

Tarea 03

Raquel Eugenia Meléndez Zamudio

La serie de tiempo utilizada (figura 1) es de la producción eléctrica por mes desde 1985 hasta el 2018, esta tiene un total de 388 instancias, no tiene valores perdidos y a primera vista parece que tiene valores atípicos no tan extremos.

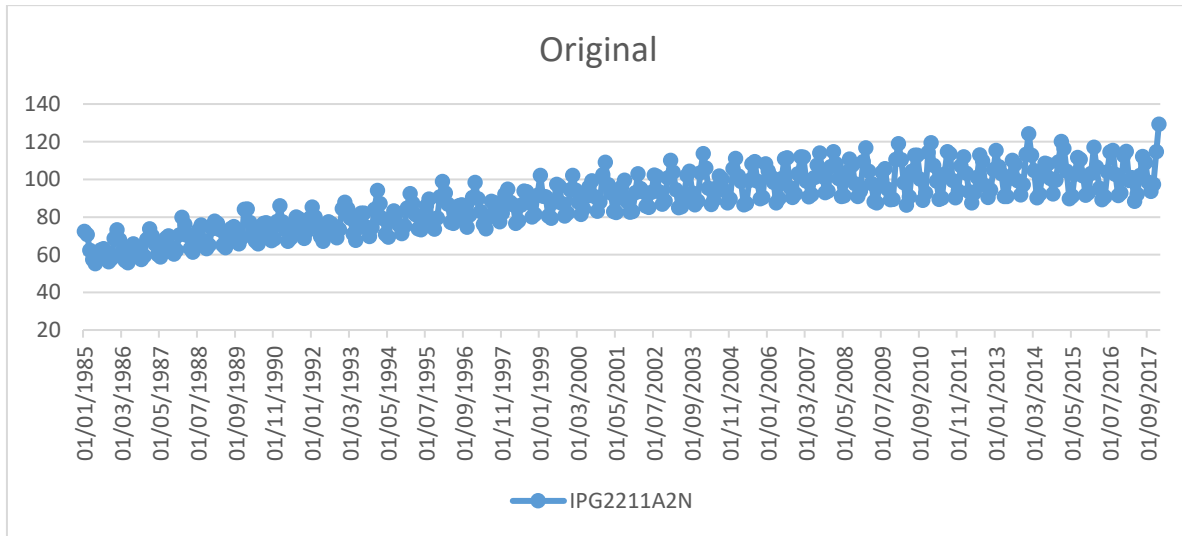


Figura 1. Serie de tiempo original.

Se introdujeron dos valores atípicos extremos en la serie de tiempo, así como tres valores convencionales con comportamiento atípico (figura 2).

