

CubOps: Manual de Usuario

Julio Sergio Santana

19 de enero de 2018

Índice

1. Introducción	1
1.1. Bases de datos de ejemplo	2
2. Aspecto general de CubOps	2
3. Los datos y su recolección	2
3.1. Los archivos y sus nombres	4
3.2. Los archivos y su estructura	4
3.3. Empaquetado de los archivos	4
3.4. Información geográfica de estaciones	5
3.5. CubOps: colección de los datos	5
3.6. Descarga y recarga de la estructura de información	6
4. Selección de variables	6
4.1. Selección de variables y operaciones	8
5. Operaciones estadísticas sobre columnas completas	8
5.1. Operaciones con más de un resultado: cuantiles	10
6. Estratificaciones de datos y operaciones sobre ellas	12
6.1. Definición de una estratificación en CubOps	12
6.2. Estratificaciones y operaciones con resultados múltiples	13
6.3. Aplicación recurrente de operaciones	13
7. Gráficas sencillas con CubOps	15
7.1. Gráficos de información completa de estaciones (columnas)	15
7.2. Algunos usos de los gráficos	17

1. Introducción

CubOps es una aplicación Web que le permitirá operar sobre un conjunto de puntos, típicamente estaciones meteorológicas, cada uno de los cuales contiene información temporal de un número finito de variables, como pueden ser, precipitación, temperaturas mínimas y máximas.

En general, el sistema trabaja con puntos a los que se le ha asociado un número de variables cualquiera, que dependen de una fecha secuencial en días.

Para el análisis y la aplicación de operaciones o producción de algunos gráficos, el sistema permite generar una tabla completa para una variable seleccionada, cada una de cuyas columnas corresponde a cada uno de los puntos o estaciones de la base de datos. Posteriormente, se pueden aplicar operaciones estadísticas (media, rango, mediana, *quantile*, etc.), a los datos de la columna, o a una selección de esta de acuerdo con las fechas establecidas. El sistema también permite la producción de varios tipos de gráficos simples para las columnas de las tablas resultantes, a saber: histogramas, *boxplots*, o series de tiempo.

En cualquier momento, el usuario puede descargar las tablas resultantes de sus operaciones en un formato estándar (csv), que podrá ser leído fácilmente por hojas de cálculo, tales como *LibreOffice Calc*, o *Excel*, o por otros sistemas o programas para desarrollar procesos adicionales con la información.

1.1. Bases de datos de ejemplo

Para que el usuario pueda experimentar con el sistema se proveen las siguientes bases de datos para descargarse en un equipo local:

1. Datos de más de 60 estaciones en la cuenca del río Usumacinta.
2. Datos de 14 estaciones en la cuenca del río Conchos.
3. Información geográfica de ubicación las estaciones registradas en CLICOM.

Estos datos pueden utilizarse a partir de las acciones que se describen a partir de la sección 3.5 (CubOps: colección de los datos).

2. Aspecto general de CubOps

Para ejecutar el sistema basta con introducir la siguiente URL en alguno de los *browsers* comunes de la Web, tales como Chrome o Firefox:

<http://172.16.19.27:3838/CubOps>

El sistema desplegará una interfaz al usuario, en la cual podrá introducir sus peticiones y recibir sus respuestas. La Fig. 1 Muestra la apariencia general del sistema una vez que se han desarrollado varias operaciones. Al inicio de la sesión, sólo se muestra el menú etiquetado como **Acción**, en el área **A** de la Fig. 1, y las áreas **B** y **C** se muestran vacías, aunque, al pulsar en la pestaña etiquetada como **Manual**, en el área **B**, se desplegará el presente documento.

En esa figura se destacan la tres áreas principales del sistema, a saber:

1. **A. Panel de acciones.** Aquí el usuario define diversas acciones que van desde la captura de sus datos, hasta la operación sobre algún subconjunto de ellos para definir alguna estadística.
2. **B. Panel de resultados tabulares.** Los resultados tabulares de las operaciones se muestran en este espacio.
3. **C. Panel de operaciones y resultados gráficos.** Aquí el usuario puede definir diversos tipos de gráfico para alguna de las columnas del resultado tabular.

3. Los datos y su recolección

Generalmente, los datos de cada estación estarán dados como una tabla en la que se proporciona una fecha en cada renglón y con una columna para cada una de las variables que se han registrado para la estación a lo largo del tiempo. En seguida se muestran los diez primeros renglones de una estación con tres variables: Prec (precipitación), Tmax (temperatura máxima) y Tmin (temperatura mínima).

##	Year	Month	Day	Prec	Tmax	Tmin
## 1	1951	1	1	0.0	31.3	17.6
## 2	1951	1	2	0.0	32.1	18.7
## 3	1951	1	3	0.0	32.1	18.3
## 4	1951	1	4	3.5	30.7	20.3
## 5	1951	1	5	0.0	30.7	19.7
## 6	1951	1	6	0.0	30.9	18.4
## 7	1951	1	7	0.0	30.8	17.9

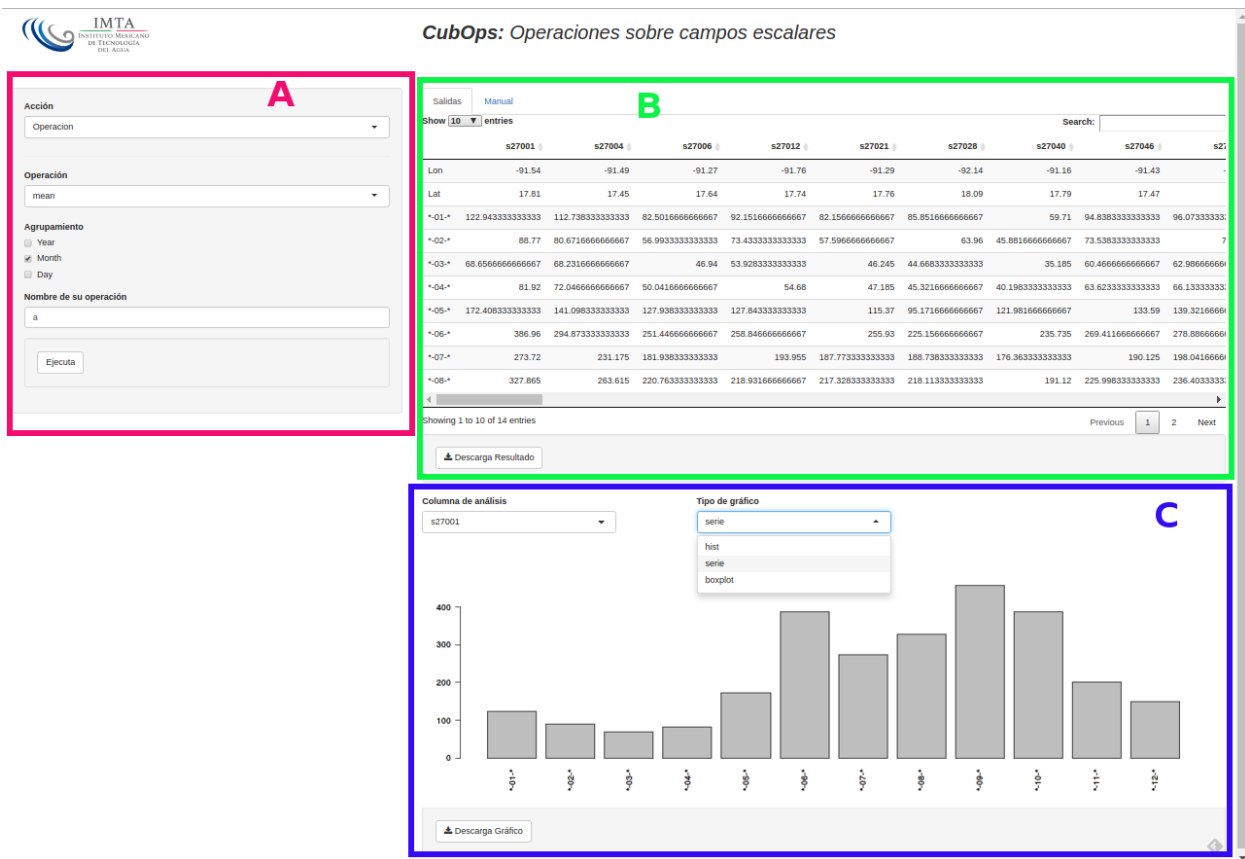


Figura 1: Apariencia general de CubOps

```
## 8 1951      1   8  0.0 21.8 18.3
## 9 1951      1   9 20.0 23.6 14.3
## 10 1951     1  10  0.0 27.6 14.4
```

De esta forma los datos de cada estación se pueden mantener en un archivo de hoja de cálculo, como LibreOffice Calc o Excel. Para su manejo en el presente sistema, dichos archivos deberán exportarse al formato estándar CSV (*comma separated value*), aunque ellos podrán estar dispersos en una estructura de directorios o carpetas.

