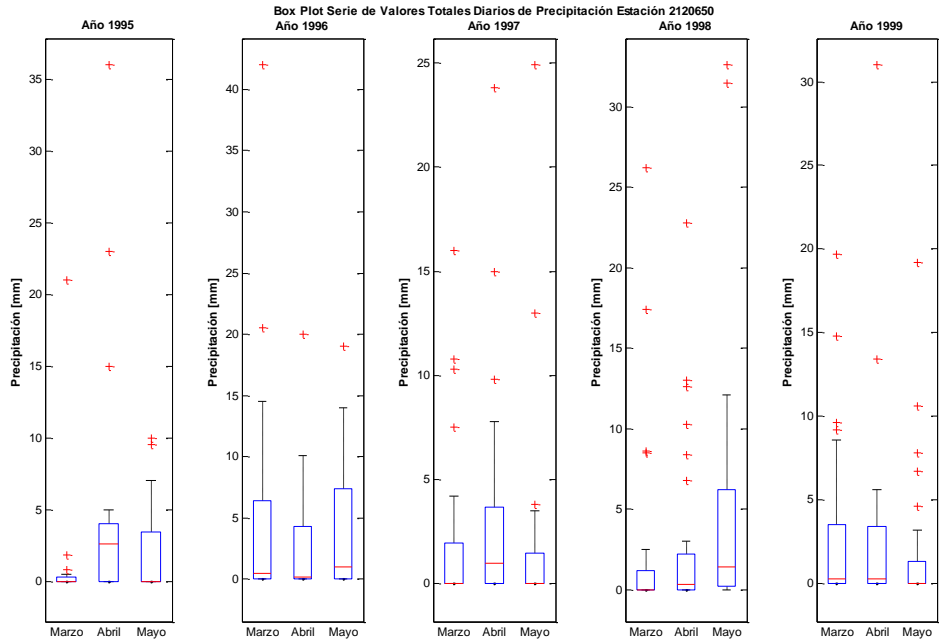
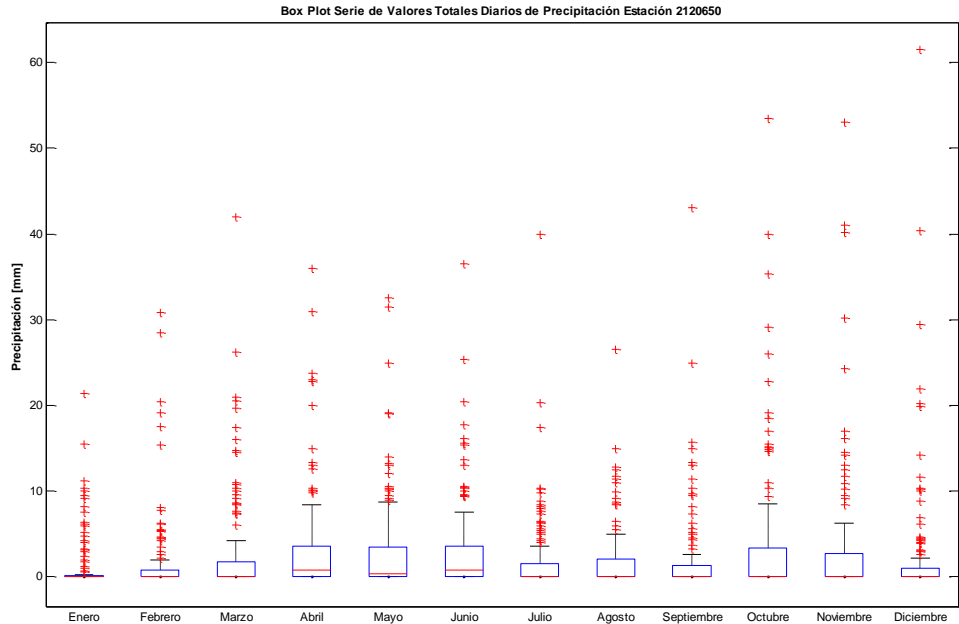


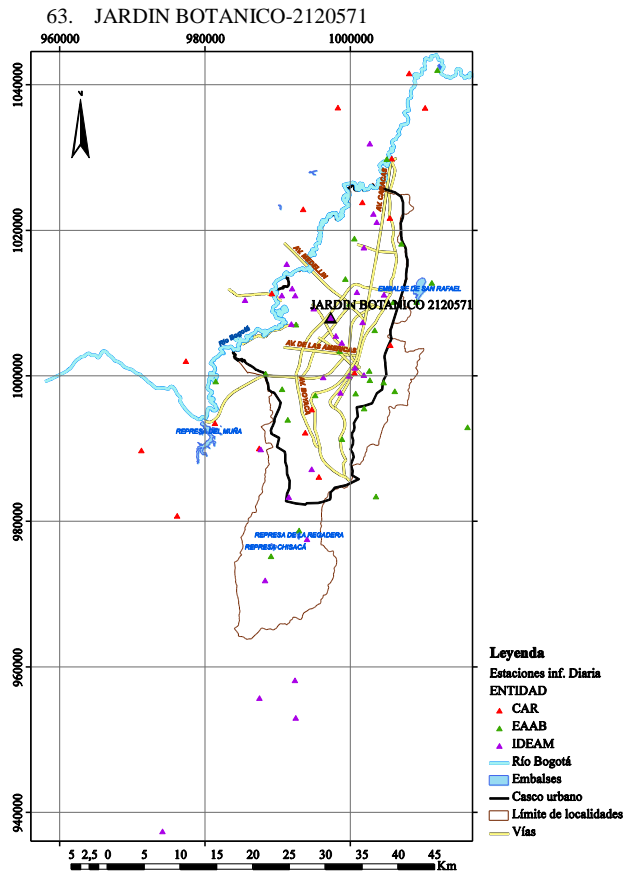




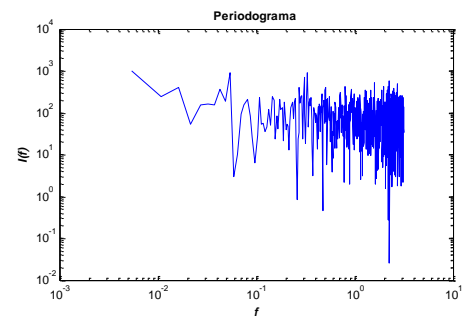
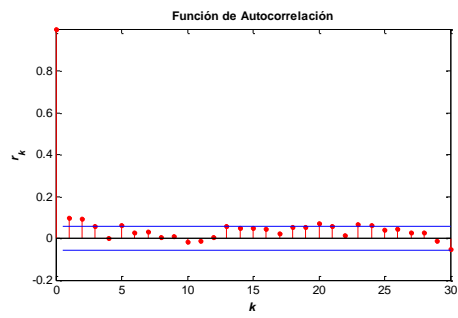
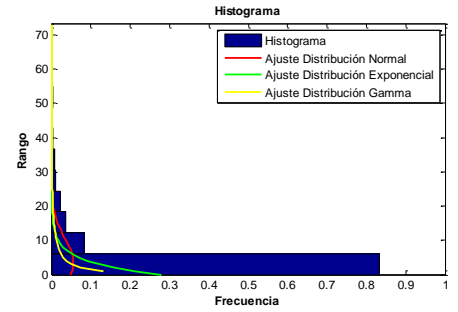
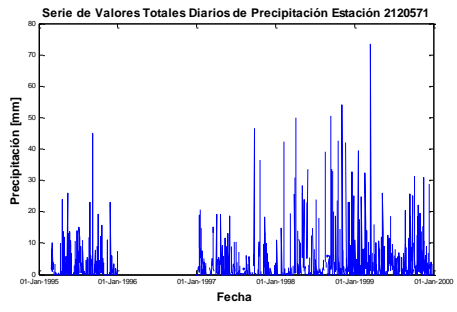
Este	1001767
Norte	1017742
Cota	2570
Número de Datos	1774
Número de Datos Sin Registro	52
Media	2.5
Mediana	0.0
Máximo	61.5
Desviación Estándar	5.8
Varianza	33.1
Coefficiente de Variación	9.6E-04
Coefficiente de Asimetría	4.4
Coefficiente de Curtosis	28.4

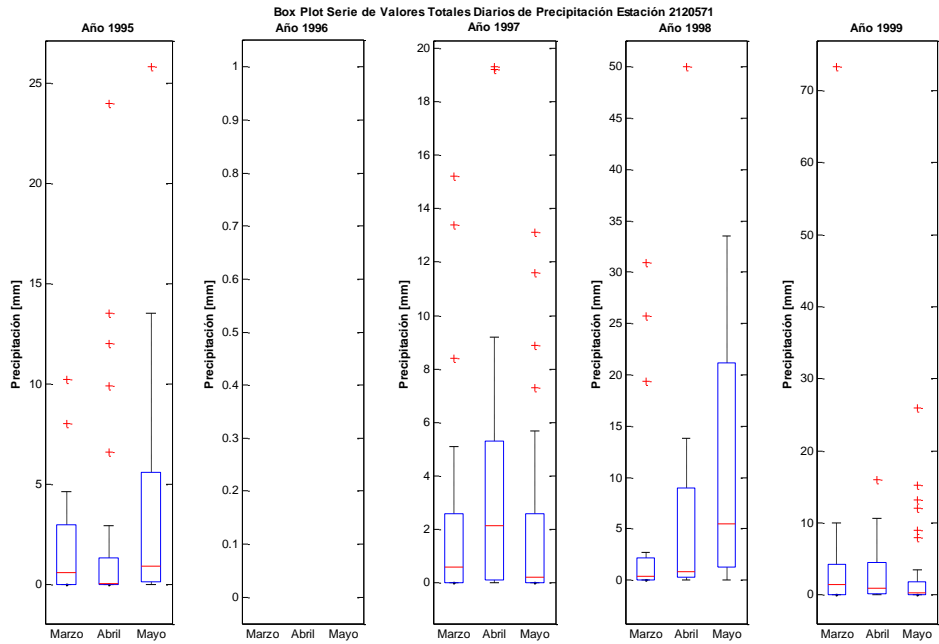
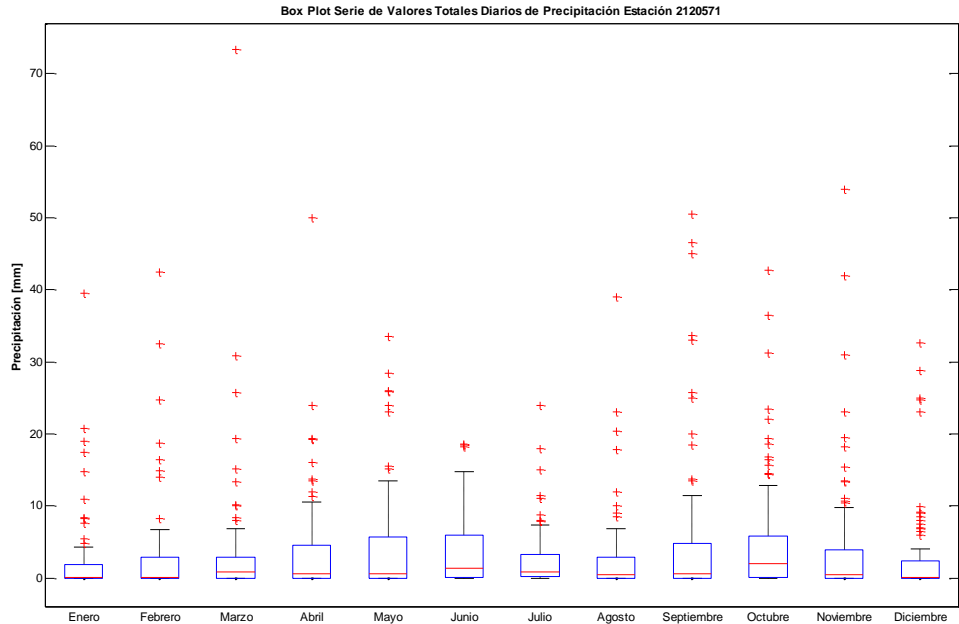




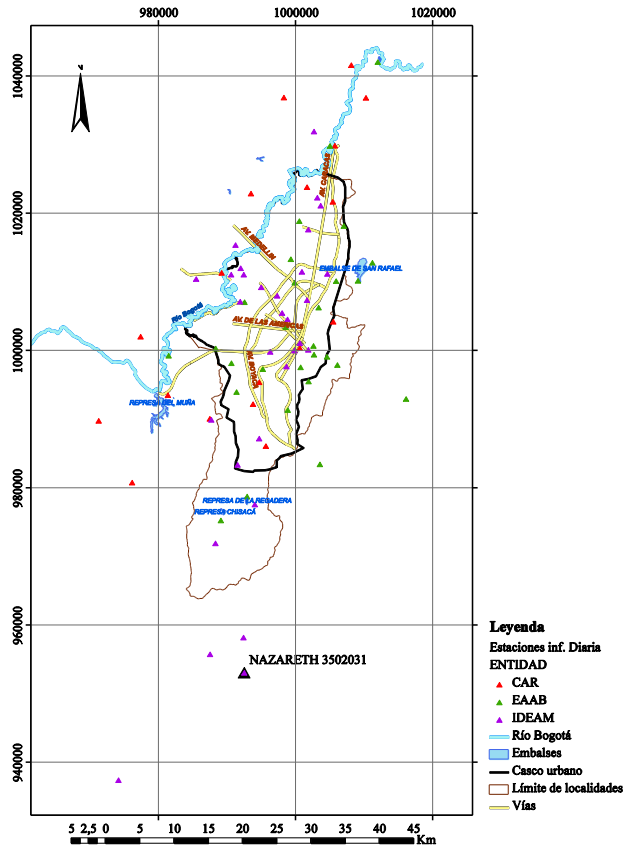


Este	997209
Norte	1008087
Cota	2552
Número de Datos	1167
Número de Datos Sin Registro	659
Media	3.6
Mediana	0.6
Máximo	73.3
Desviación Estándar	7.2
Varianza	51.5
Coefficiente de Variación	1.4E-03
Coefficiente de Asimetría	3.8
Coefficiente de Curtosis	22.2

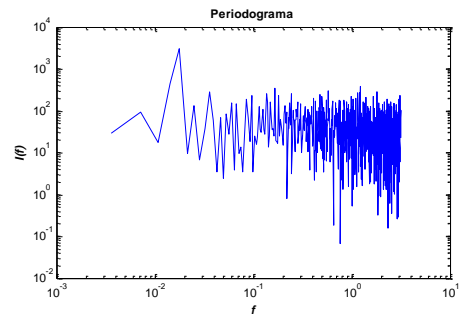
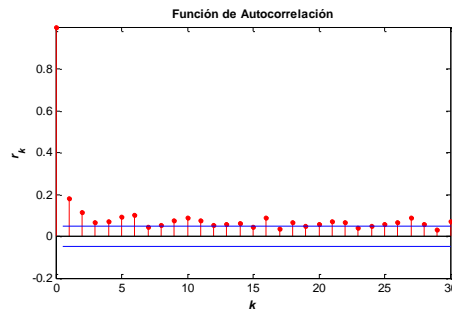
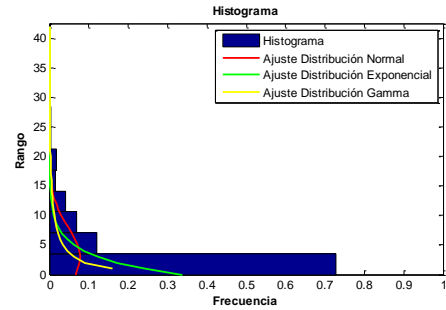
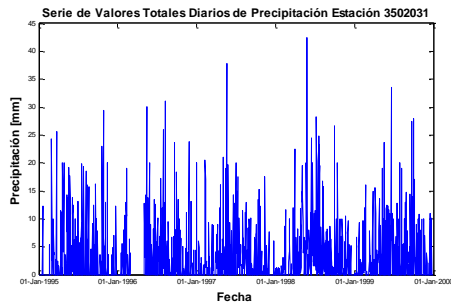