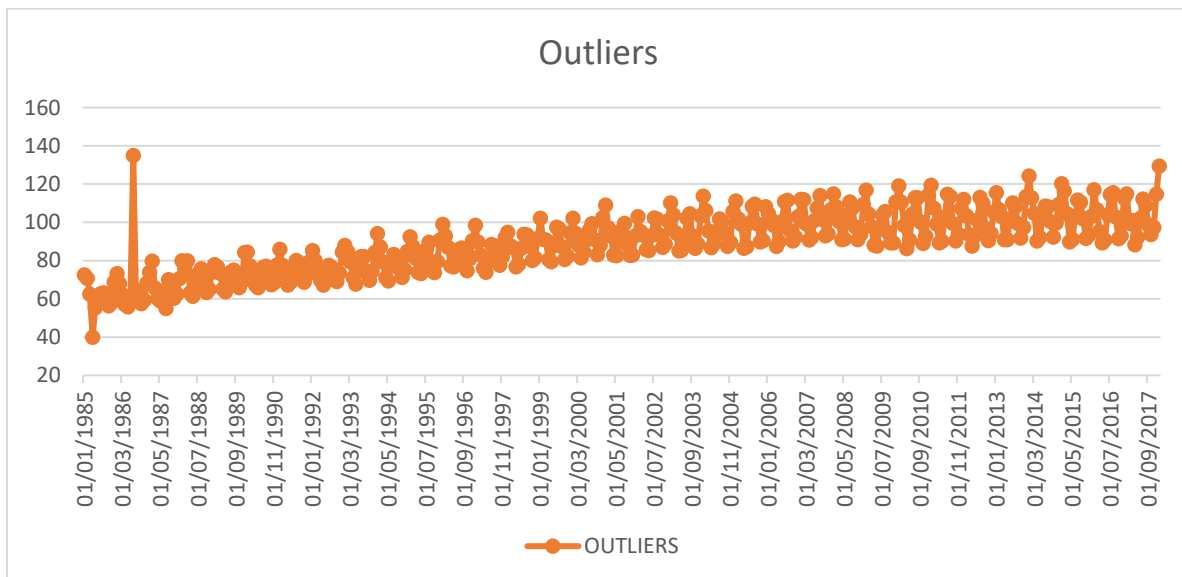


Figura 2. Serie de tiempo con Outliers.

Instituto Politécnico Nacional
Centro de Investigación en Computación
Semestre B22

Así mismo, como se observa en la figura tres, además de los valores atípicos se introdujeron valores perdidos aislados, así como, una serie de 5 valores perdidos.

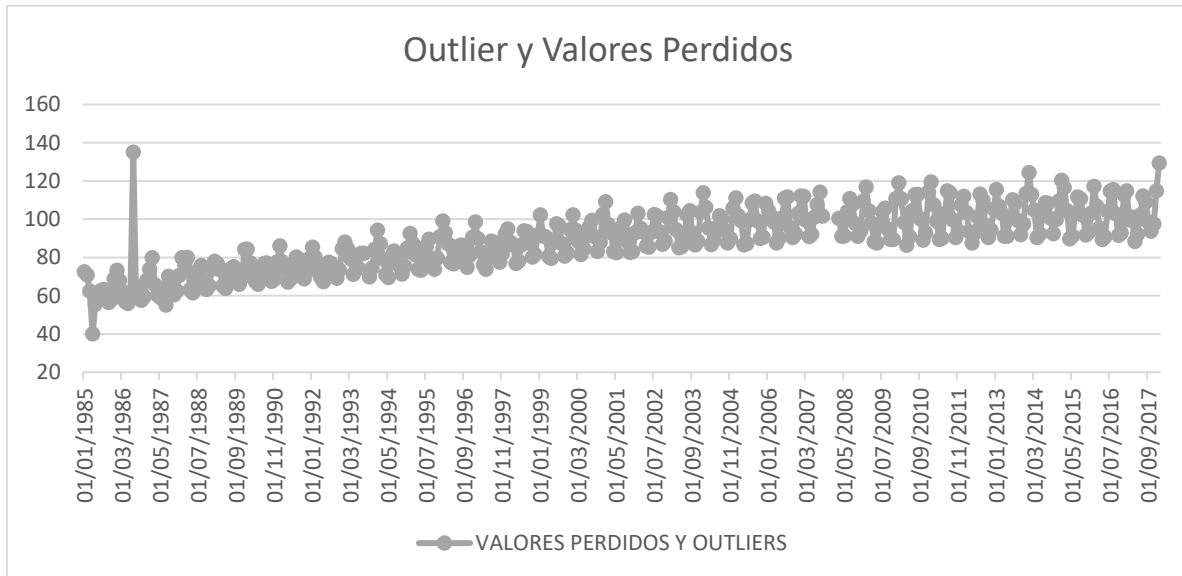


Figura 3. Serie de tiempo con Outliers y valores perdidos.

De las técnicas utilizadas para eliminar valores perdidos y outliers, la primera utilizada fue la de recortar la serie de tiempo (figura 4). Se observa que de los 398 valores existentes para la serie de tiempo únicamente se conservaron 119, así mismo, se observa que al recortar la serie de tiempo se pierde información que podría ser relevante para realizar predicciones.