Los archivos CSV pudieran no traer el encabezado que identifica a cada una de las columnas, pero en ese caso el sistema entenderá que, aparte de la fecha que se debe dar en las primeras tres columnas, las columnas restantes serán sólo tres que corresponderán a las variables pp (precipitación), tmax (temperatura máxima) y tmin (temperatura mínima), en ese orden.

3.1. Los archivos y sus nombres

Para poder manejar dentro del sistema un conjunto de archivos de tablas como las descritas en la sección 3, éstas se pueden encontrar dispersas dentro de una estructura de directorios/subdirectorios (carpetas), del sistema operativo, tal como se muestra en la Fig. 2. Sin embargo, para que el sistema pueda reconocer los archivos en cuestión, ellos deben ser nombrados de acuerdo con una estructura sintáctica particular que se describe a continuación:

1. El primer caracter del nombre debe ser una letra “s”, minúscula.
2. En seguida, una secuencia de caracteres alfanuméricos que representan el *identificador* de la estación correspondiente al archivo.
3. Después una letra “H”, mayúscula y la secuencia de caracteres “.csv”, que identifican el tipo de archivo.

Por ejemplo, el archivo de nombre “s3394H.csv” contendría la tabla de información de la estación 3394 en formato CSV. A estos archivos se les denominará *archivos de estaciones*.

3.2. Los archivos y su estructura

Externamente, como se muestra en la Fig. 2, los archivos de estaciones podrán estar dispersos en una estructura de directorios. Internamente, la condición es que la estructura de las tablas sea la misma para todos los archivos, esto es, todos deben tener el mismo número de columnas, con las tres primeras para la fecha en orden: año, mes y día, y las restantes para las variables que se registren en las estaciones, en el mismo orden para todas las tablas. Si las tablas llevan o no encabezado, esto debe ser igual para todas ellas.

3.3. Empaquetado de los archivos

Para que el sistema pueda digerir la información correspondiente a un conjunto de estaciones como la que se ha ilustrado en las secciones anteriores y que se muestra gráficamente en la Fig. 2, es necesario empaquetarlo en un solo archivo de tipo ZIP.

En un sistema operativo Linux o Unix, el comando para efectuar esa operación es:

```
zip -r arch.zip path/de/inicio
```

Al final de la operación toda la estructura de archivos se tendría en el paquete de nombre “arch.zip”, en el caso del ejemplo.

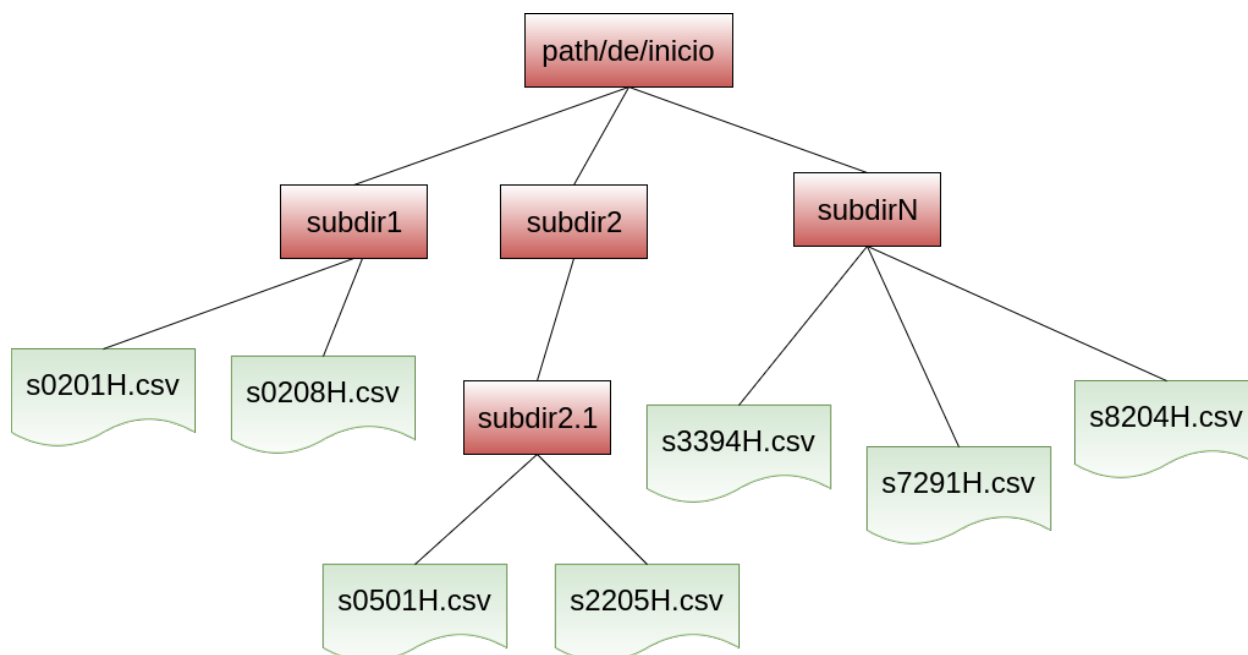


Figura 2: Ejemplo de estructura de directorios con archivos CSV

3.4. Información geográfica de estaciones

Aparte de la información de diversas variables de las estaciones, que se empaqueta en un archivo de tipo ZIP, como se ha mostrado en las secciones anteriores, es necesario proveer también de una tabla que registre las coordenadas geográficas en formato CSV. Un ejemplo de la estructura de esta tabla se muestra a continuación:

##	Clave	Nombre	Latitud	Longitud
## 1	1003	CALVILLO (SMN)	21.88333	-102.7189
## 2	1004	CA?ADA HONDA	22.00778	-102.1981
## 3	1005	PRESA EL NIAGARA	21.77889	-102.4414
## 4	1006	EL TULE (SMN)	22.08167	-102.0914
## 5	1007	JESUS MARIA (SMN)	21.96222	-102.3447
## 6	1008	PUERTO DE LA CONCEPCION	22.20250	-102.1344
## 7	1009	LA LABOR (SMN)	21.96222	-102.6947
## 8	1010	LA TINAJA	22.16528	-102.5558
## 9	1011	MALPASO	21.85833	-102.6628
## 10	1012	PRESA MEDIA LUNA	21.79250	-102.8017

La tabla debe contener por lo menos tres columnas, la correspondiente a la “Clave” de la estación, y las correspondientes a la “Longitud” y “Latitud” de la estación, con esos nombres de columna. Además, la tabla debe contener *al menos* la información de las estaciones que se consignan en el empaquetamiento de la sección 3.3.

3.5. CubOps: colección de los datos

Para que el sistema colectione toda la información de las estaciones se debe seleccionar, en el menú de Acciones, la entrada “**Colecta Datos**”, como se muestra en la Fig. 3.