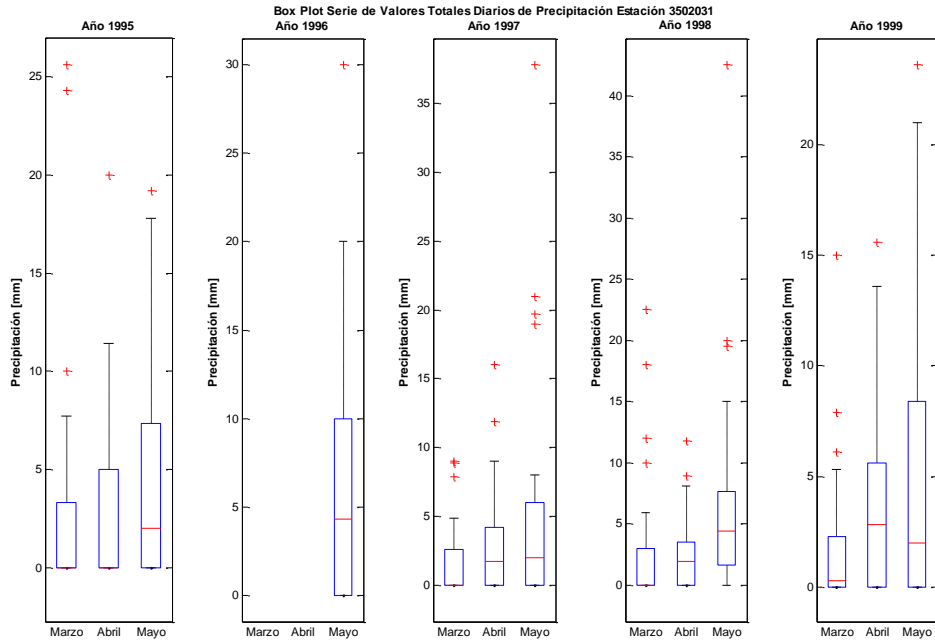
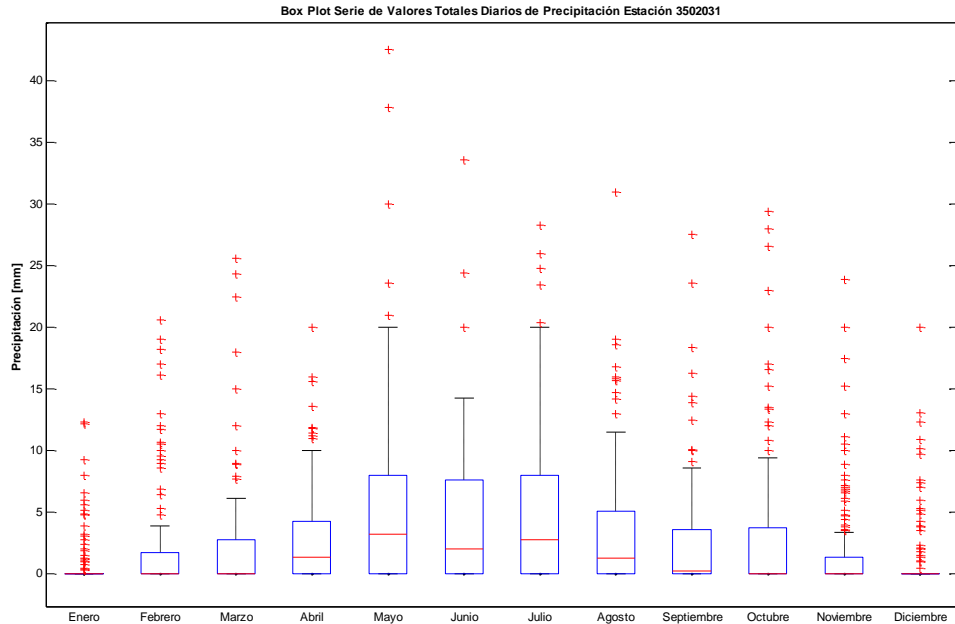


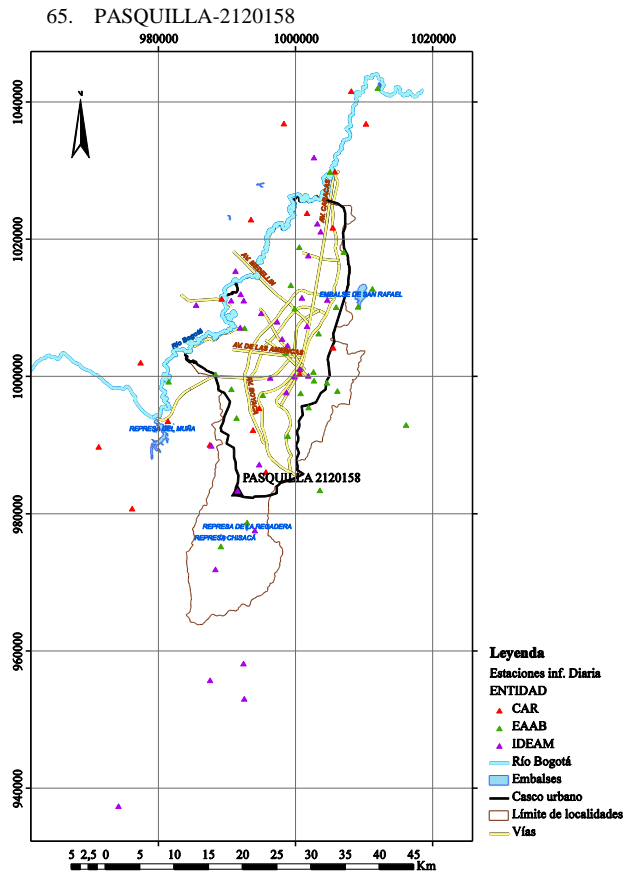
64. NAZARETH-3502031



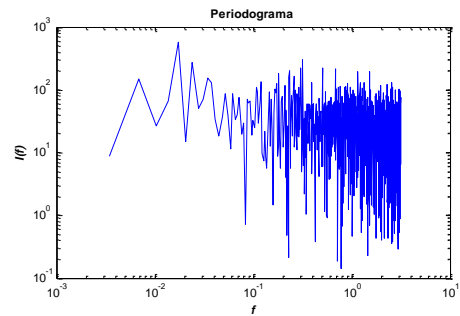
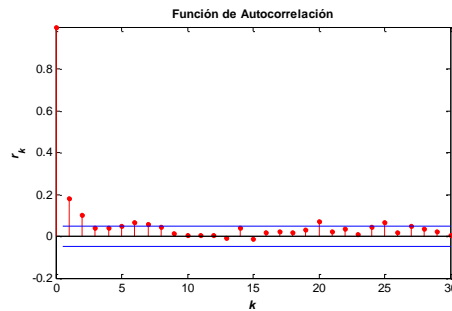
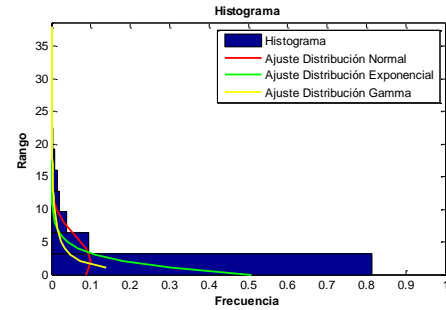
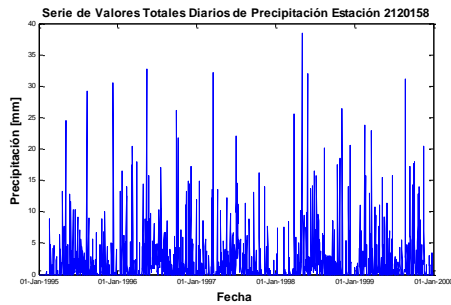
Este	992396
Norte	953120
Cota	2800
Número de Datos	1765
Número de Datos Sin Registro	61
Media	3.0
Mediana	0.0
Máximo	42.5
Desviación Estándar	5.1
Varianza	26.4
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	2.6
Coefficiente de Curtosis	11.3

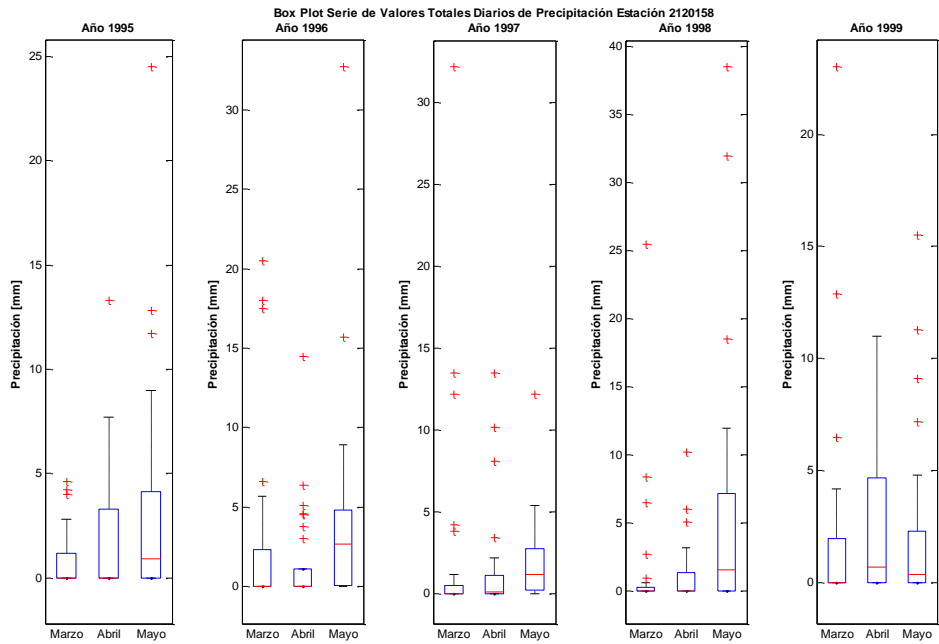
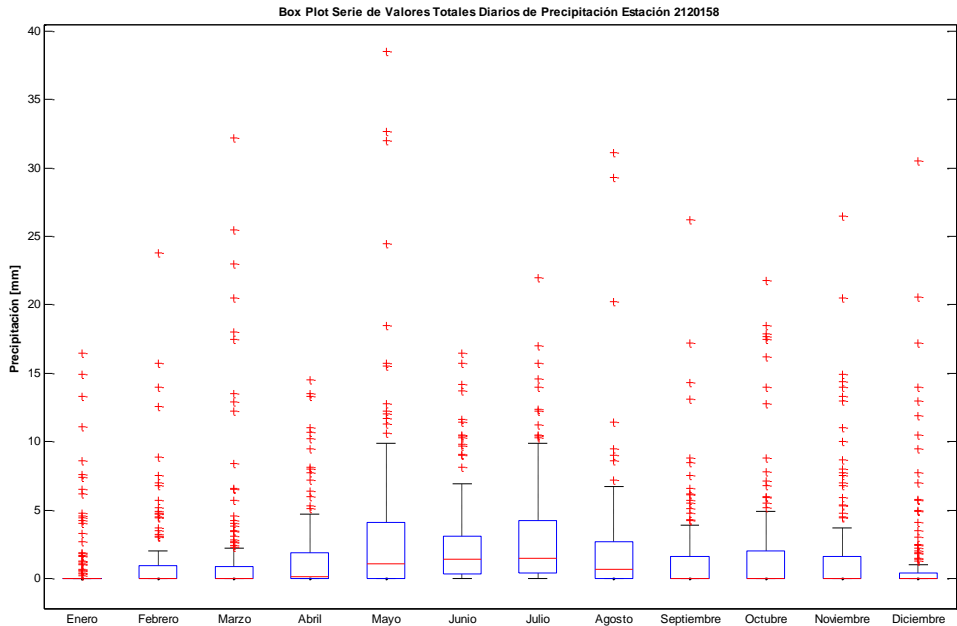




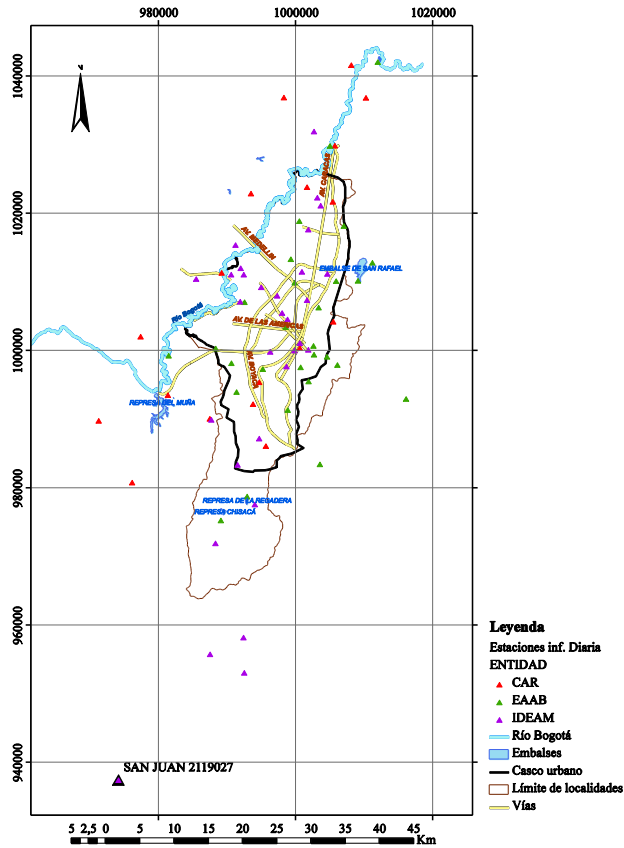


Este	991418
Norte	983446
Cota	3000
Número de Datos	1825
Número de Datos Sin Registro	1
Media	2.0
Mediana	0.0
Máximo	38.5
Desviación Estándar	4.1
Varianza	16.4
Coefficiente de Variación	6.6E-04
Coefficiente de Asimetría	3.7
Coefficiente de Curtosis	21.3

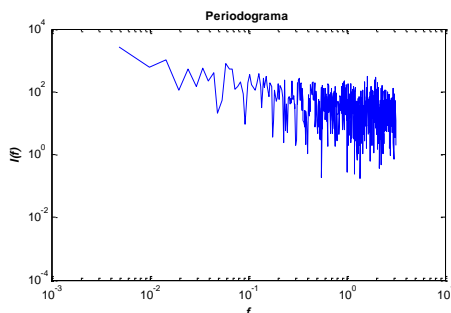
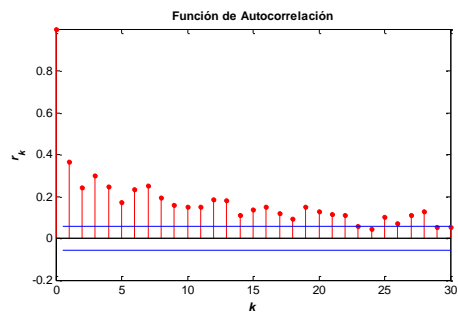
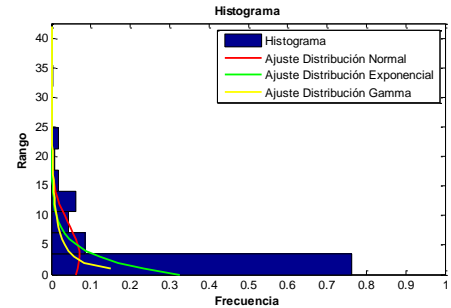
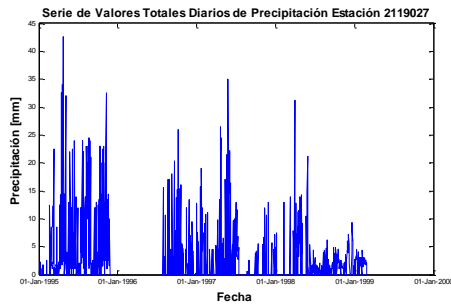


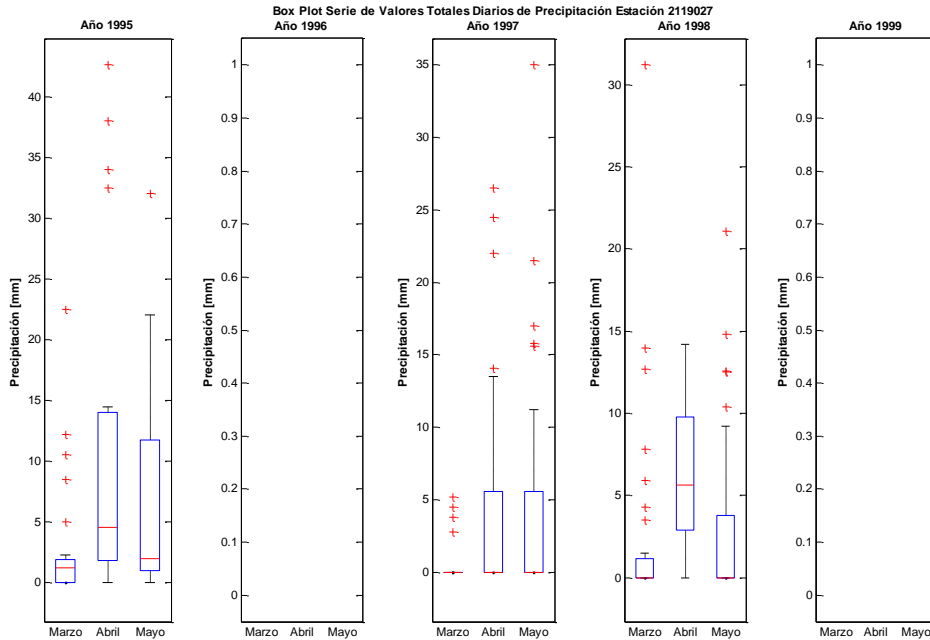
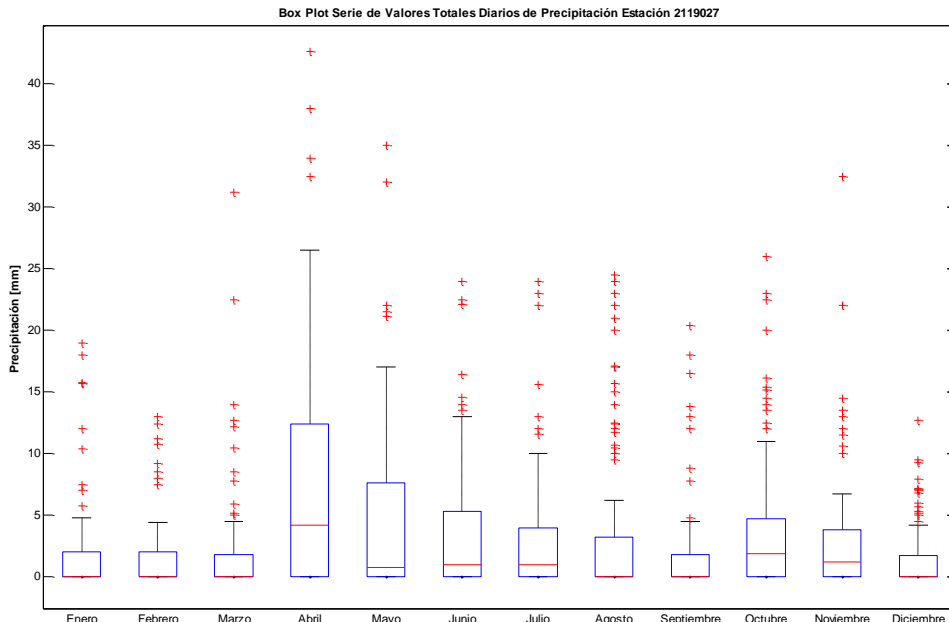