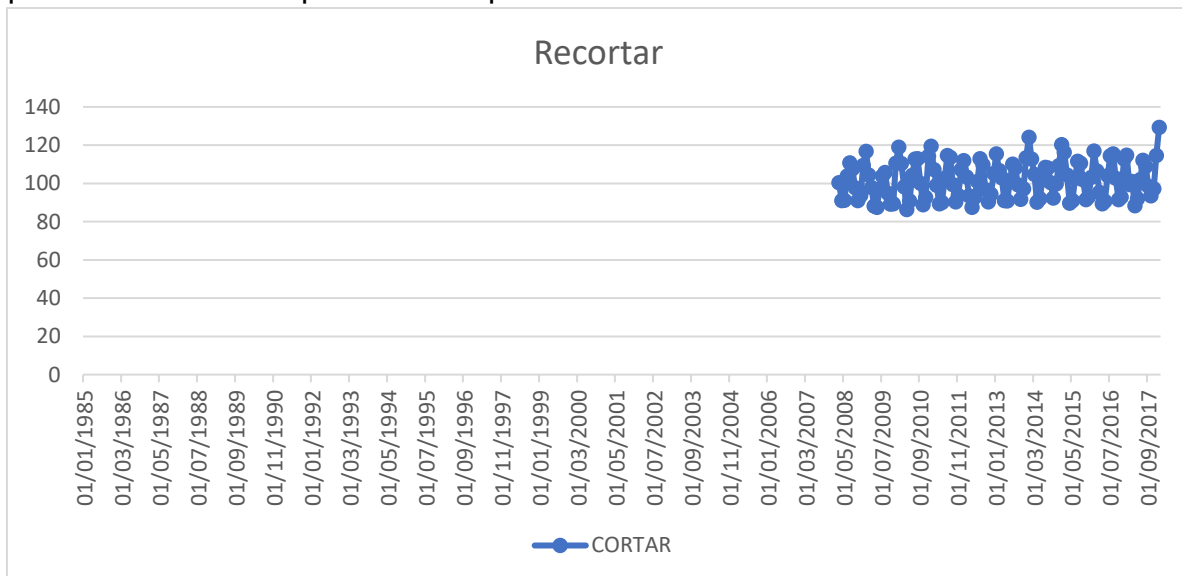


Figura 4. Serie de tiempo recortada.

Instituto Politécnico Nacional
Centro de Investigación en Computación
Semestre B22

Por otro lado, se observa que el remuestreo (figura 5) podría llegar a ser una opción para tratar los outliers y algunos valores perdidos; sin embargo, cuando son segmentos de valores perdidos estos permanecerán de esa forma.