Una vez seleccionada esta acción el sistema provee sendos botones para la *subida* de los dos archivos de interés, descritos en las secciones 3.3 y 3.4; esto es, el paquete de archivos ZIP con la información de las

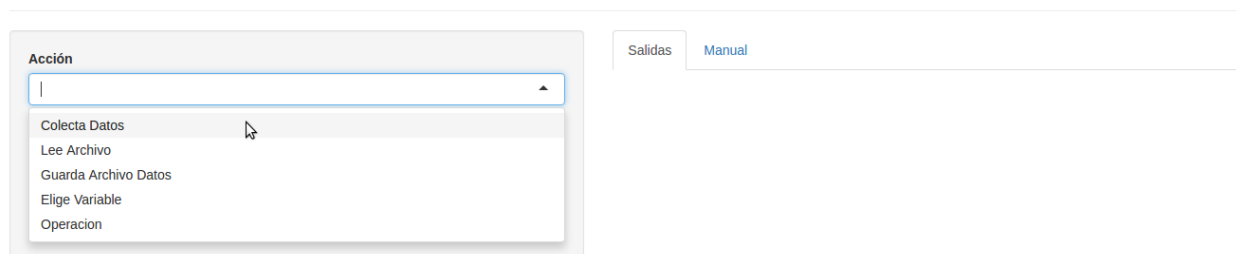


Figura 3: Colección de datos

variables en las estaciones y el archivo CSV con la información geográfica de las estaciones. El despliegue de estos botones se muestra en la Fig. 4, a continuación.

Ya que se han subido o cargado estos datos, lo cual se muestra en las barras de progreso de la carga en la Fig. 5, se debe oprimir el botón “**Ejecuta**”, en la parte inferior del panel de Acciones, para proceder al armado completo de la estructura de información que utiliza internamente el sistema. En la Fig. 5, se muestra el resultado, donde, en el panel de resultados tabulares, se muestra un texto que da alguna pista del armado de dicha estructura.

3.6. Descarga y recarga de la estructura de información

Una vez construída la estructura de información, como se ha mostrado en las secciones anteriores, ésta se puede descargar de manera local para reutilizarse posteriormente por el sistema, sin necesidad de reconstruirla a partir de un archivo de tipo ZIP. La descarga se hace seleccionando la entrada “**Guardar Archivo Datos**”, en el menú de acciones (ver Fig. 3). La estructura de información queda guardada en un archivo de tipo RDS, que es un archivo binario del lenguaje de programación R.


La recarga de la estructura de información, se hace seleccionando la entrada “**Lee Archivo**”, en el menú de Acciones (ver Fig. 3). Una vez leído algún archivo de este tipo (RDS), el resultado sería semejante al de la construcción de la estructura que se muestra en la Fig. 5.

4. Selección de variables

El sistema maneja una estructura de información que se ha construído como se ha señalado en las secciones anteriores. La Fig. 6 muestra de manera gráfica la apariencia de dicha estructura.

La estructura puede ser concebida como *un número determinado de tablas*, tantas como variables haya registradas, cuyos renglones serían las fechas y cuyas columnas corresponderían con cada una de las estaciones (ver Fig. 6). El presente sistema permite operar sobre cada una de las tablas de la estructura; esto es, con cada una de las variables registradas. De modo que el primer paso, antes de cualquier operación, es la selección de una tabla o variable para trabajar con ella.

Para proceder a la selección de la variable, se debe entrar al menú de Acciones y seleccionar de ahí la entrada “**Elige Variable**” (ver Fig. 3). Esta acción abre un menú, “**Variable a revisar**”, con el conjunto de variables



IMTA
INSTITUTO MEXICANO
DE TECNOLOGÍA
DEL AGUA

CubOps: Operaciones sobre campos escalares

Acción

Colecta Datos

Archivo ZIP

Browse...

No file selected

Info Estaciones (Arch. csv)

Browse...

No file selected

Ejecuta

Salidas

Manual

Figura 4: Botones para la *subida* de archivos de datos



IMTA
INSTITUTO MEXICANO
DE TECNOLOGÍA
DEL AGUA

CubOps: Operaciones sobre campos escalares

Acción

Colecta Datos

Archivo ZIP

Browse...

arch.zip

Upload complete

Info Estaciones (Arch. csv)

Browse...

Estaciones001.csv

Upload complete

Ejecuta

Salidas

Manual

CubOps: Operación exitosa

```

List of 2
$ Coords:'data.frame':      66 obs. of  2 variables:
..$ Lon: num [1:66] -91.5 -91.5 -91.3 -91.8 -91.3 ...
..$ Lat: num [1:66] 17.8 17.4 17.6 17.7 17.8 ...
$ Cubote: num [1:21915, 1:66, 1:3] 0 0 0 3.5 0 0 0 20 0 ...
.. attr(,"dimnames")=List of 3
.. ..$ : chr [1:21915] "1951-01-01" "1951-01-02" "1951-01-03" "1951-01-04" ...
.. ..$ : chr [1:66] "s27001" "s27004" "s27006" "s27012" ...
.. ..$ : chr [1:3] "Prec" "Tmax" "Tmin"

```

Figura 5: Armado de la estructura de información

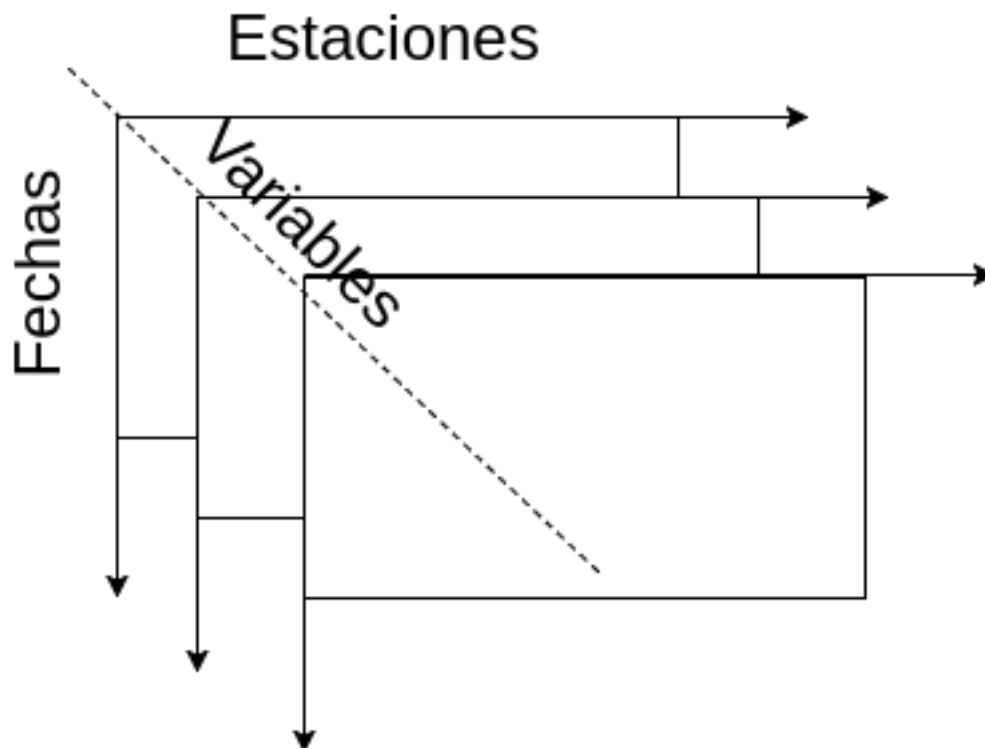


Figura 6: La estructura de la información

que se pueden revisar, como se muestra en la Fig. 7.

Una vez que se elige la variable deseada, y luego de que se oprime el botón **“Ejecuta”**, en el panel de resultados tabulares, aparece la tabla correspondiente a la variable, con los dos primeros renglones mostrando las coordenadas geográficas de las estaciones. La tabla se puede navegar por páginas. Este resultado se muestra abajo en la Fig. 8.

La tabla resultante, se puede descargar a su equipo local, en el formato estándar CSV, oprimiendo el botón **“Descarga Resultado”**, que aparece en la parte inferior del panel de resultados tabulares (Ver Fig. 8).

4.1. Selección de variables y operaciones

En la sección siguiente se empieza a tratar el asunto de la aplicación de operaciones sobre los datos contenidos en la tabla o variable seleccionada. Dado que **las operaciones se aplican sobre la tabla visible**, cada vez que se quiera aplicar una nueva operación sobre los mismos datos, será necesario *regresar* y elegir nuevamente la misma variable para restablecer la tabla de origen. Esto, como se verá más adelante en la sección 6.3, es porque hay resultados que se obtienen de aplicar consecutivamente varias operaciones sobre las tablas consecutivas resultantes, por ejemplo, en el caso de los promedios de sumas acumulativas de precipitación.