66. SAN JUAN-2119027

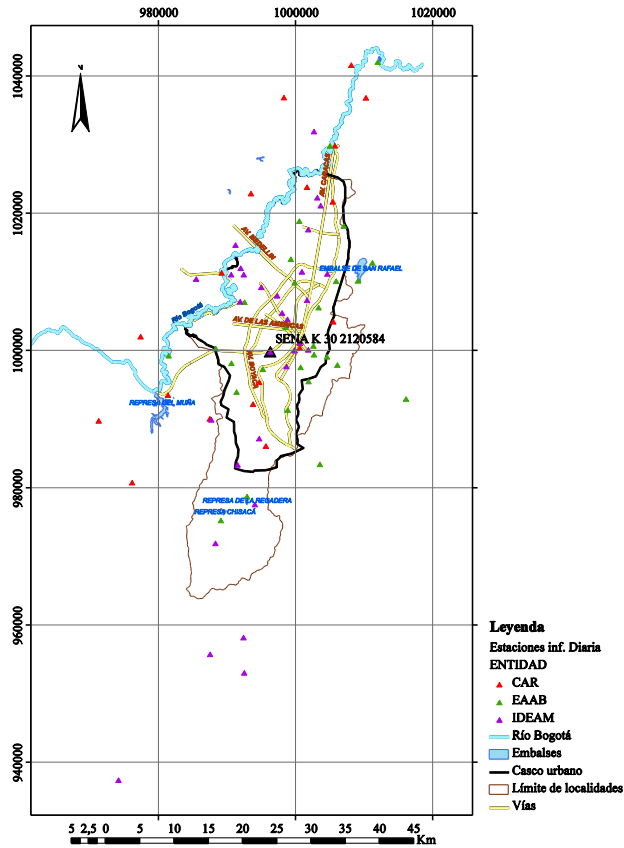


Este	974053
Norte	937507
Cota	2900
Número de Datos	1280
Número de Datos Sin Registro	546
Media	3.1
Mediana	0.0
Máximo	42.6
Desviación Estándar	5.6
Varianza	31.3
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	2.7
Coefficiente de Curtosis	11.7

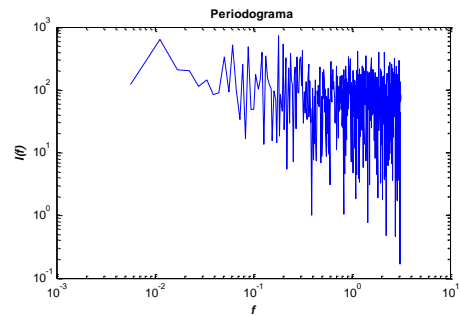
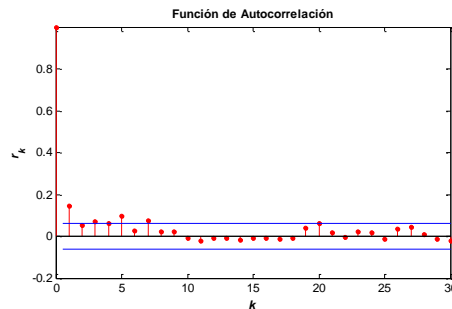
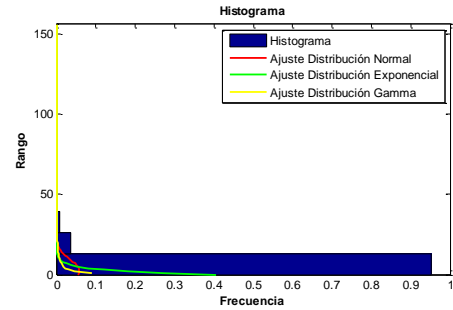
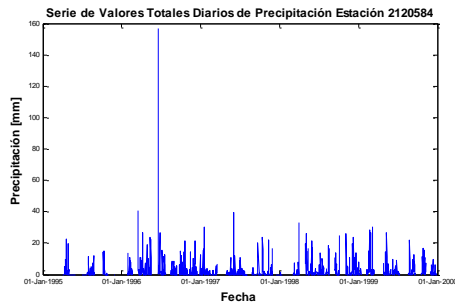


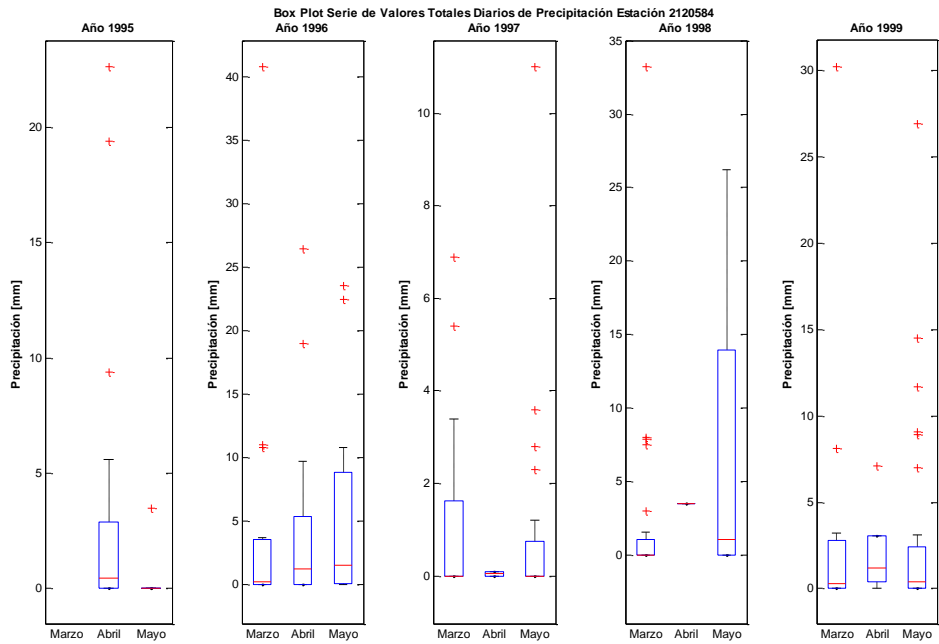
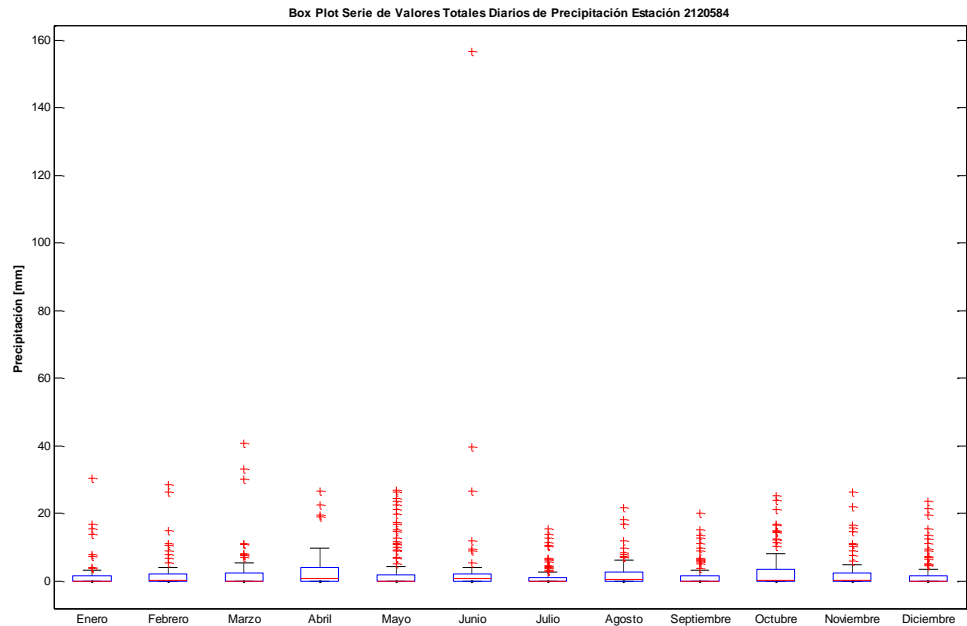


67. SENA K 30-2120584

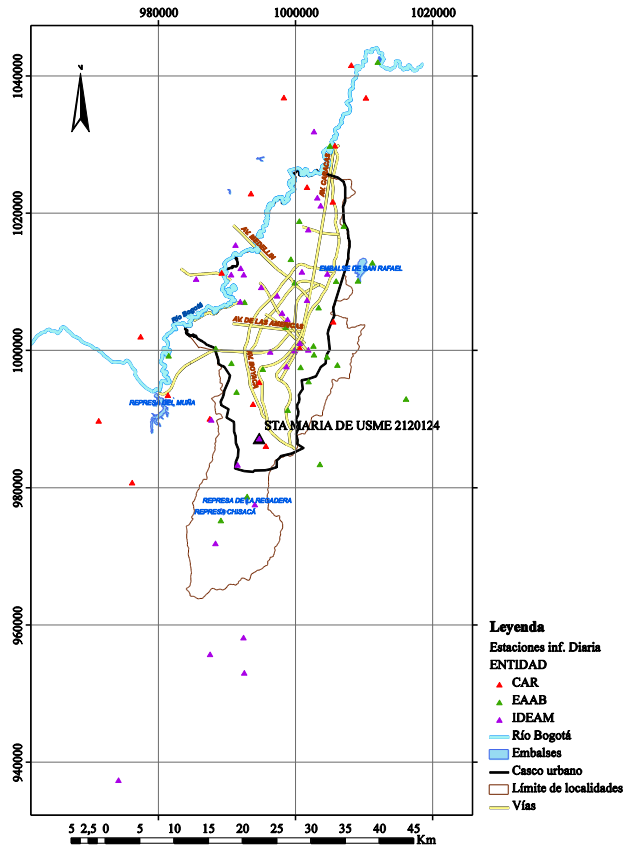


Este	996191
Norte	999907
Cota	2553
Número de Datos	1122
Número de Datos Sin Registro	704
Media	2.5
Mediana	0.1
Máximo	156.6
Desviación Estándar	6.8
Varianza	46.5
Coefficiente de Variación	9.6E-04
Coefficiente de Asimetría	11.6
Coefficiente de Curtosis	237.6

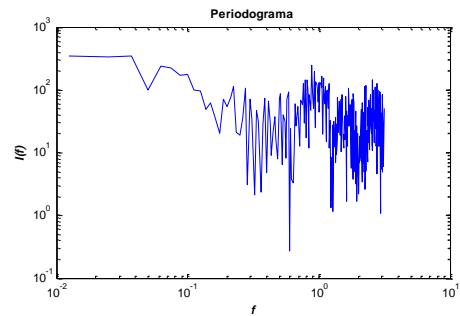
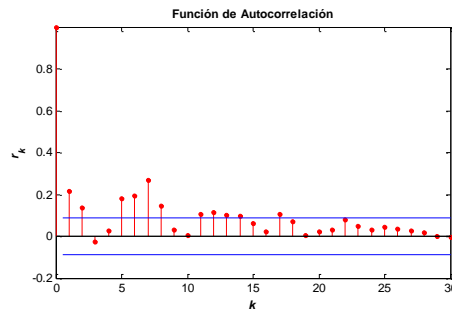
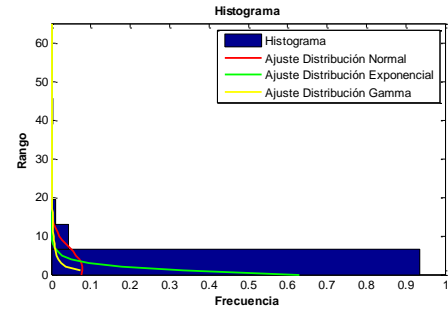
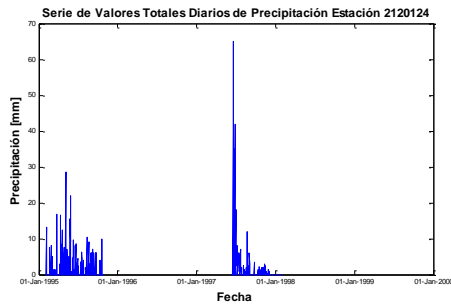


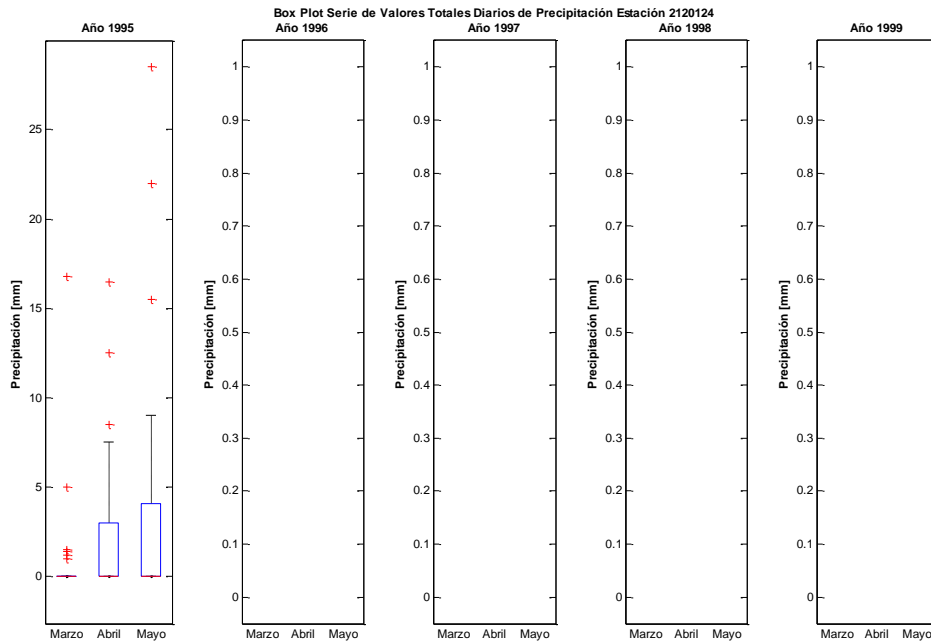
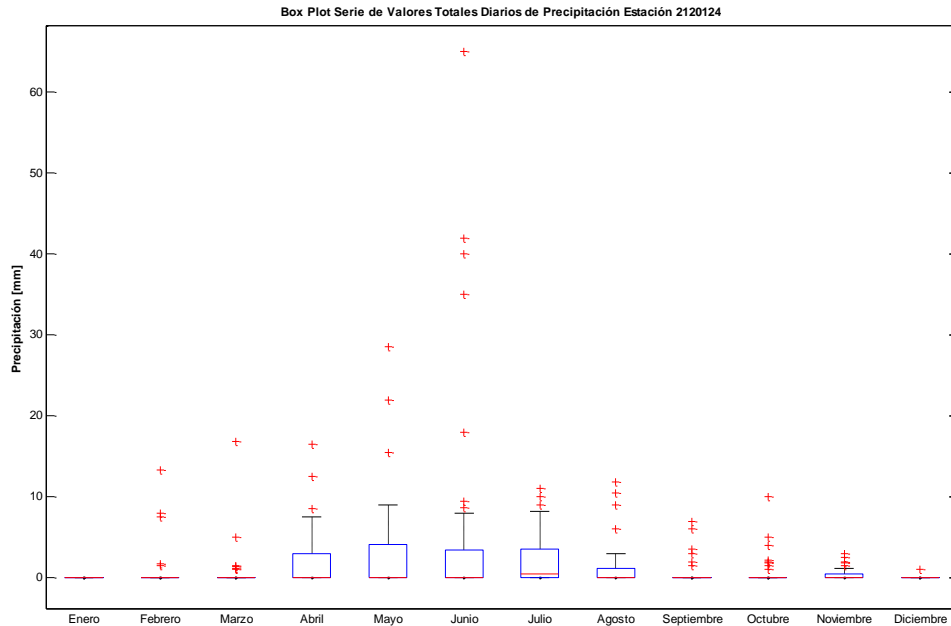