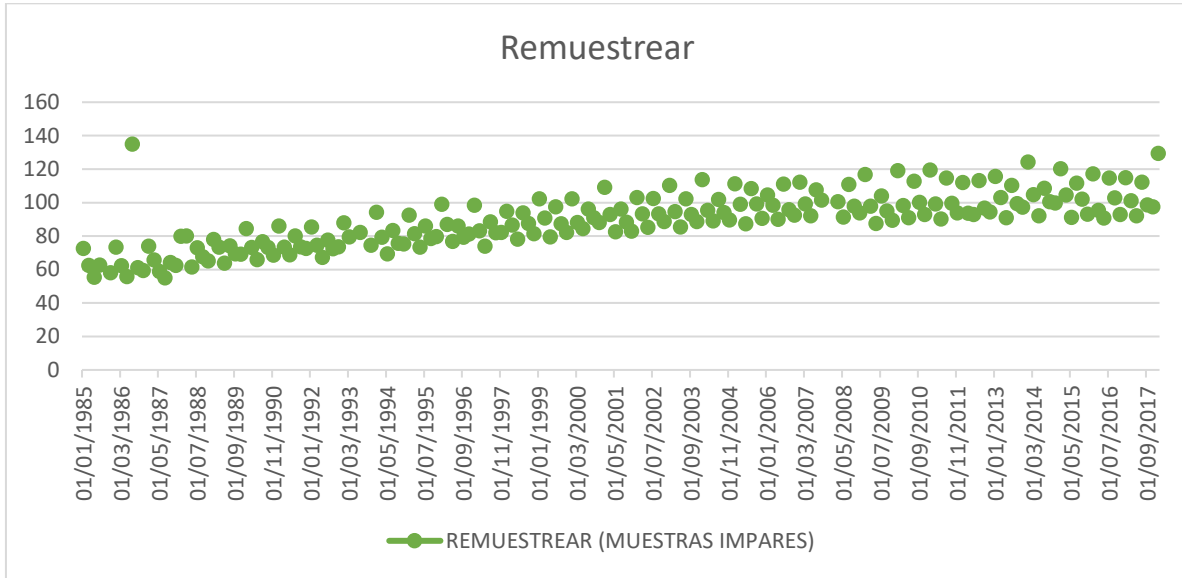


Figura 5. Serie de tiempo con remuestreo por pares.

El recorrer la serie de tiempo (figura 6) no podría ser una opción muy viable, dado que, al ser un fenómeno medido mensualmente, se estaría perdiendo información importante sobre este.

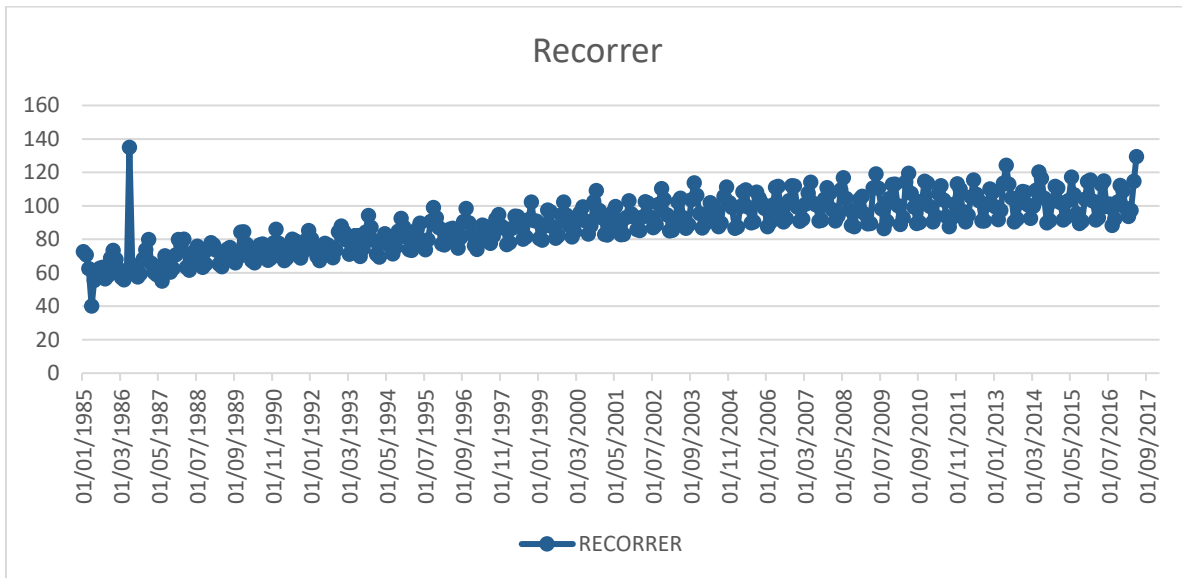


Figura 6. Serie de tiempo recorrida.

Instituto Politécnico Nacional
Centro de Investigación en Computación
Semestre B22

En cuanto a las imputaciones LOCF y NOCF, son técnicas útiles para la eliminación de outliers, así como, de valores perdidos; no se observa una diferencia significativa entre las gráficas de estas dos técnicas (figura 7 y 8).