5. Operaciones estadísticas sobre columnas completas

El sistema permite ejecutar alguna de las operaciones estadísticas disponibles, sobre la serie completa de datos de todas las columnas exhibidas, que corresponden a las estaciones registradas. Como se verá más adelante, algunas operaciones, como la media, la varianza, o la mediana, reportarán un sólo resultado numérico para cada serie de datos, mientras que otras, como el rango o los cuantiles, reportarán dos o más resultados.

Acción

Elige Variable

Variable a revisar

Tmin

Prec

Tmax

Tmin

Salidas Manual

CubOps: Operación exitosa

```

List of 2
$ Coords: 'data.frame':      66 obs. of  2 variables:
..$ Lon: num [1:66] -91.5 -91.5 -91.3 -91.8 -91.3 ...
..$ Lat: num [1:66] 17.8 17.4 17.6 17.7 17.8 ...
$ Cubote: num [1:21915, 1:66, 1:3] 0 0 0 3.5 0 0 0 20 0 ...
..- attr(*, "dimnames")=List of 3
.. ..$ : chr [1:21915] "1951-01-01" "1951-01-02" "1951-01-03" "1951-01-04" ...
.. ..$ : chr [1:66] "s27001" "s27004" "s27006" "s27012" ...
.. ..$ : chr [1:3] "Prec" "Tmax" "Tmin"

```

Figura 7: Conjunto de variables que pueden ser revisadas

Acción

Elige Variable

Variable a revisar

Tmin

Ejecuta

Salidas Manual

Show 10 entries

Search:

	s27001	s27004	s27006	s27012	s27021	s27028	s27040	s27046	s27047	s2705
Lon	-91.54	-91.49	-91.27	-91.76	-91.29	-92.14	-91.16	-91.43	-91.43	-9
Lat	17.81	17.45	17.64	17.74	17.76	18.09	17.79	17.47	17.47	18
1951-01-01	17.6	18		16.6		17.3	16.1	15.9	15.9	1
1951-01-02	18.7	19.3		17.6		20	17.1	16.9	17	1
1951-01-03	18.3	20.3		17.2		20.4	16.8	16.5	16.6	1
1951-01-04	20.3	22		19		22.8	18.7	18.4	18.7	
1951-01-05	19.7	21.1		18.5		22	18.1	17.8	18	1
1951-01-06	18.4	17.8		17.4		19.4	16.7	16.4	16.5	2
1951-01-07	17.9	17		17		19	16.2	15.9	15.9	2
1951-01-08	18.3	18.4		17.3		19.1	16.7	16.4	16.5	1

Showing 1 to 10 of 21,917 entries

Previous 1 2 3 4 5 ... 2192 Next

Descarga Resultado

Columna de análisis: s27001

Tipo de gráfico:

Figura 8: Tabla correspondiente a una variable

Acción

Operacion

Operación

mean|

mean

median

var

sd

sum

min

max

range

Ejecuta

Salidas **Manual**

Show 10 entries

Search:

	s27001	s27004	s27006	s27012	s27021	s27028	s27040	s27046	s27047	s27050	s27056	s27059	s27063	s27066
Lon	-91.54	-91.49	-91.27	-91.76	-91.29	-92.14	-91.16	-91.43	-91.43	-92.6	-91.54	-91.18	-91.07	-91.1
Lat	17.81	17.45	17.64	17.74	17.76	18.09	17.79	17.47	17.47	18.38	17.81	17.94	17.6	17.7
1951-01-01	17.6	18		16.6		17.3	16.1	15.9	15.9	17.6	16.4	15.5	17.3	17
1951-01-02	18.7	19.3		17.6		20	17.1	16.9	17	17.7	17.4	16.6	18	
1951-01-03	18.3	20.3		17.2		20.4	16.8	16.5	16.6	18.3	17.1	16.4	17.9	17
1951-01-04	20.3	22		19		22.8	18.7	18.4	18.7	19	19	18.4	19.2	16
1951-01-05	19.7	21.1		18.5		22	18.1	17.8	18	19.8	18.4	17.8	18.8	16
1951-01-06	18.4	17.8		17.4		19.4	16.7	16.4	16.5	20.9	17.2	16.3	17.7	
1951-01-07	17.9	17		17		19	16.2	15.9	15.9	20.3	16.7	15.8	17.3	17
1951-01-08	18.3	18.4		17.3		19.1	16.7	16.4	16.5	19.6	17.1	16.2	17.7	17

Showing 1 to 10 of 21,917 entries

Previous 1 2 3 4 5 ... 2192 Next

Descarga Resultado

Columna de análisis: s27001

Tipo de gráfico:

Figura 9: Menú de operaciones

Para seleccionar una operación cualquiera, primeramente se toma del menú de Acciones, la entrada “**Operación**”, lo que abre un menú de “**Operación**” con todas las operaciones estadísticas disponibles, tal como se muestra en la Fig. 9. Al momento, las operaciones disponibles son: la media (**mean**), la mediana (**median**), la varianza (**var**), la desviación estándar (**sd**), la sumatoria (**sum**), el valor mínimo (**min**), el valor máximo (**max**), el rango (**range**), compuesto por los valores mínimo y máximo, y los cuantiles (**quantile**).

Nótese, en la Fig. 9, que en la tabla sobre la cuál se va a operar, hay columnas sin datos, tales como las etiquetadas “s27006” y “s27021”.

Ya que se ha elegido la operación, por ejemplo *la mediana*, y se le ha dado un nombre, en el campo textual provisto para ello, por ejemplo “mm”, se procede a ejecutar dicha operación, para todas las columnas con datos al oprimir el botón “**Ejecuta**” que aparece en la parte inferior del panel de acciones. Los resultados de la operación se muestran en una tabla como la que se muestra en la Fig. 10, donde la operación en cuestión arroja un único resultado para cada serie de datos (columna).

Nuevamente, la tabla resultante puede ser descargada en un equipo local, en el formato estándar CSV, al oprimir el botón “**Descarga Resultado**”.

5.1. Operaciones con más de un resultado: cuantiles

Como se mencionó antes, hay operaciones estadísticas que arrojarán más de un dato a la salida. Tal es el caso de los cuantiles. Si se elige esta operación en el menú de “**Operación**”, la interfaz del sistema despliega un campo de texto etiquetado como “**Probabilidades separadas por**,”, en el que se pueden especificar las probabilidades de *ruptura*, separadas por comas, de los cuantiles para el cálculo. Por omisión, estas probabilidades son: 0, 0.25, 0.50, 0.75 y 1.0, lo que corresponde a los así denominados *cuartiles* de cada serie. El resultado, que se muestra en la Fig. 11, estaría dado, en ese caso, por 5 renglones, nombrados como: qq.0 %, qq.25 %, qq.50 %, qq.75 % y qq.100 %. El prefijo “qq” en todos los nombres de los renglones de la tabla, se ha tomado del campo de texto “**Nombre de su operación**”, que es un nombre arbitrario que provee el usuario del sistema.

La interpretación de la información en esa tabla es más o menos sencilla. Por ejmplo, para la primera columna, etiquetada como “s27001”, la serie de datos de temperatura mínima, que fue la variable elegida, tienen un valor mínimo y máximo de 10° y 34°, que en la columna se muestran como los renglones qq.0 % y qq.100 %, respectivamente, lo que se interpreta como 0 % de los datos están tienen un valor menor que 10° y 100 % de

Acción

Operacion

Operación

median

Agrupamiento

☒ Year

☒ Month

☒ Day

Nombre de su operación

mm

Ejecuta

Salidas **Manual**

Show 10 entries

Search:

	s27001	s27004	s27006	s27012	s27021	s27028	s27040	s27046	s27047	s27050
Lon	-91.54	-91.49	-91.27	-91.76	-91.29	-92.14	-91.16	-91.43	-91.43	-92.
Lat	17.81	17.45	17.64	17.74	17.76	18.09	17.79	17.47	17.47	18.3
mm.1	23.1	23.7		22		22.2	21.6	21.7	22	22.