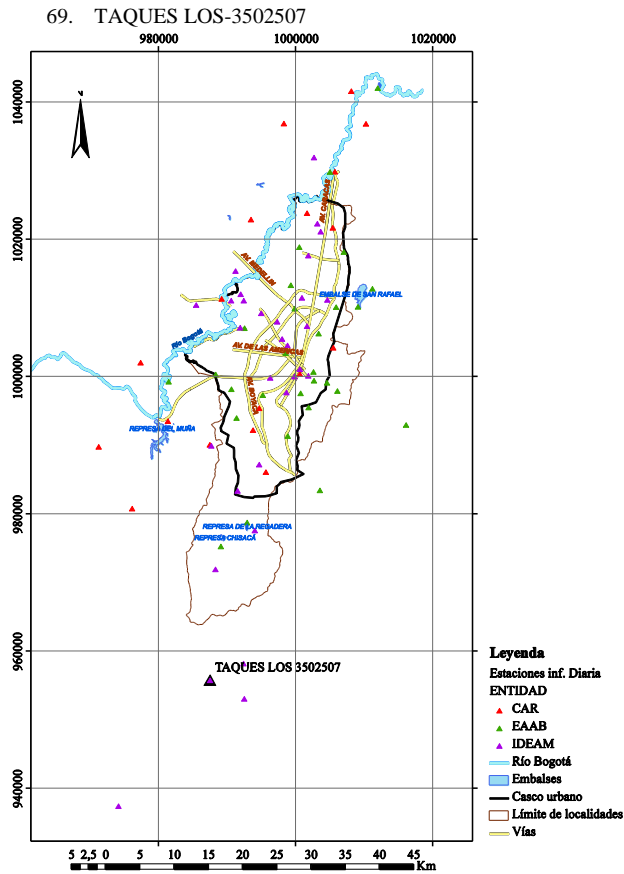
68. STA MARIA DE USME-2120124



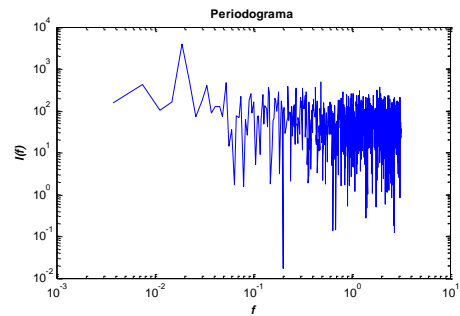
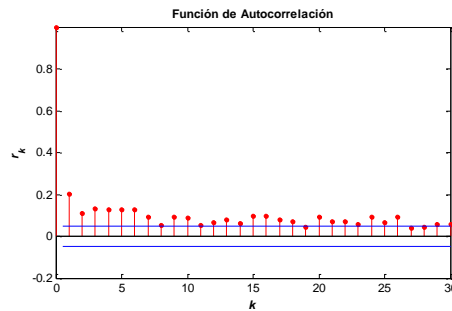
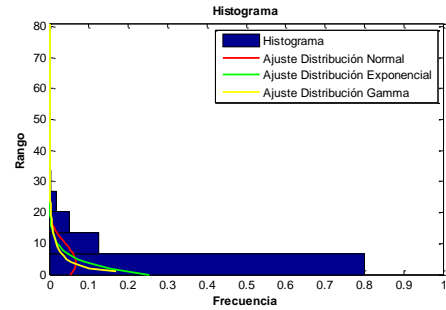
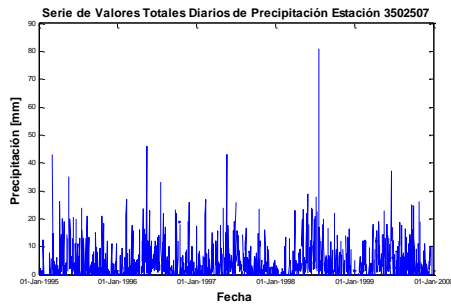
Este	994588
Norte	987296
Cota	2800
Número de Datos	505
Número de Datos Sin Registro	1321
Media	1.6
Mediana	0.0
Máximo	65.0
Desviación Estándar	5.1
Varianza	25.7
Coefficiente de Variación	5.7E-04
Coefficiente de Asimetría	7.1
Coefficiente de Curtosis	69.8

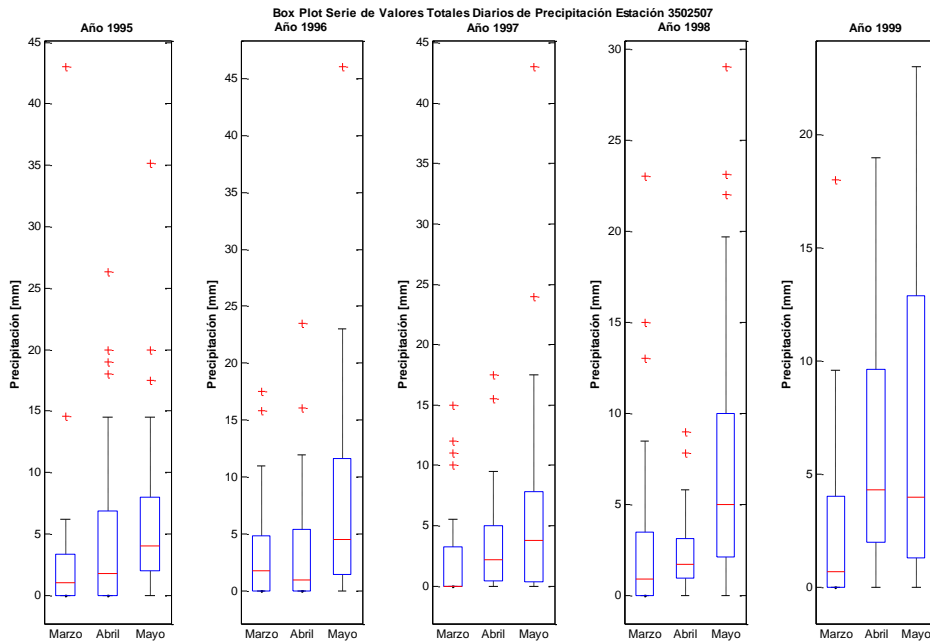
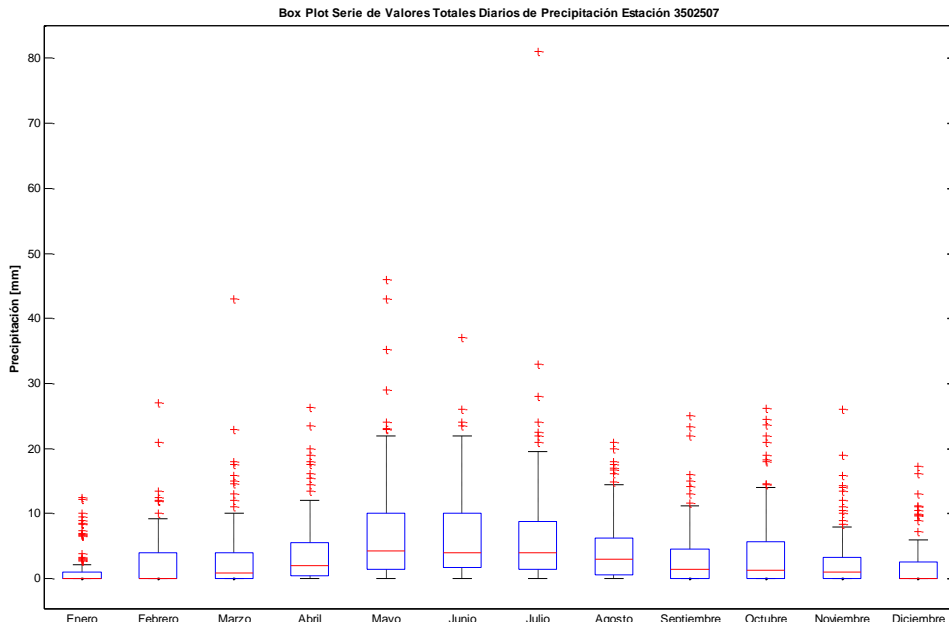


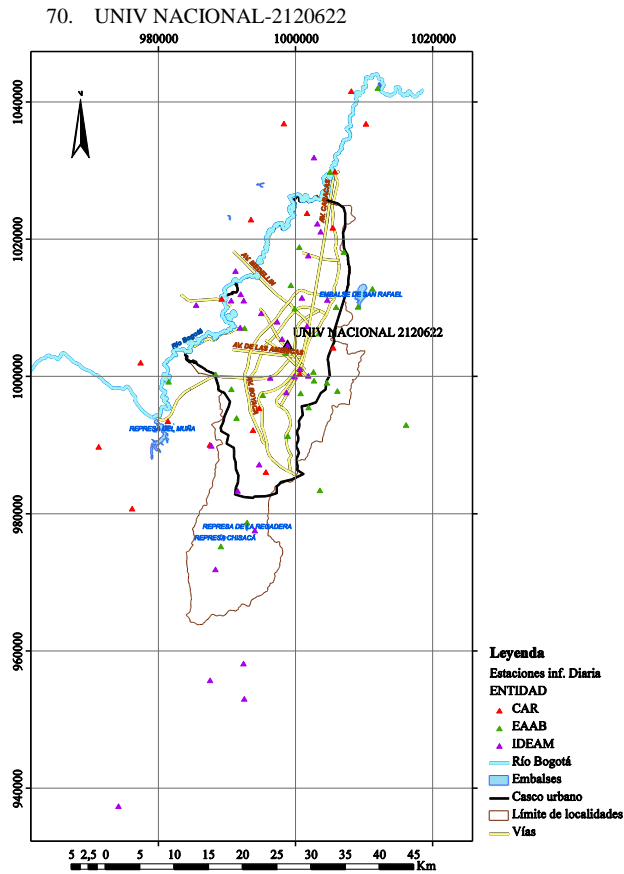




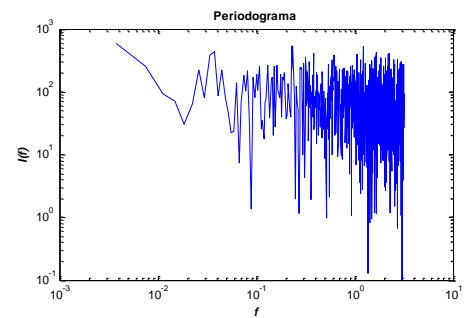
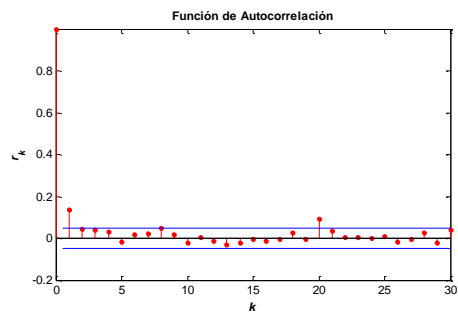
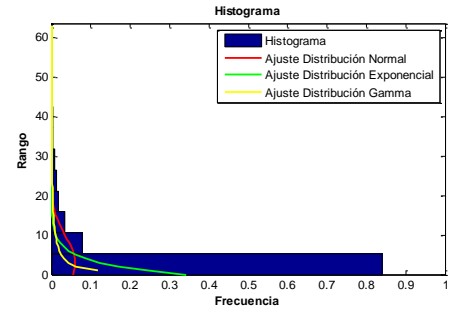
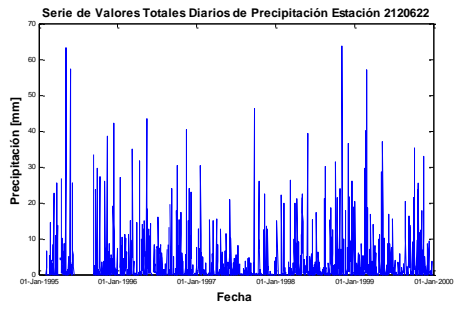
Este	987406
Norte	955820
Cota	3150
Número de Datos	1671
Número de Datos Sin Registro	155
Media	4.0
Mediana	1.6
Máximo	81.0
Desviación Estándar	6.0
Varianza	35.7
Coefficiente de Variación	1.3E-03
Coefficiente de Asimetría	3.3
Coefficiente de Curtosis	25.5

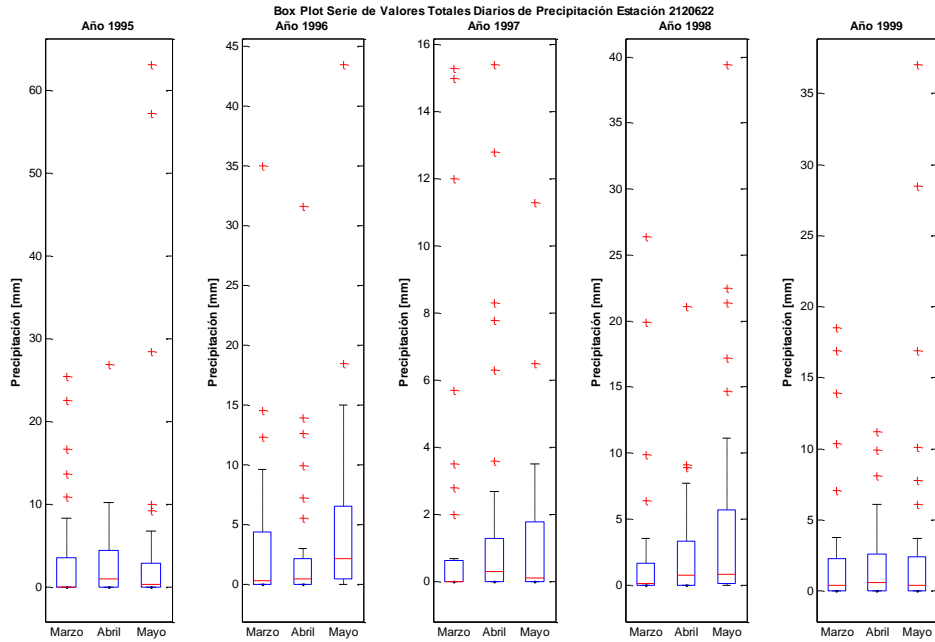
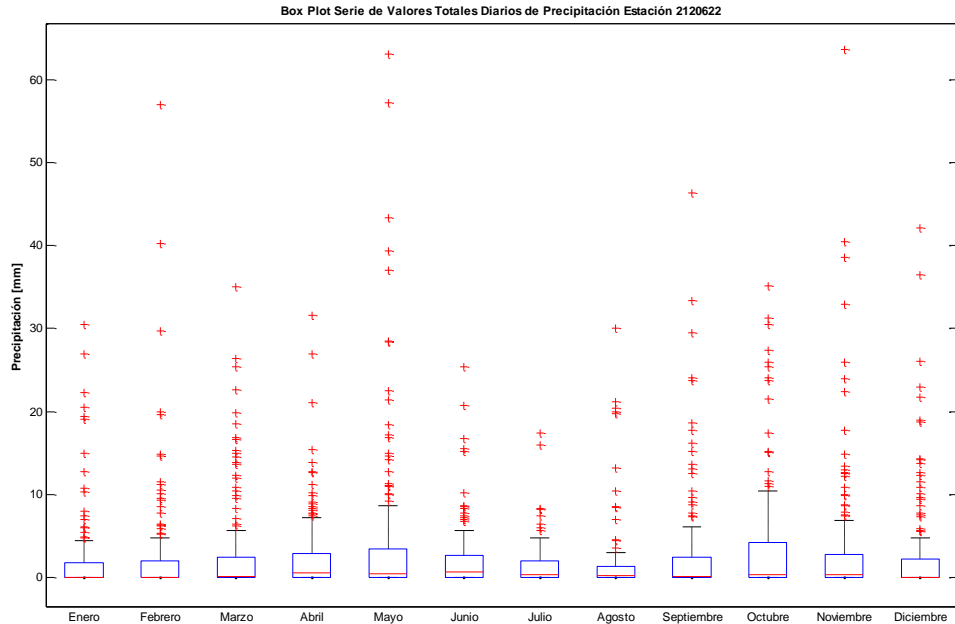




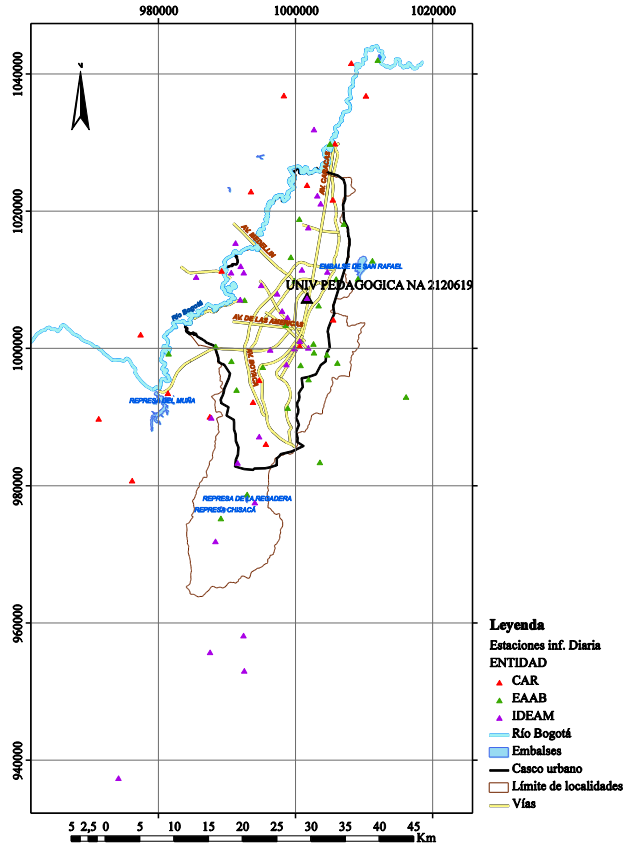


Este	998717
Norte	1004632
Cota	2556
Número de Datos	1687
Número de Datos Sin Registro	139
Media	2.9
Mediana	0.3
Máximo	63.6
Desviación Estándar	6.6
Varianza	43.3
Coefficiente de Variación	1.1E-03
Coefficiente de Asimetría	4.1
Coefficiente de Curtosis	25.1