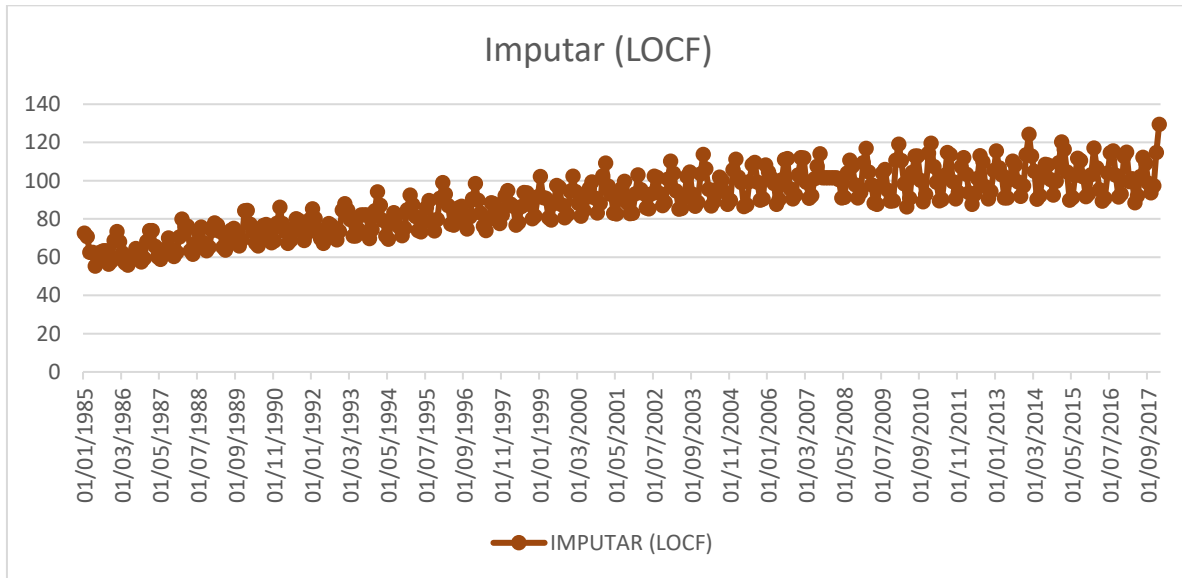


Figura 7. Serie de tiempo con Imputación (LOCF).