Showing 1 to 3 of 3 entries

Previous 1 Next

Descarga Resultado

Columna de análisis

s27001

Tipo de gráfico

Figura 10: Operación de un solo resultado (mediana)

Acción

Operacion

Operación

quantile

Probabilidades separadas por '

0,0.25,0.5,0.75,1

Agrupamiento

☒ Year

☒ Month

☒ Day

Nombre de su operación

qq

Ejecuta

Salidas **Manual**

Show 10 entries

Search:

	s27001	s27004	s27006	s27012	s27021	s27028	s27040	s27046	s27047	s27050	s27056	s27059	s27063	s27069
Lon	-91.54	-91.49	-91.27	-91.76	-91.29	-92.14	-91.16	-91.43	-91.43	-92.6	-91.54	-91.18	-91.07	-91.78
Lat	17.81	17.45	17.64	17.74	17.76	18.09	17.79	17.47	17.47	18.38	17.81	17.94	17.6	17.86
qq.0%	10	10		9.5		11	-3	8.1	7.7	13	9.2	7.9	11.5	11.5
qq.25%	21.5	21.7		20.3		20.7	19.7	19.8	20	20.9	20	19.5	20	20
qq.50%	23.1	23.7		22		22.2	21.6	21.7	22	22.3	22	21.4	21.4	21.4
qq.75%	24.4	24.6		23		23.2	23	23	23.1	23.2	23	22.6	22.2	22.2
qq.100%	34	29.4		28		29.7	34.8	29	29	30	27.5	29.5	26	26.5

Showing 1 to 7 of 7 entries

Previous 1 Next

Descarga Resultado

Columna de análisis

s27001

Tipo de gráfico

Figura 11: Cuartiles para cada una de las columnas

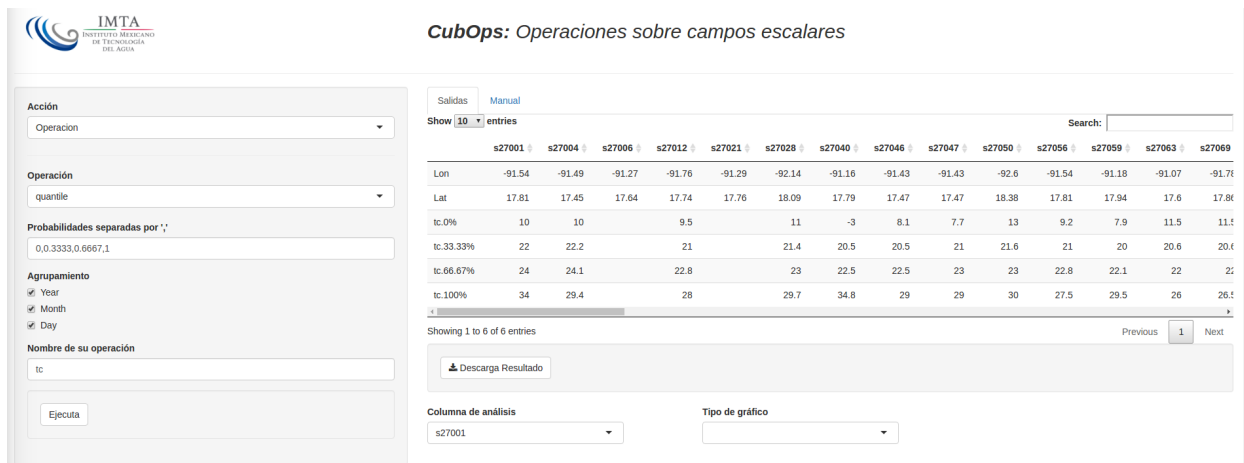


Figura 12: Terciles para cada una de las columnas

los datos, esto es, todos, tienen un valor menor o igual que 34° . En este caso, la mediana de los datos es 23.1° , lo que significa que el 50 % de los datos es menor o igual a este valor (qq.50 %).

Cambiar las probabilidades en el campo de texto, permite desarrollar el calculo de otro tipo de cuantiles. Por ejemplo, si se quisieran calcular los terciles, se podría establecer la siguiente serie de probabilidades: 0, 0.3333, 0.6667, y 1.0. El resultado de esta operación se muestra en la Fig. 12.

Nuevamente, en cualquiera de estos casos, la tabla resultante, por ejemplo, con cuartiles o terciles, puede ser descargada en un equipo local, en el formato estándar CSV, al oprimir el botón “**Descarga Resultado**”.

Nótese además, que las columnas que no tenían información originalmente, tampoco reportan resultados en ninguno de los dos casos ejemplificados.

6. Estratificaciones de datos y operaciones sobre ellas

Al aplicar las distintas operaciones estadísticas, las fechas, que obran como nombre de los renglones de la tabla, se pueden usar para ejecutar distintas *estratificaciones* de los datos. Se puede definir como un *estrato* de una *estratificación*, al conjunto de renglones que comparten alguna porción de la fecha, por ejemplo, el año y el mes; esto es, por ejemplo, los 31 renglones del mes de marzo de 1961, constituirían un estrato para ese caso. La *estratificación* sería entonces el conjunto de todos los estratos que la componen.

6.1. Definición de una estratificación en CubOps

Después de haber seleccionado alguna variable, y al momento de elegir alguna operación, se puede definir también alguna estratificación a partir de las fechas. Esto se hace, mediante el agrupamiento en los campos de fecha, como se muestra en la Fig. 13.

En el caso de la Fig. 13, se ha elegido el año y el mes, mientras que el día se ha descartado. Esto significa que los estratos se definirán a partir del año y el mes solamente. Esto es, la operación que se elija, la media en este caso, se ejecutará para los conjuntos de renglones que tengan el mismo año y mes, o sea, se estarán obteniendo los promedios mensuales de temperatura mínima a lo largo de todos los años en el registro. Así, la Fig. 14, muestra el resultado de esta operación. Nótese que en las fechas, el campo correspondiente al día se ha reemplazado por un asterisco “*”. Entonces, por ejemplo, el primer dato de la columna etiquetada como “s27001”, se encuentra en el renglón “1951-01-*”; ello indica que la media de temperaturas mínimas durante

Acción

Operacion ▼

Operación

mean ▼

Agrupamiento

☒ Year

☒ Month

☐ Day

Nombre de su operación

Salidas Manual

Show 10 entries

	s27001	s27004	s27006	s27012	s27015
Lon	-91.5366666667	-91.4930555556	-91.2747222222	-91.7630555556	-91.2747222222
Lat	17.81	17.4494444444	17.6377777778	17.7419444444	17.7566666667
1951-01-01	17.6	18		16.6	
1951-01-02	18.7	19.3		17.6	
1951-01-03	18.3	20.3		17.2	
1951-01-04	20.3	22		19	
1951-01-05	19.7	21.1		18.5	
1951-01-06	18.4	17.8		17.4	
1951-01-07	17.9	17		17	
1951-01-08	18.3	18.4		17.3	

Showing 1 to 10 of 21,917 entries

Figura 13: Agrupamientos

el primer mes de año 1951 es de 16.887°, y así son también todas las interpretaciones de todas las columnas y renglones de la tabla resultante.

Nuevamente, esta tabla resultante, puede ser descargada en un equipo local, en el formato estándar CSV, al oprimir el botón “**Descarga Resultado**”.

6.2. Estratificaciones y operaciones con resultados múltiples

Cuando se ejecutan operaciones combinadas con estratificaciones de la tabla, se puede ver en la Fig. 14, que la tabla resultante arroja un renglón para cada estrato de los datos originales. Esto no representa ningún problema si la operación arroja un único resultado, como es el caso de la media, la mediana, o la desviación estándar, por ejemplo. En este punto surge la pregunta: ¿cómo maneja CubOps el caso de operaciones que arrojan más de un resultado?, ya que por cada estrato y estación se tendrían resultados múltiples. La respuesta es que el sistema *expande la columna correspondiente a una estación, a tantas columnas como resultados arroje la operación en cuestión*. La Fig. 15 muestra, como ejemplo, el caso de la producción de *terciles*, por medio de la operación “**quantile**”, para una tabla original de temperaturas mínimas. Es de notarse aquí que los nombres de las columnas están formados por el nombre de la estación original, seguido del carácter “:”, y luego un conjunto de caracteres que permitirán identificar cada caso, que para el ejemplo de los terciles son los porcentajes de probabilidad elegidos.