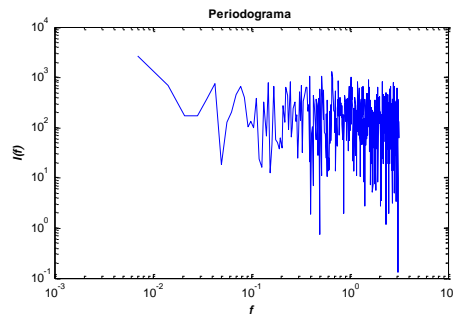
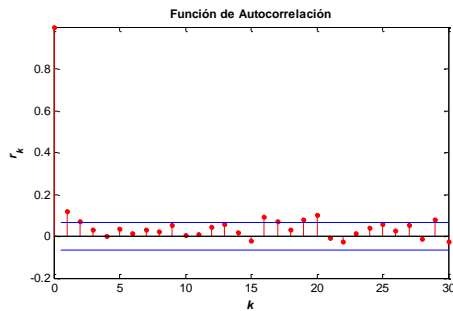
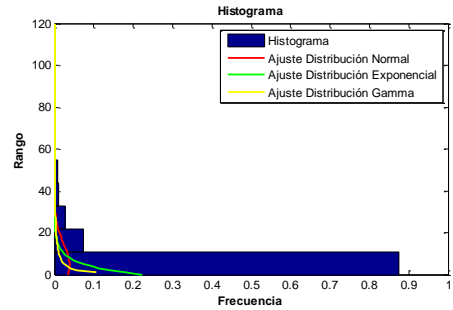
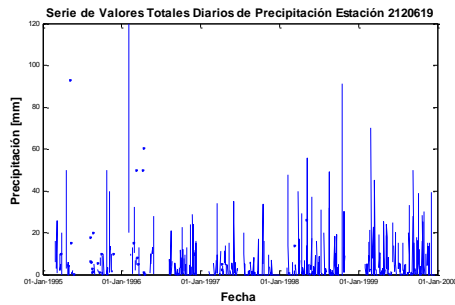


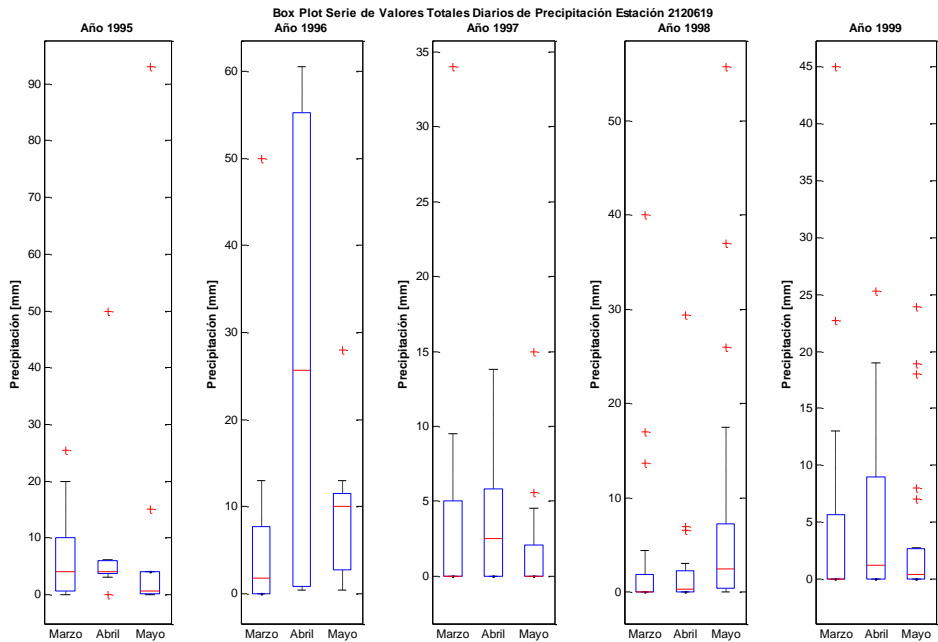
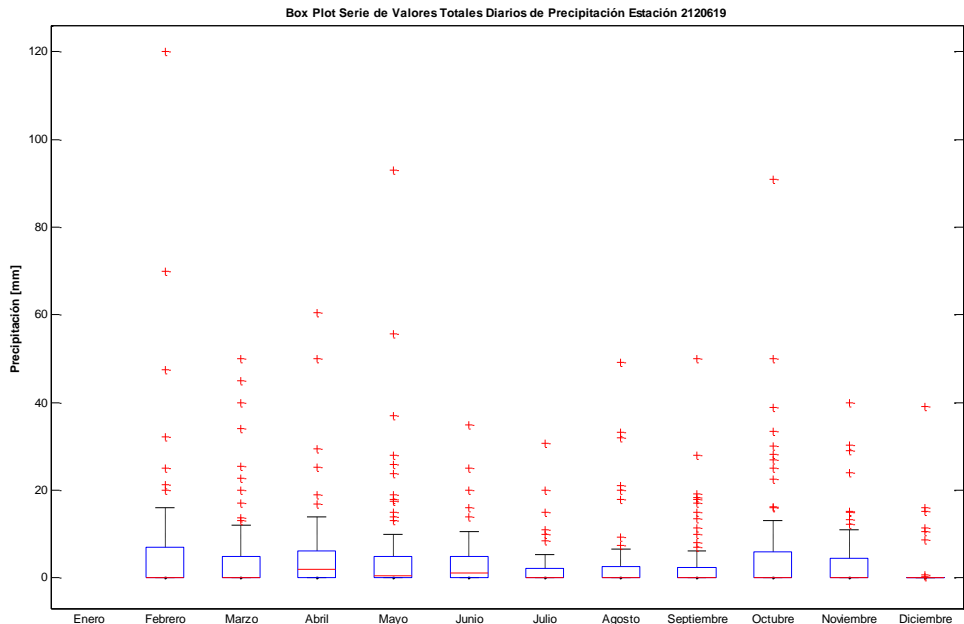


71. UNIV PEDAGOGICA NA-2120619

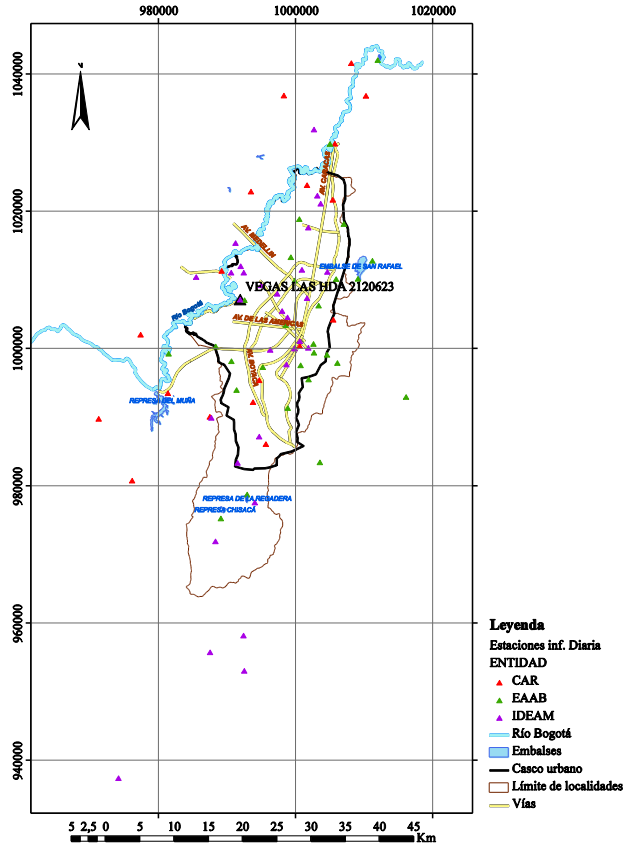


Este	1001581
Norte	1007478
Cota	2570
Número de Datos	889
Número de Datos Sin Registro	937
Media	4.5
Mediana	0.0
Máximo	120.0
Desviación Estándar	10.4
Varianza	108.6
Coefficiente de Variación	1.7E-03
Coefficiente de Asimetría	4.8
Coefficiente de Curtosis	36.4

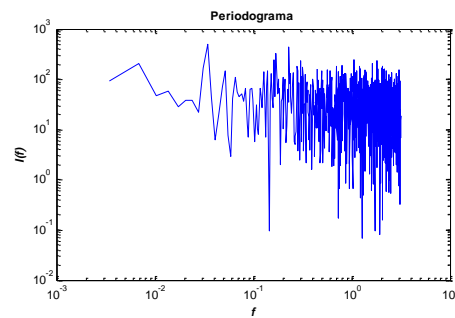
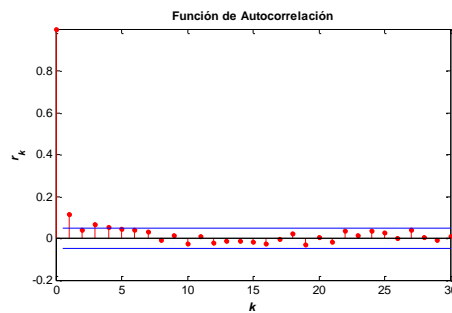
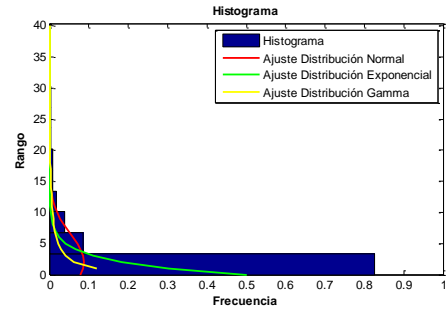
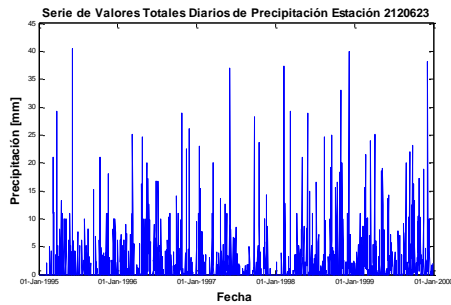


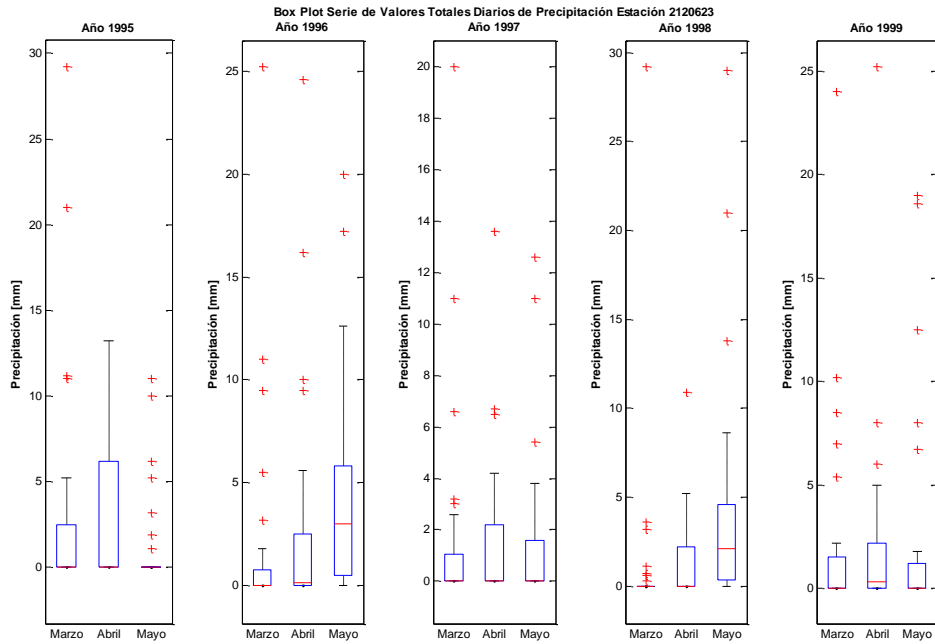
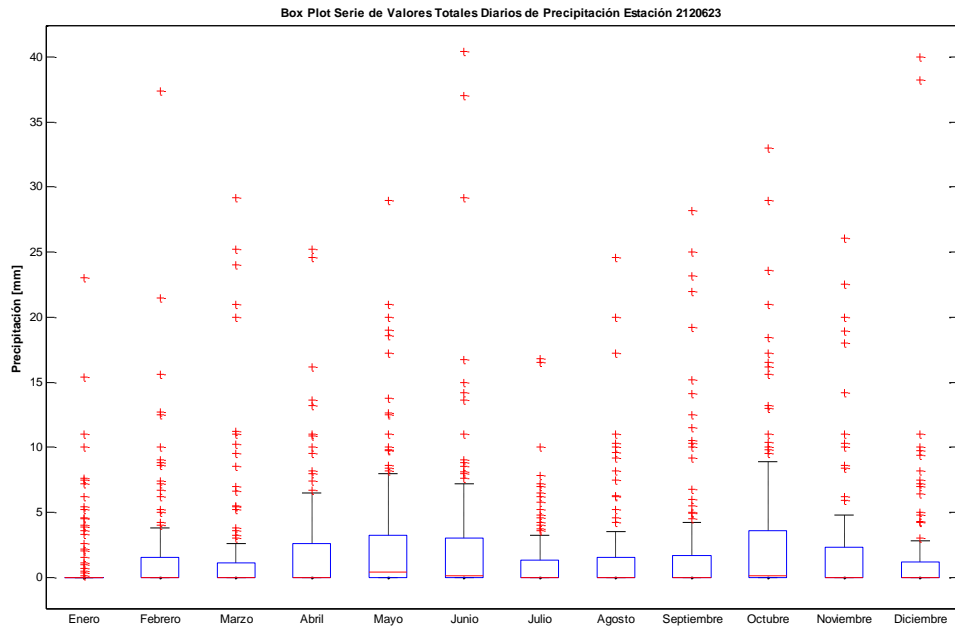


72. VEGAS LAS HDA-2120623

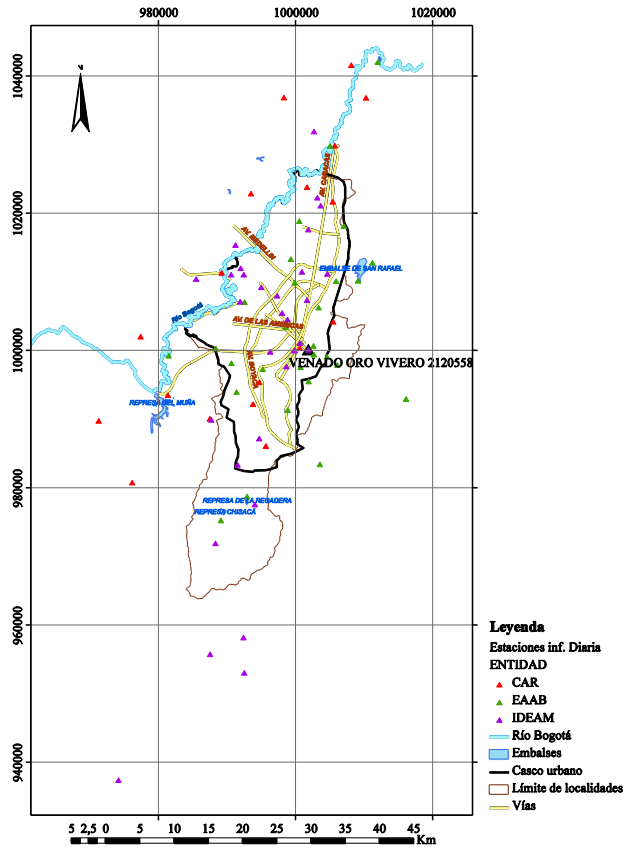


Este	991799
Norte	1007240
Cota	2543
Número de Datos	1826
Número de Datos Sin Registro	0
Media	2.0
Mediana	0.0
Máximo	40.4
Desviación Estándar	4.6
Varianza	21.0
Coefficiente de Variación	7.8E-04
Coefficiente de Asimetría	4.0
Coefficiente de Curtosis	23.3

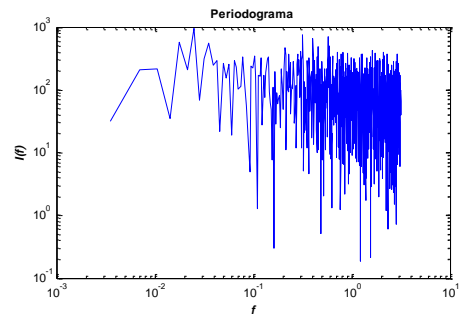
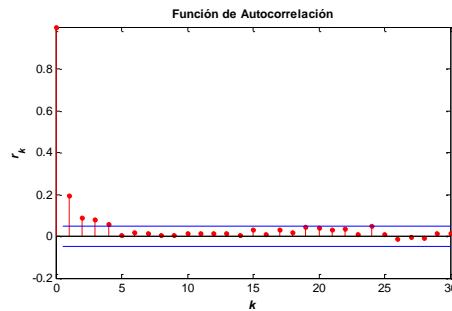
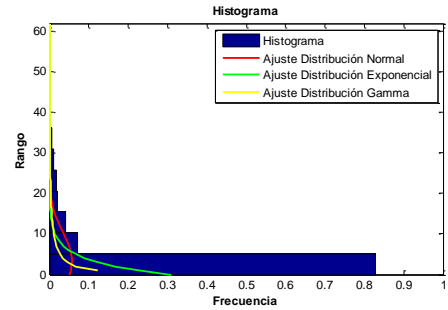
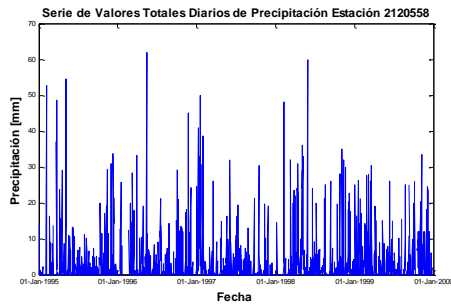