6.3. Aplicación recurrente de operaciones

Hay ocasiones en que es necesario aplicar una nueva operación al resultado de otra operación. Tal es el caso de los **promedios mensuales de precipitaciones acumuladas durante el mes**. Con este propósito, CubOps ha sido diseñado para aplicar cualquiera de las operaciones disponibles sobre la tabla actualmente visible en el panel de resultados tabulares.



Figura 14: Resultados de una operación estratificada



Figura 15: Terciles para datos estratificados

Acción

Elige Variable

Variable a revisar

Prec

Ejecuta

Salidas Manual

Show 10 entries

Search:

	s27001	s27004	s27006	s27012	s27021	s27028	s27040	s27046	s27047	s27050	s270
Lon	-91.54	-91.49	-91.27	-91.76	-91.29	-92.14	-91.16	-91.43	-91.43	-92.6	-91
Lat	17.81	17.45	17.64	17.74	17.76	18.09	17.79	17.47	17.47	18.38	17
1951-01-01	0	0	0	0	0	0	0	0	0	0	
1951-01-02	0	0	0	0	0	0	0	0	0	0	
1951-01-03	0	0	0	0	0	0	0	0	0	0	
1951-01-04	3.5	0.1	0.7	0.7	0.8	0	0.7	0.7	0.8	0.1	
1951-01-05	0	0	0	0	0	0	0	0	0	0	
1951-01-06	0	0	0	0	0	0	0	0	0	0	
1951-01-07	0	2	4.7	6.2	4.8	7.4	3.9	5.6	5.7	20	
1951-01-08	0	18.5	8.4	9.1	8.3	5.2	6.9	9.9	10.1	9.4	

Showing 1 to 10 of 21,917 entries

Previous 1 2 3 4 5 ... 2192 Next

Descarga Resultado

Figura 16: Selección de la variable precipitación

Por ejemplo, para producir justamente los **promedios mensuales de precipitaciones acumuladas durante el mes**, primeramente se selecciona la variable precipitación de la estructura de información vigente, como se muestra en la Fig. 16.

Ahora se procede a acumular las precipitaciones que ocurrieron cada mes, para ello se selecciona la operación **sum** (sumatoria), y en el **Agrupamiento** se elimina el campo correspondiente al día, con el fin de producir estratos mensuales para todos los años, tal como se muestra en la Fig. 17. Ahí, por ejemplo, en el renglón “1951-03-*” y la columna “s27004”, se encuentra el valor 133.6, lo que significa que durante el mes de marzo del año 1951, se observó una precipitación acumulada durante todo el mes de 133.6 mm en la estación 27004.

Finalmente, para producir los promedios mensuales de esos resultados para toda la serie, se procede a elegir la operación **mean** y se elimina ahora el campo correspondiente al año en el **Agrupamiento**. Después de oprimir el botón **Ejecuta**, se obtiene el resultado que se muestra en la Fig. 18. En el renglón “*-09-*” y columna “s27021”, se observa un valor de 322.705, lo que significa que el promedio de precipitaciones acumuladas mensuales en el mes de septiembre, para la estación 27021, es de 322.705 mm.

7. Gráficas sencillas con CubOps

Como una cualidad adicional, **CubOps** permite generar algunos tipos de gráficas sencillas, a saber: series, histogramas y *boxplots* (diagramas de cajas). Estas gráficas permiten tener una mayor visión general de los resultados que se presentan en las tablas del panel de resultados tabulares. A continuación se indica cómo explotar esta característica del sistema.

7.1. Gráficos de información completa de estaciones (columnas)

Después de elegir alguna de las variables disponibles, como es el caso que se ha presentado en la Fig. 8, se puede proceder a producir alguna de las gráficas disponibles. Para ello es necesario primeramente, seleccionar alguna de las columnas en la tabla y luego el tipo de gráfico deseado. Por ejemplo, en la Fig. 19 se ha seleccionado la columna “s27040” y se ha elegido la producción de un gráfico de tipo “hist” (histograma);