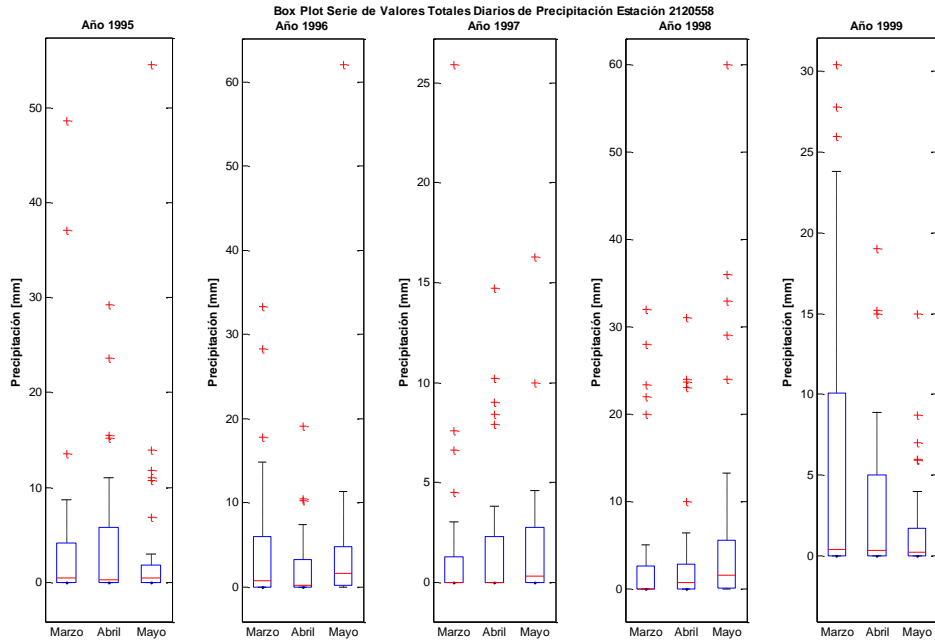
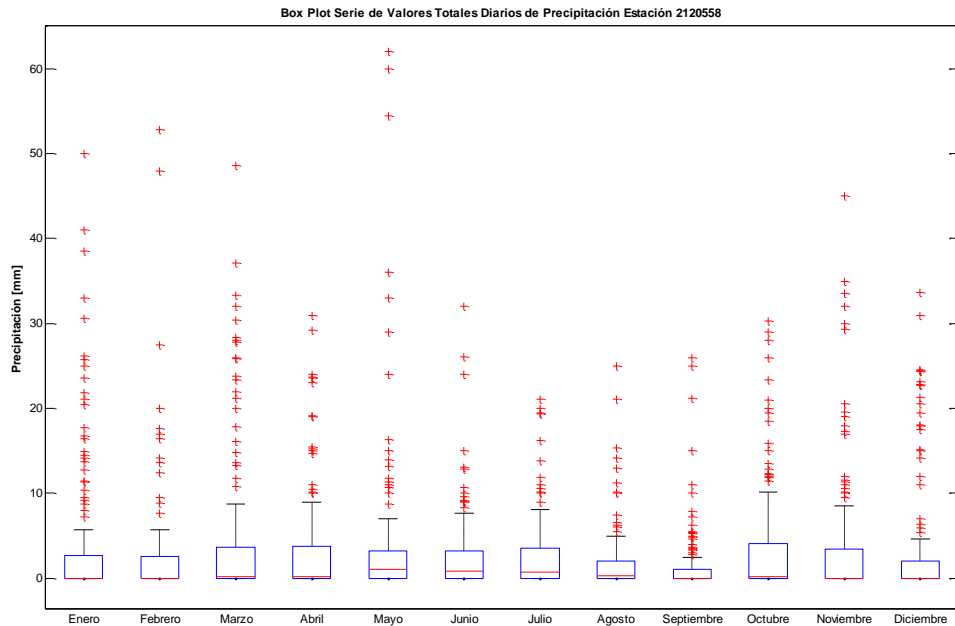


73. VENADO ORO VIVERO-2120558

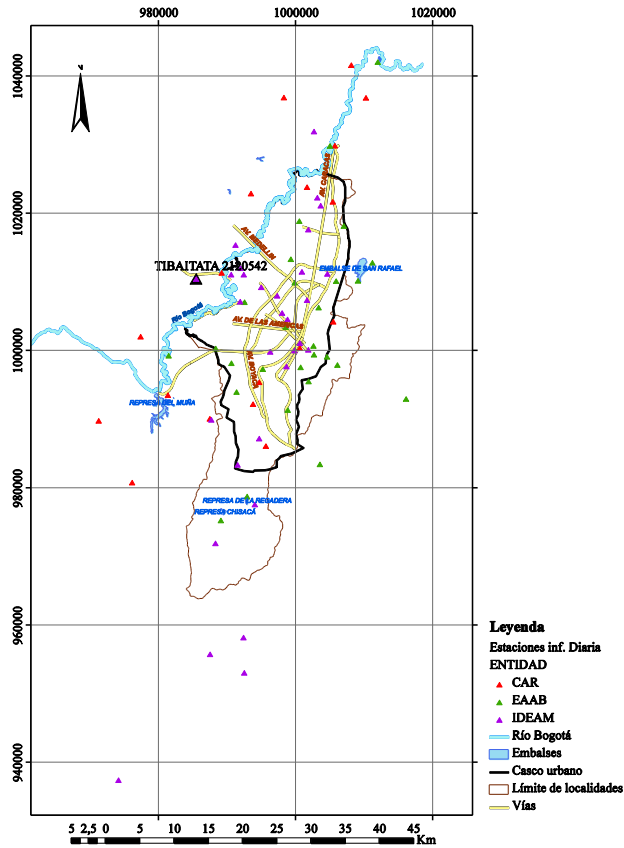


Este	1001770
Norte	1000239
Cota	2725
Número de Datos	1774
Número de Datos Sin Registro	52
Media	3.2
Mediana	0.2
Máximo	62.0
Desviación Estándar	7.0
Varianza	49.1
Coefficiente de Variación	1.2E-03
Coefficiente de Asimetría	3.6
Coefficiente de Curtosis	18.9

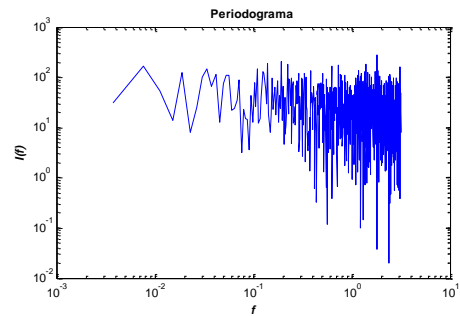
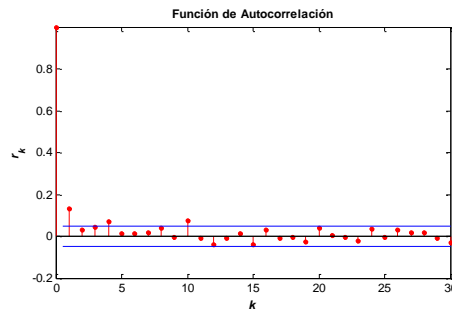
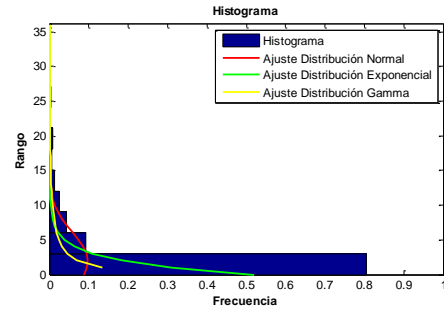
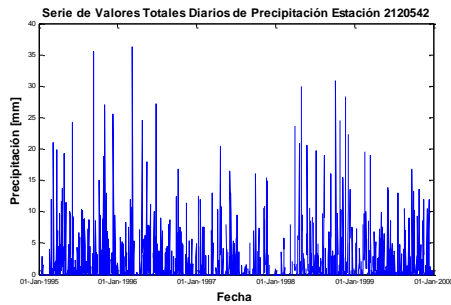


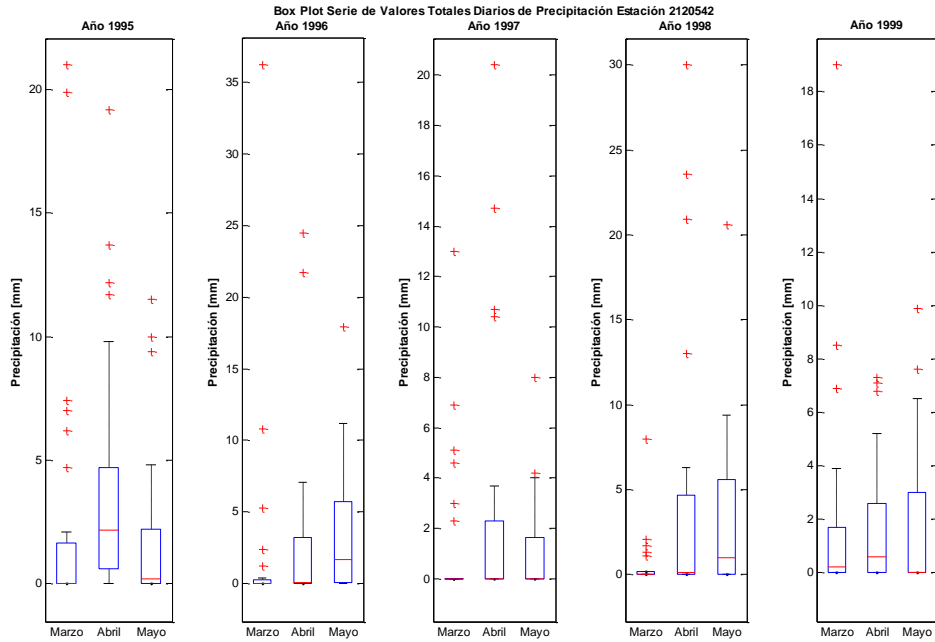
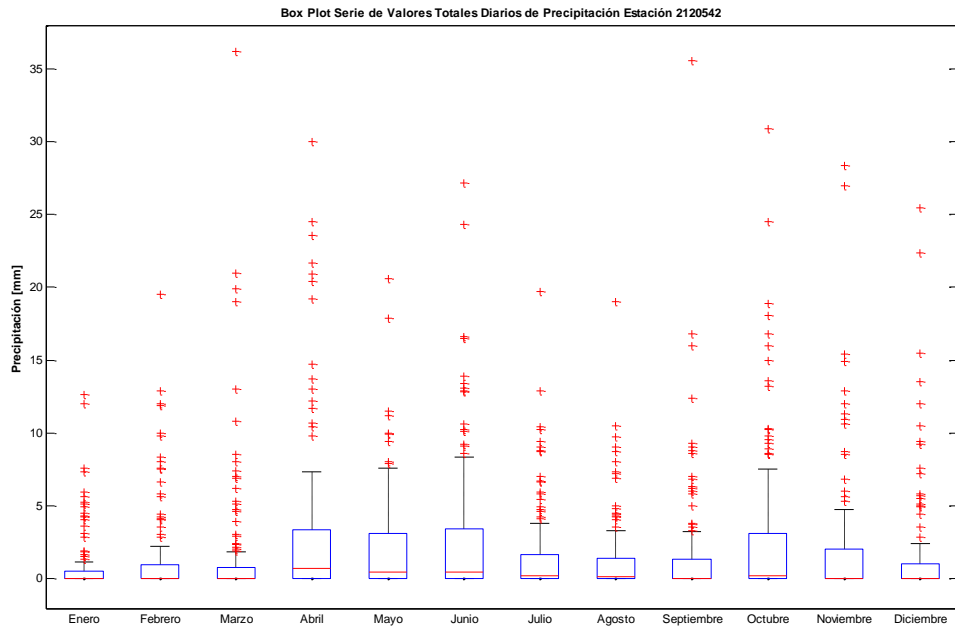


74. TIBAITATA-2120542

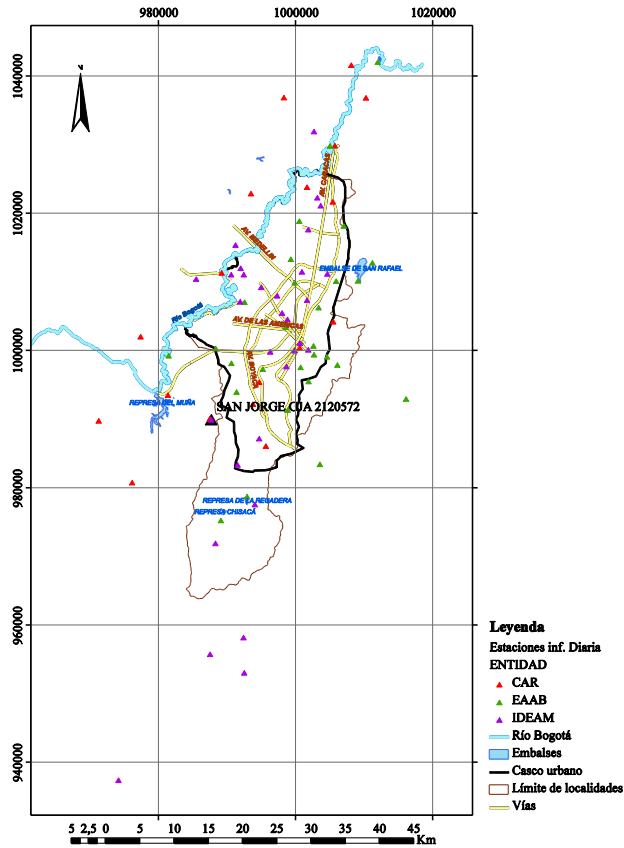


Este	985411
Norte	1010531
Cota	2543
Número de Datos	1660
Número de Datos Sin Registro	166
Media	1.9
Mediana	0.0
Máximo	36.2
Desviación Estándar	4.1
Varianza	16.8
Coefficiente de Variación	7.6E-04
Coefficiente de Asimetría	3.6
Coefficiente de Curtosis	19.7

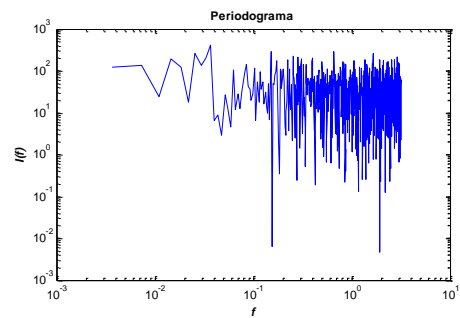
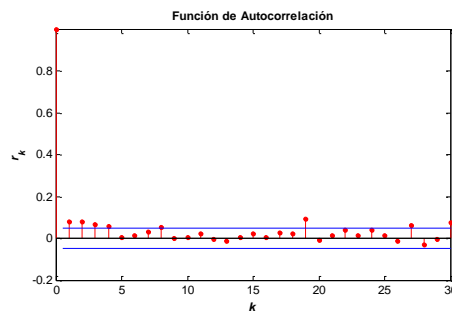
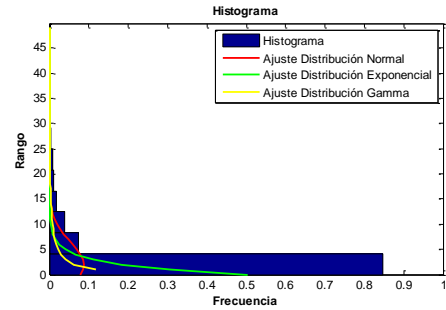
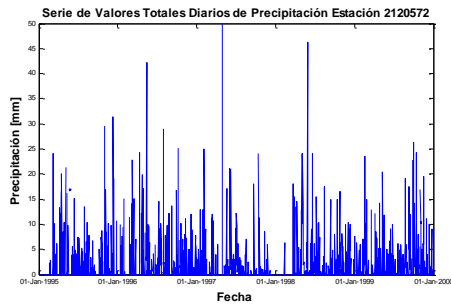


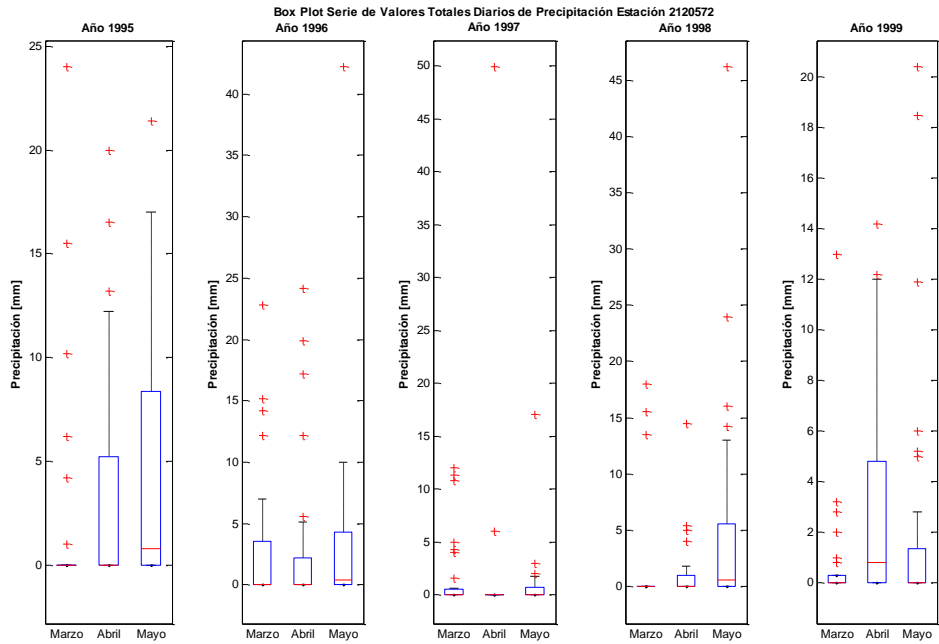
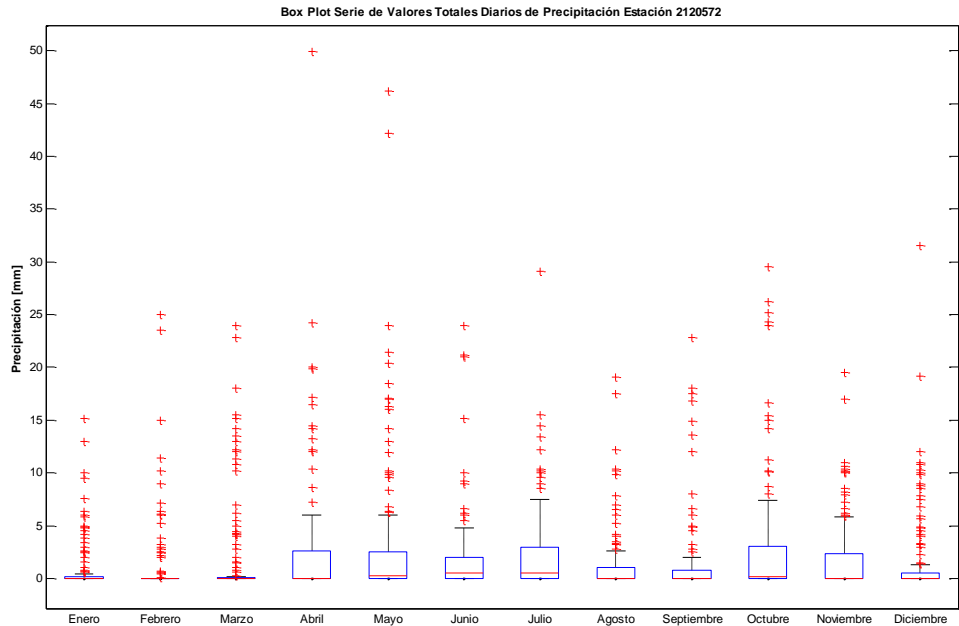


75. SAN JORGE GJA-2120572

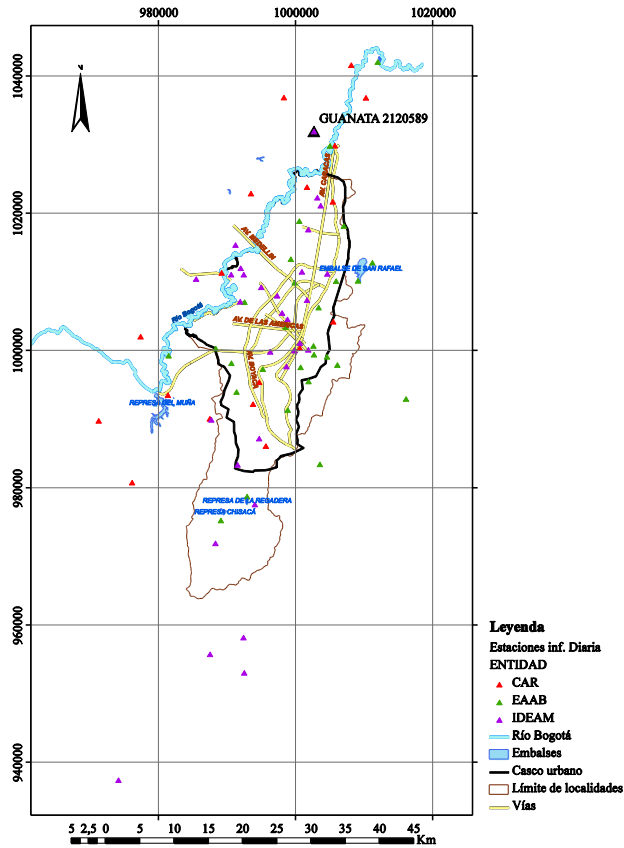


Este	987596
Norte	989999
Cota	2900
Número de Datos	1709
Número de Datos Sin Registro	117
Media	2.0
Mediana	0.0
Máximo	49.9
Desviación Estándar	4.6
Varianza	21.1
Coefficiente de Variación	6.8E-04
Coefficiente de Asimetría	4.0
Coefficiente de Curtosis	25.8

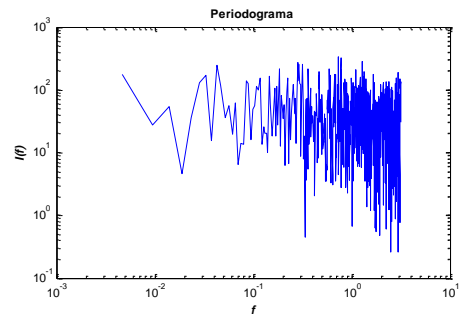
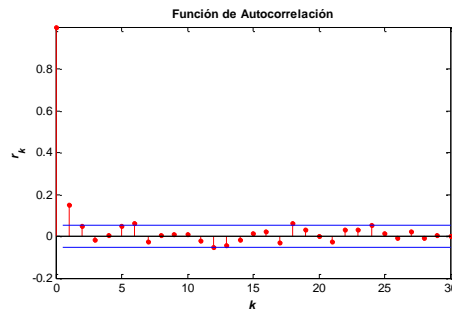
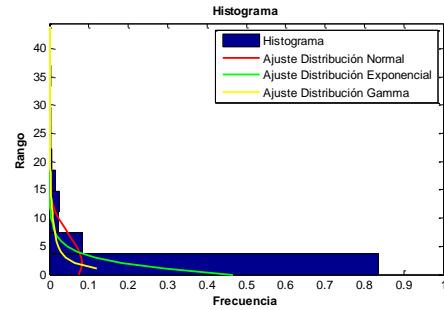
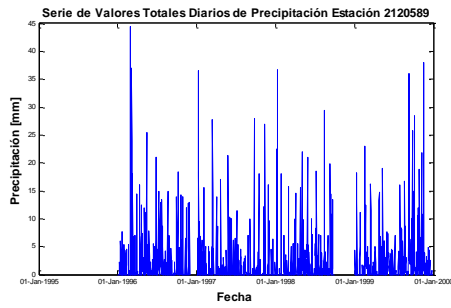


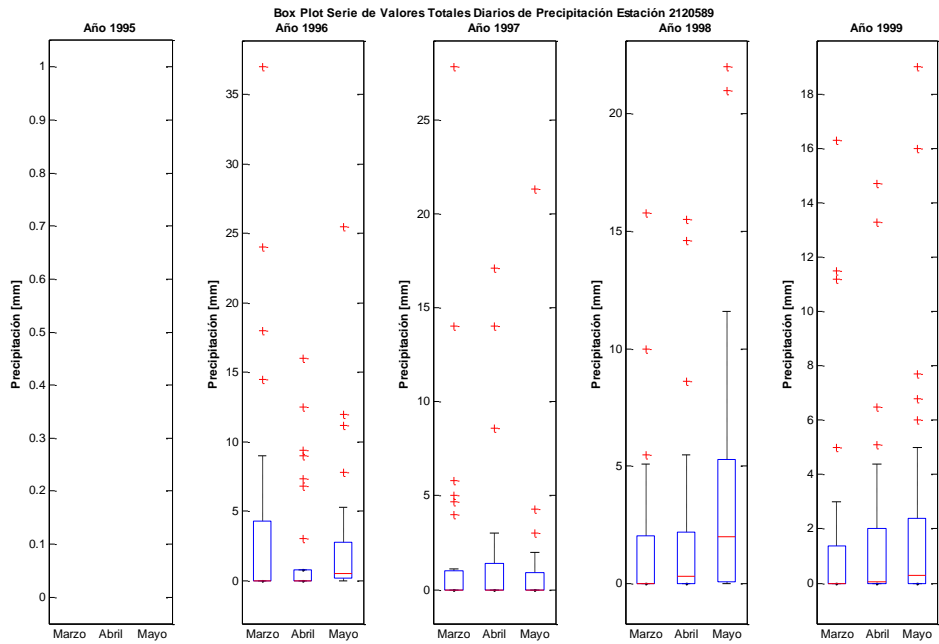
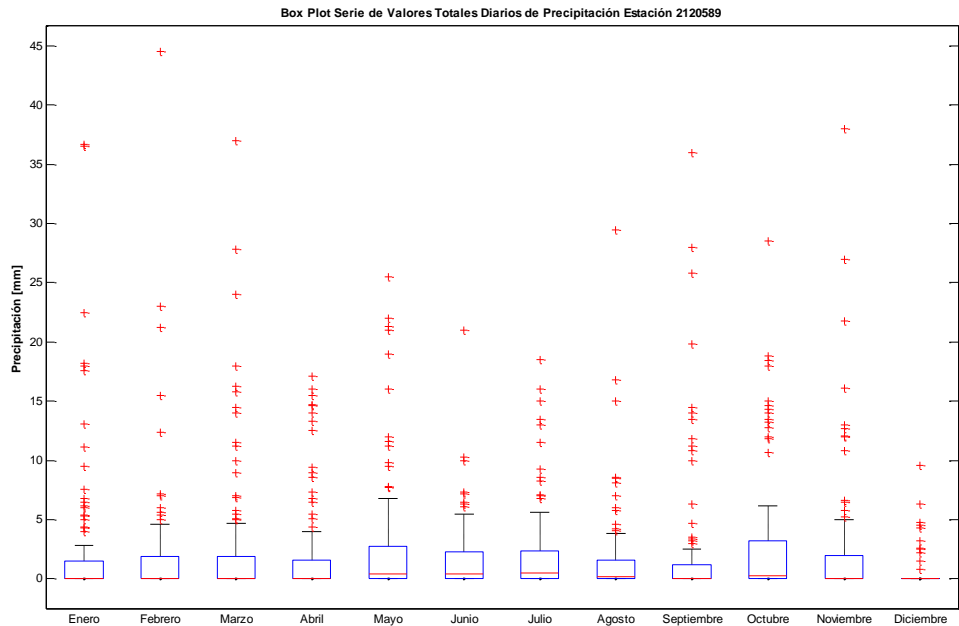


76. GUANATA-2120589

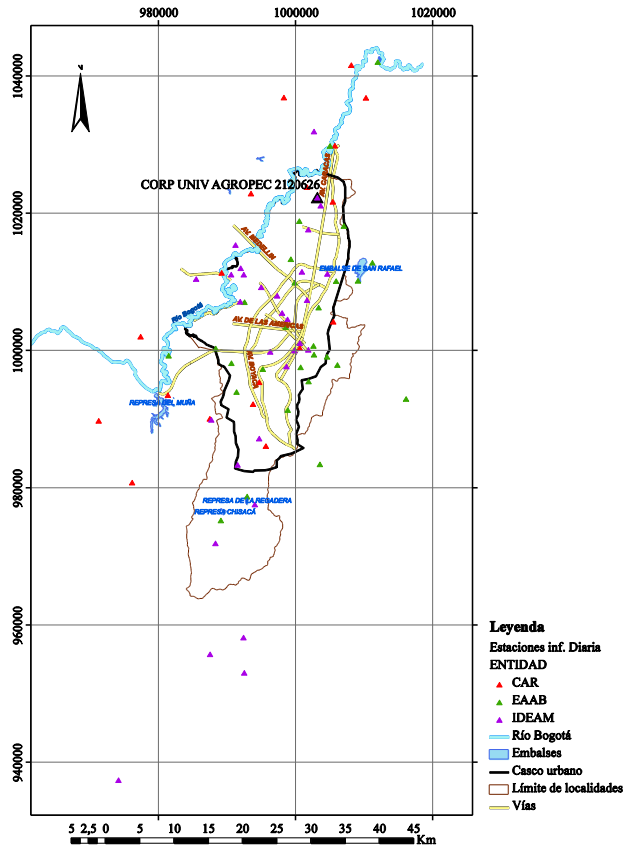


Este	1002570
Norte	1032040
Cota	2550
Número de Datos	1336
Número de Datos Sin Registro	490
Media	2.2
Mediana	0.0
Máximo	44.5
Desviación Estándar	4.9
Varianza	23.7
Coefficiente de Variación	8.4E-04
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	22.1

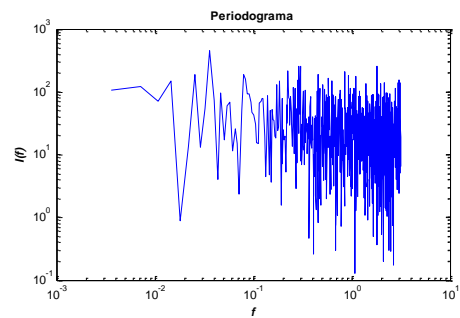
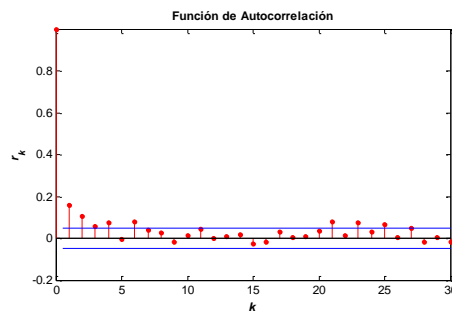
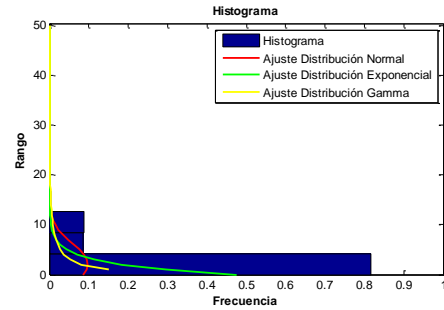
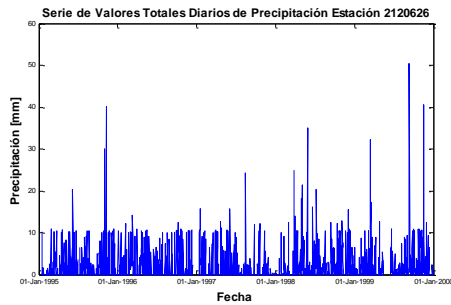


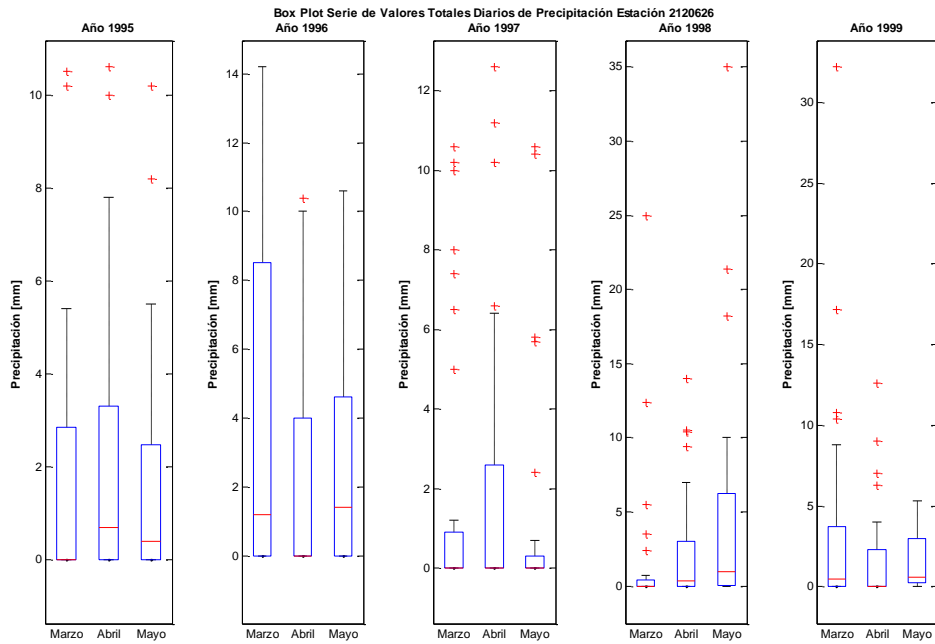
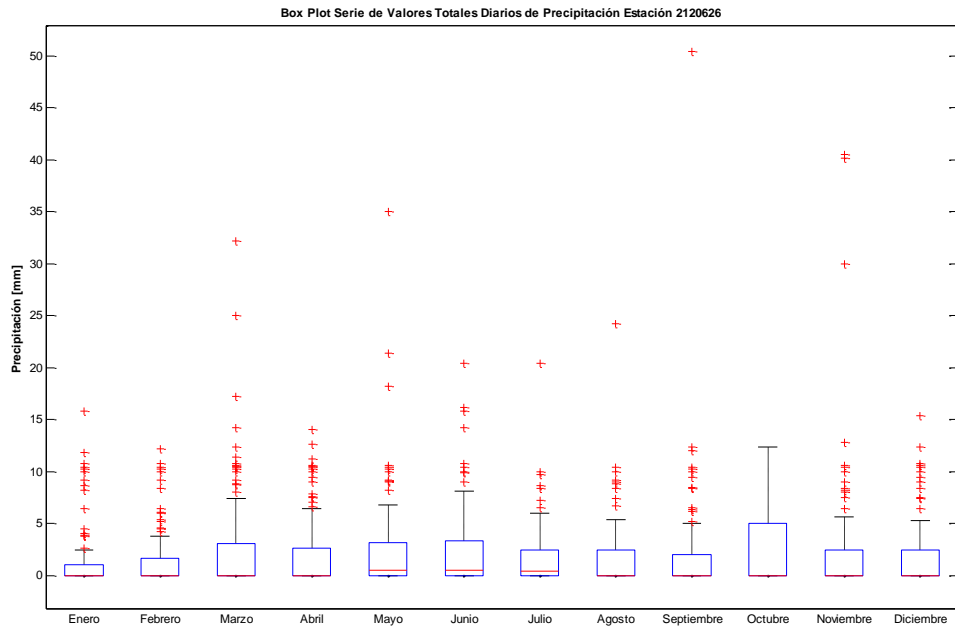