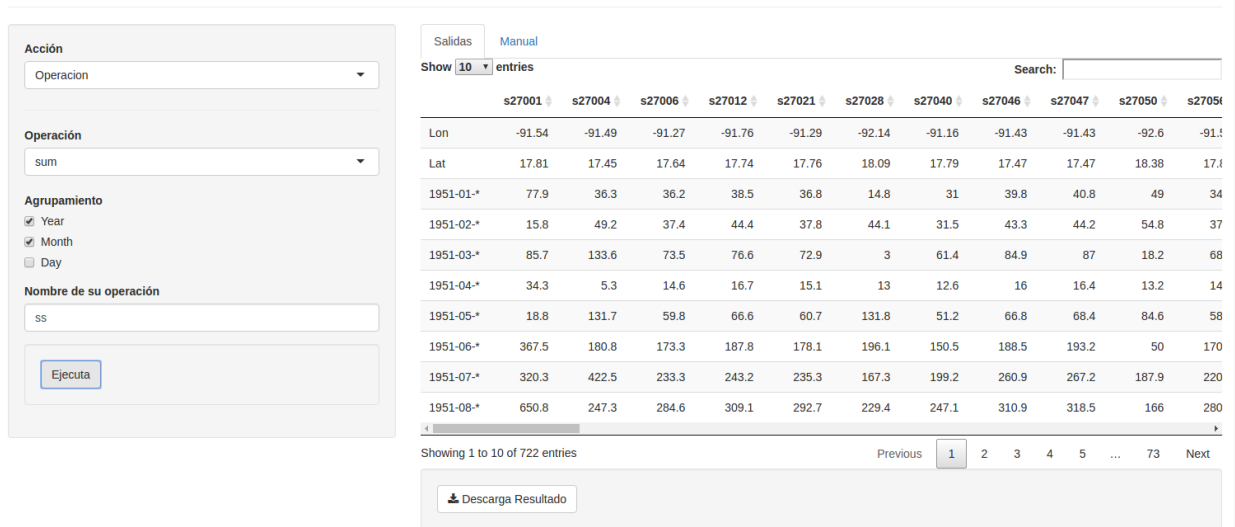


Figura 17: Acumulación de precipitaciones por mes

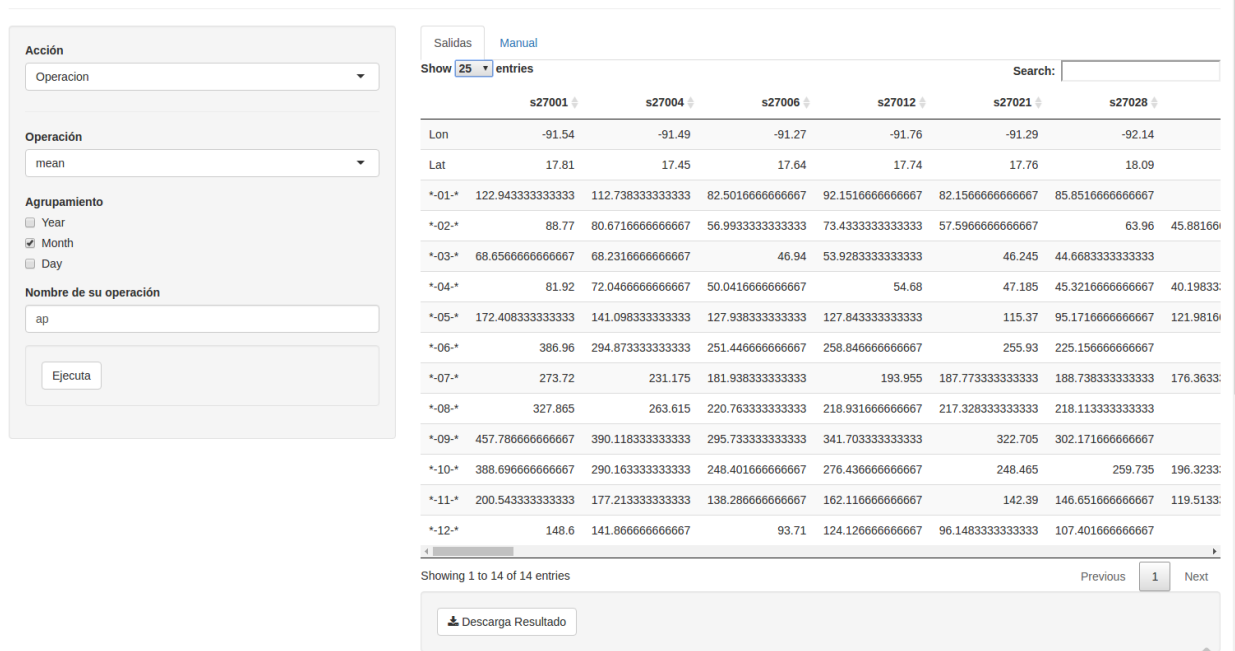


Figura 18: Promedios de precipitaciones acumuladas mensuales

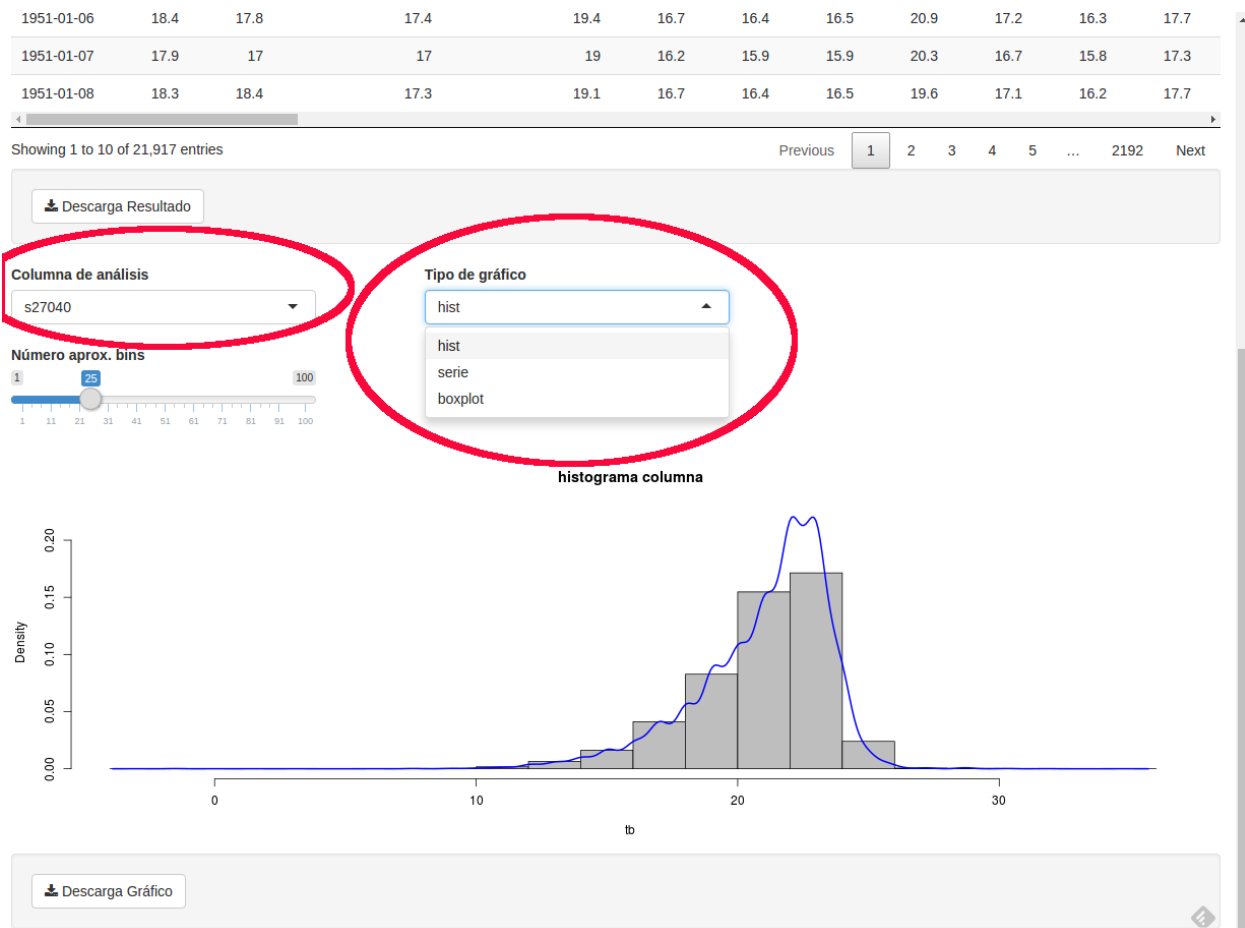


Figura 19: Producción de gráficos con **CubOps**

después de esa selección el gráfico correspondiente se muestra en la parte inferior, conocida como *panel de resultados gráficos*. El gráfico producido se puede salvar en un archivo gráfico, si se oprime el botón etiquetado como **Descarga Gráfico**.

Para el caso particular de los histogramas, aparece además un control, conocido como *slider*, que permite incrementar o disminuir el número de barras que se presentan en el histograma. Al desplazar de botoncito, dinámicamente se incrementan o decrementan las barras que aparecen, como se puede apreciar en la Fig. 20.

7.2. Algunos usos de los gráficos

Cuando ya se ejecutan operaciones en las estaciones o columnas, los gráficos producidos pueden ser más interesantes. Por ejemplo, en la Fig. 21, se observa un gráfico de tipo “serie”, que se ha producido después de haber ejecutado la operación “quantil” sobre todos los datos de temperatura mínima de la columna “s27040”. Aquí, en el eje de las abscisas se tienen los rangos de cuantil, y en las ordenadas, los valores de las temperaturas.

Otro interesante gráfico a considerar son los *boxplots*, que dan también una idea de la distribución de la información. La Fig. 22 muestra un gráfico de este tipo para las temperaturas mínimas de la estación “s27004”.

Un caso interesante es la graficación de la climatología de las precipitaciones acumuladas mensuales, cuyo cálculo se ha presentado en forma tabular en la Fig. 18. Para graficar esto, basta con seleccionar alguna estación y el tipo de gráfico “serie”; la gráfica resultante se muestra en la Fig. 23.