77. CORP UNIV AGROPEC-2120626



Este	1003082
Norte	1022386
Cota	2570
Número de Datos	1737
Número de Datos Sin Registro	89
Media	2.1
Mediana	0.0
Máximo	50.4
Desviación Estándar	4.1
Varianza	16.5
Coefficiente de Variación	8.2E-04
Coefficiente de Asimetría	3.9
Coefficiente de Curtosis	30.0





APÉNDICE B. MÉTODOS DE INTERPOLACIÓN ESPACIAL

1. Método del inverso de la distancia (IDW)

El método inverso de la distancia, que también se llama método de la distancia inversa ponderada (IDW *inverse distance weighted* por sus siglas en inglés), es una técnica general y sencilla para realizar interpolación espacial. Este método ha sido utilizado ampliamente en diferentes campos, tales como hidrología, ciencias de la tierra (Ware *et al.*, 1991; Ashraf *et al.*, 1997; Cheng, 1998), entre otros. La ecuación básica para el método inverso de la distancia está definida como:

$$k_{xy} = \frac{\sum_{i=1}^N k_i w_i}{\sum_{i=1}^N w_i} \quad \text{Ecuación 9-1}$$

donde k_i es el valor de control para el punto de muestreo i , w_i representa un peso para determinar la importancia relativa de cada punto de control k_i en el proceso de la interpolación, k_{xy} es el punto a estimar, y N es el número de puntos de muestreo (Bartier y Keller, 1996). Este método se aplica comúnmente para estimar precipitación media y precipitación desconocida en zonas donde no se conoce el valor de la precipitación. Cuando cada punto de control tiene la misma importancia relativa, el método inverso de la distancia es idéntico al método de la media aritmética también utilizado para estimación de las precipitaciones.

Una alternativa que para conocer el valor del peso w_i , consiste en reconocer la influencia de los puntos más cercanos y distantes, para lo que se estima las distancias desde el punto a estimar y los puntos con valores conocidos. Entonces a la distancia entre los puntos se le asigna una potencia:

$$w_i = d_{xyi}^{-m} \quad \text{Ecuación 9-2}$$

donde d_{xyi} es la distancia entre k_{xy} y k_i , y m es un exponente dado por los usuarios (Bartier y Keller, 1996), y también se llama el orden de las distancias. El método de distancia inversa es flexible debido a la naturaleza ajustable de la orden de las distancias (Chang *et al.*, 2003). Entonces, la Ecuación 9-1 puede ser reescrita como:

$$k_{xy} = \frac{\sum_{i=1}^N k_i d_{xyi}^{-m}}{\sum_{i=1}^N d_{xyi}^{-m}} \quad \text{Ecuación 9-3}$$

Para el cálculo de la precipitación la ecuación se escribe en los siguientes términos:

$$P_p = \sum_{i=1}^N (W_i P_i) = \frac{\sum_{i=1}^N P_i d_{pi}^{-m}}{\sum_{i=1}^N d_{pi}^{-m}} \quad \text{Ecuación 9-4}$$

donde P_p es la precipitación interpolados en el punto p ; P_i es la precipitación conocida en el punto i ; W_i es el factor de ponderación que representa la influencia relativa de la estación conocida del punto i , y d_{pi} es la distancia entre el punto p y la estación con valor conocido i . La interpolación IDW es univariante con un solo factor de influencia, es decir, la distancia horizontal. Esta técnica supone que la zona es la interpolación uniforme en vez de variable (Ware et al., 1991). Por lo tanto, no puede aplicarse en zonas con una topografía escarpada, de ser utilizada la Ecuación 9-4 sufre algunas modificaciones. En el caso de la interpolación de precipitación en la ciudad de Bogotá el valor del exponente m es de dos.

2. Método V4

El método de interpolación V4 de Matlab se basa en un algoritmo que encuentra una superficie de curvatura mínima que pasa por un conjunto de puntos espaciados de forma irregular (Sandwell, 1987). Draftsmen en el siglo 19 solucionó el problema para una dimensión, asignándole pesos a una viga lo suficientemente elástica para deformarla y hacerla pasar por todos los puntos. Análogamente, para N puntos y para m dimensiones se tiene:

$$\nabla^4 \mathbf{w}(\mathbf{x}) = \sum_{j=1}^N \alpha_j \delta(\mathbf{x} - \mathbf{x}_j) \quad \text{Ecuación 9-5}$$

$$\mathbf{w}(\mathbf{x}_i) = \mathbf{w}_i \quad \text{Ecuación 9-6}$$

Donde ∇^4 es el operador biarmónico y \mathbf{x} es la posición en el espacio m -dimensional. La solución general es:

$$\mathbf{w}(\mathbf{x}) = \sum_{j=1}^N \alpha_j \phi_m(\mathbf{x} - \mathbf{x}_j) \quad \text{Ecuación 9-7}$$

Los pesos α_j 's son encontrados solucionando el siguiente sistema lineal

$$\mathbf{w}_i = \sum_{j=1}^N \alpha_j \phi_m(\mathbf{x}_i - \mathbf{x}_j) \quad \text{Ecuación 9-8}$$

Donde ϕ_m representa las funciones Green biarmónicas. Como en el caso unidimensional, las pendientes pueden ser usadas en el espacio m-dimensional cada pendiente está asociada con una dirección n . Cuando las pendientes son utilizadas para solucionar el problema de la interpolación, se debe resolver el siguiente sistema lineal:

$$s_i = (\nabla \mathbf{w} \cdot \mathbf{n})_i = \sum_{j=1}^N \alpha_j \nabla \phi_m(x_j - x_j) \cdot \mathbf{n}_i \quad \text{Ecuación 9-9}$$

En siguiente tabla se muestran las funciones de Green y las ecuaciones para el cálculo del gradiente en función de la dimensión en la que se esté trabajando.

Número de dimensiones m	Función de Green $\phi_m(\mathbf{x})$	Gradiente Función de Green $\nabla \phi_m(\mathbf{x})$
1	$ x ^3$	$x x $
2	$ x ^2(\ln x -1)$	$x(2\ln x -1)$
3	$ x $	$x x ^{-1}$
4	$\ln x $	$x x ^{-2}$
5	$ x ^{-1}$	$-x x ^{-3}$
6	$ x ^{-2}$	$-2x x ^{-4}$
m	$ x ^{4-m}$	$-(4-m)x x ^{2-m}$

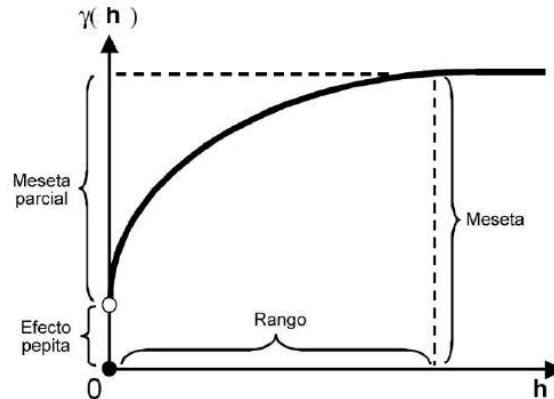
3. Método de Kriging

La geoestadística brinda una herramienta para la caracterización de esta correlación espacial, la cual se denomina semivariograma, (Carr, 1995). El semivariograma es una función que mide el grado de disimilaridad de una variable en función de la distancia de separación (Clark, 1979). Para determinar esta disimilaridad se usan las diferencias de los valores elevadas al cuadrado, con lo cual el semivariograma se define como:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^n [Z(u) - Z(u+h)]^2 \quad \text{Ecuación 9-10}$$

donde $Z(u)$ corresponde al valor del atributo de interés medido en el punto u , $Z(u+h)$ es el valor del atributo medido en el punto $u+h$ y $N(h)$ es el número de pares encontrados a una distancia de separación h . Por lo tanto, el semivariograma es una función que depende de la separación espacial h . En general, el semivariograma para distancias de separación pequeñas tiende a presentar valores pequeños, y a medida que la distancia de separación aumenta los valores del semivariograma son mayores hasta que para cierta distancia de separación este deja de crecer y se estabiliza. La distancia de separación para la cual ocurre esto se denomina rango de influencia y marca la distancia de separación máxima para la cual el fenómeno presenta correlación espacial. El valor para el cual se estabiliza el semivariograma se denomina meseta y en general este valor tiende a ser igual a

la varianza original de los datos. De acuerdo con la definición de semivariograma Ecuación 9-10, para una distancia de separación igual a 0 se tiene que $\gamma(0)$ debería ser igual a 0, sin embargo esto no ocurre en algunas oportunidades; situación denominada efecto pepita. Para una colección de datos para diversos valores de h y graficando $\gamma(h)$ vs. h de la Ecuación 9-10, se obtiene la representación del semivariograma como se observa en la siguiente figura.



Unos de los problemas con el análisis de información espacial consiste en la reconstrucción de la variación espacial del fenómeno a partir de medidas puntuales, es decir, la elaboración de mapas a partir de muestras. A nivel geoestadístico, la metodología de Kriging permite cumplir este objetivo, asegurando que las estimaciones en cada punto tienen la mínima varianza de error posible (Isaaks y Srivistava, 1989). El valor en un punto donde no se tenga muestra u se puede determinar a partir de los valores de las muestras cercanas Z_i , por medio de un estimador:

$$Z^* = \sum_{i=1}^n \lambda_i Z_i \quad \text{Ecuación 9-11}$$

Donde λ_i son los factores de ponderación que miden la influencia de cada muestra en el valor estimado. Estos pesos dependen de la correlación espacial del fenómeno (semivariograma), así como de la configuración geométrica de las muestras cercanas. Estos pesos se determinan por medio de la solución de un sistema de ecuaciones dado por:

$$C(u_\alpha - u_\beta)\lambda - \mu = C(u - u_\alpha) \quad \text{Ecuación 9-12}$$

En donde $C(u - u_\alpha)$ es la covarianza entre el punto de estimación y las muestras cercanas, $C(u_\alpha - u_\beta)$ es la covarianza entre las muestras cercanas y μ es el multiplicador de Lagrange. Estos valores de covarianza se determinan a partir del semivariograma (Ecuación 9-10) por medio de la siguiente relación:

$$C(h) = VAR - \gamma(h) \quad \text{Ecuación 9-13}$$

en donde $C(h)$ es la covarianza para la distancia de separación h , VAR es la varianza de los datos y $\gamma(h)$ es el semivariograma para la distancia de separación h .

Una de las grandes ventajas que tiene el método de kriging (Ecuación 9-12) comparada con otros métodos de interpolación espacial, es la posibilidad de definir una medida de confiabilidad del valor estimado (Ecuación 9-11) denominada, varianza de estimación (Isaaks y Srivistava, 1989; Carr, 1995) y definida por:

$$\sigma_e^2 = VAR - \sum_{i=1}^n \lambda_i C(u - u_\alpha) \quad \text{Ecuación 9-14}$$

Otro componente importante en la aplicación del método de Kriging es el análisis estructural. Este consiste en identificar y caracterizar la porción de la región de estudio en que se cumple la hipótesis geoestadística asumida. Según Armstrong y Carignan (como se cita en Giraldo 2002), el análisis estructural o variográfico está compuesto por:

- El cálculo del semivariograma experimental: mediante el cual se identifican las zonas homogéneas o anómalas.
- Y el ajuste del semivariograma experimental a un modelo teórico conocido: con el fin de facilitar el manejo matemático de las expresiones.

Los modelos matemáticos más comunes son:

- Modelo Pepita Puro: Característico de fenómenos netamente aleatorios y con un comportamiento Pepita Puro, es decir: valor pepita igual a la meseta, sin área de influencia definida

$$\gamma(h) = \begin{cases} 0 & \text{Si } h = 0 \\ C, \text{ constante} & \text{Si } h > 0 \end{cases} \quad \text{Ecuación 9-15}$$

- Modelo esférico: Representativo de un comportamiento lineal en el origen, sin valor pepita y con meseta.

$$\gamma(h) = \begin{cases} C * \left[\frac{3}{2} * \left(\frac{h}{a} \right) - \frac{1}{2} * \left(\frac{h}{a} \right)^3 \right] & \text{Si } h \leq a \\ C, \text{ constante} & \text{Si } h > a \end{cases} \quad \text{Ecuación 9-16}$$

donde $3/2 \cdot (C/a)$ es la pendiente en el origen, a es el rango y C el alcance vertical.

- Modelo exponencial: Distintivo de una función lineal en el origen, sin valor pepita, con meseta finita y crecimiento menos rápido que el del modelo esférico.

$$\gamma(h) = C * \left[1 - e^{-\frac{3 \cdot h}{a'}} \right] \quad \text{Ecuación 9-17}$$

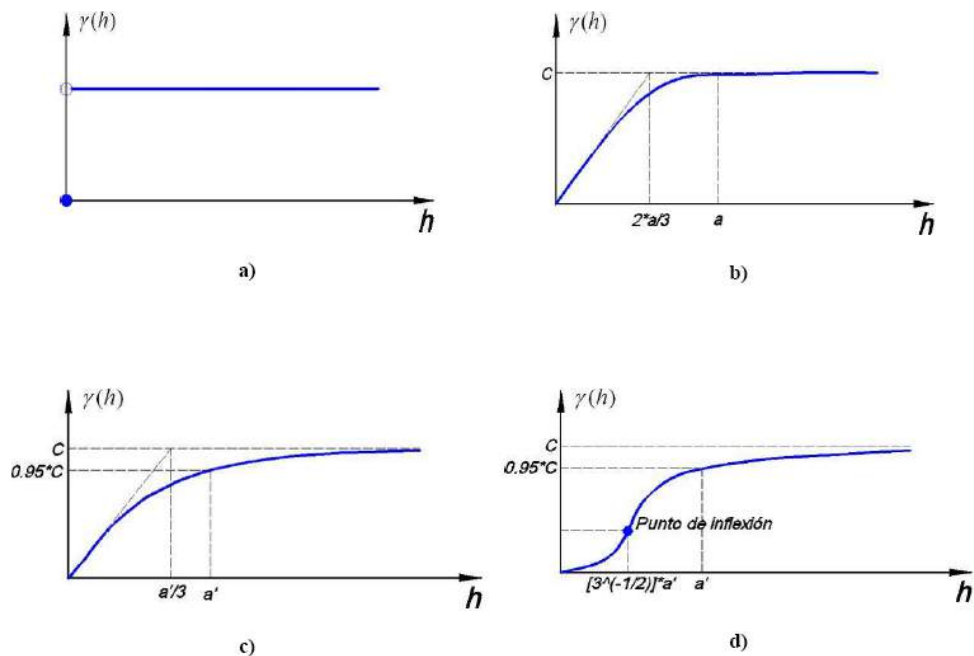
donde $3 \cdot c / a'$ es la pendiente en el origen, a' es el rango y C el alcance vertical.

- Modelo Gaussiano: Caracterizado por un comportamiento parabólico en el origen, sin valor pepita, con meseta finita y punto de inflexión.

$$\gamma(h) = C * \left[1 - e^{-3 \cdot \left(\frac{h}{a'} \right)^2} \right] \quad \text{Ecuación 9-18}$$

donde la pendiente en el origen es igual a cero, a' es el rango y C el alcance vertical.

En la **¡Error! No se encuentra el origen de la referencia.** se muestran las formas de los semivariogramas explicados.



a) Modelo Pepita puro, b) Modelo Esférico, c) Modelo Exponencial, d) Modelo Gaussiano

***APÉNDICE C. CARACTERIZACIÓN DE LOS TIPOS DE
CLASIFICACIÓN***