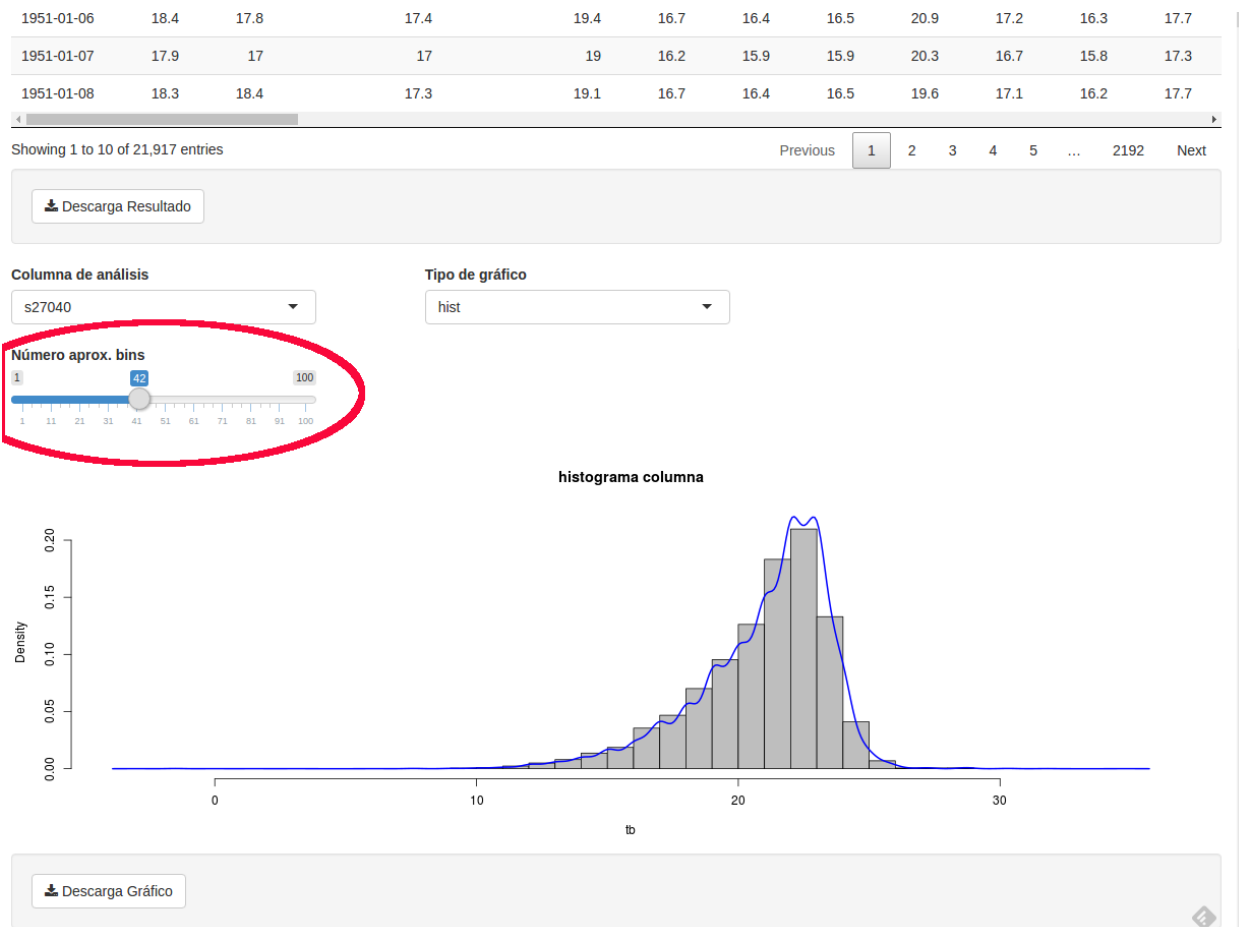


Figura 20: Otros controles en las gráficas

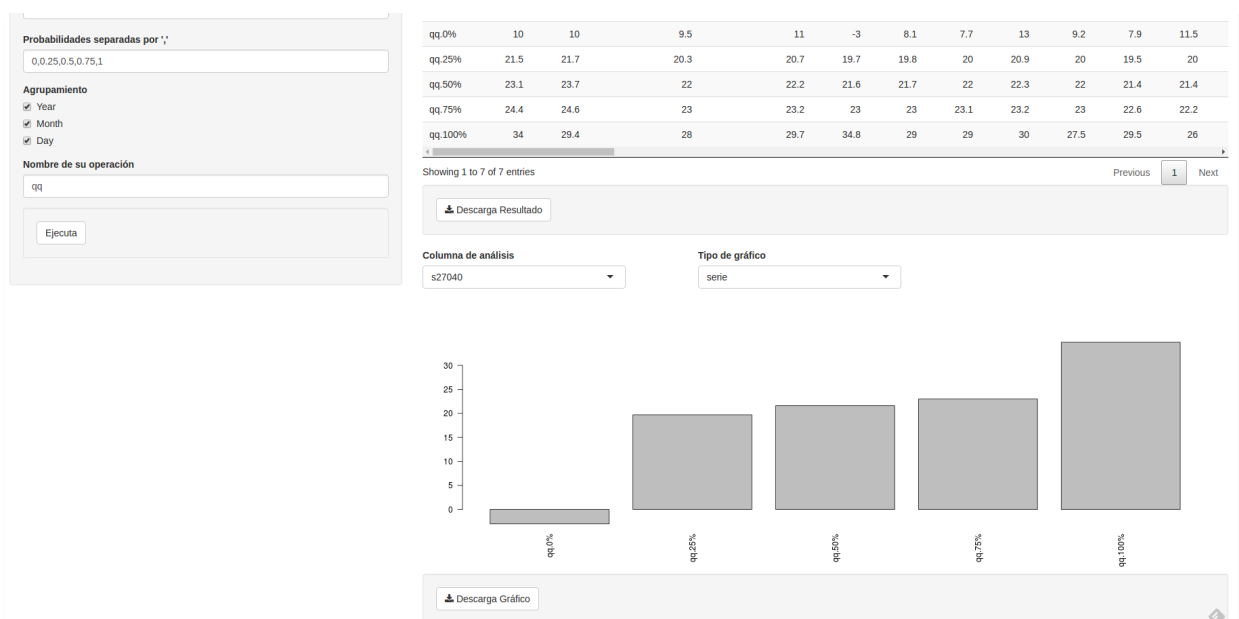


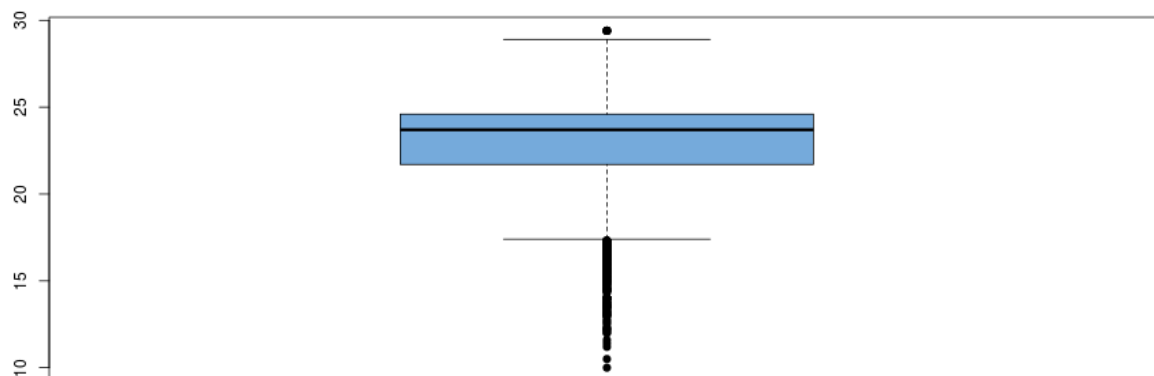
Figura 21: Gráfica de los cuantiles

Columna de análisis

s27004

Tipo de gráfico

boxplot



Descarga Gráfico

Figura 22: Gráfica tipo *boxplot*

-08-	327.865	263.615	220.763333333333	218.931666666667	217.328333333333	218.113333333333
-09-	457.786666666667	390.118333333333	295.733333333333	341.703333333333	322.705	302.171666666667
-10-	388.696666666667	290.163333333333	248.401666666667	276.436666666667	248.465	259.735
-11-	200.543333333333	177.213333333333	138.286666666667	162.116666666667	142.39	146.651666666667
-12-	148.6	141.866666666667	93.71	124.126666666667	96.1483333333333	107.401666666667

Showing 1 to 14 of 14 entries

Previous

1

Next

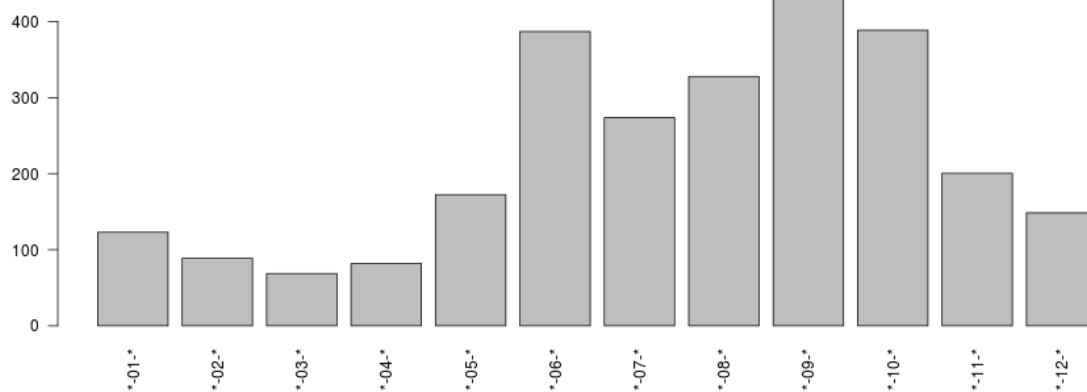
Descarga Resultado

Columna de análisis

s27001

Tipo de gráfico

serie



Descarga Gráfico

Figura 23: Climatología de precipitaciones acumuladas mensuales