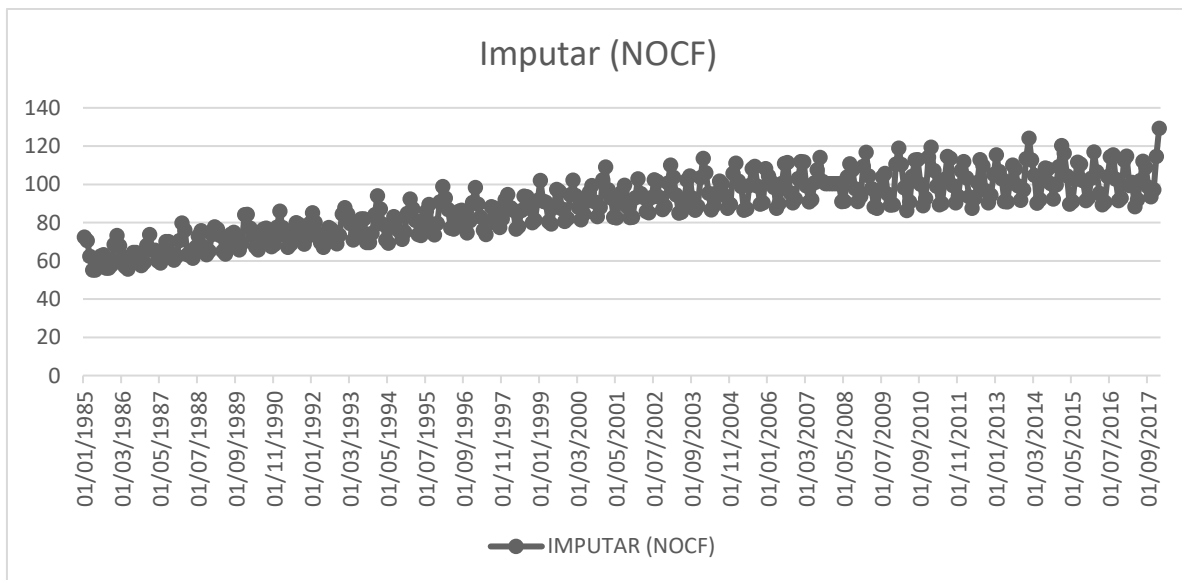


Figura 8. Serie de Tiempo con Imputación (NOCF).

En cuanto a la imputación con media se observa (figura 9) que es funcional para los valores perdidos; sin embargo, si los valores atípicos existentes son muy extremos, estos seguirán estando presente en la serie de tiempo de forma más convencional.

Instituto Politécnico Nacional
Centro de Investigación en Computación
Semestre B22

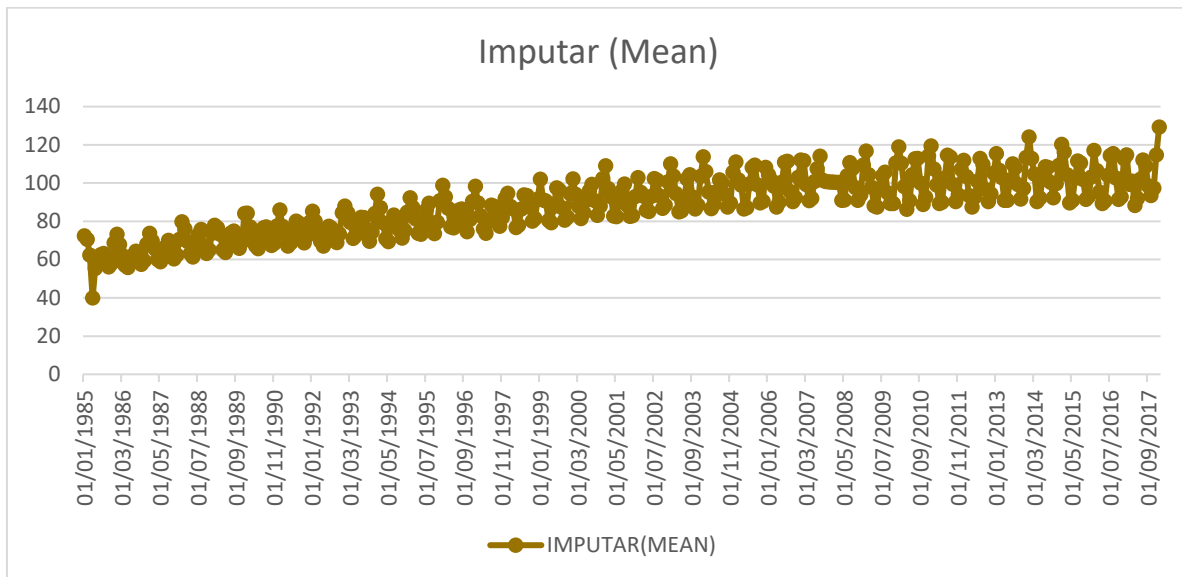


Figura 9. Serie de tiempo con Imputación (Mean).

Finalmente, la técnica de Smooth resulta ser más eficiente entre más valores sean tomados. Como se observa en la figura 10, el smooth con 3 valores muestra aún valores atípicos y si la cantidad de muestras tomadas no es mayor a los segmentos de valores perdidos existentes, estos no se podrán imputar.

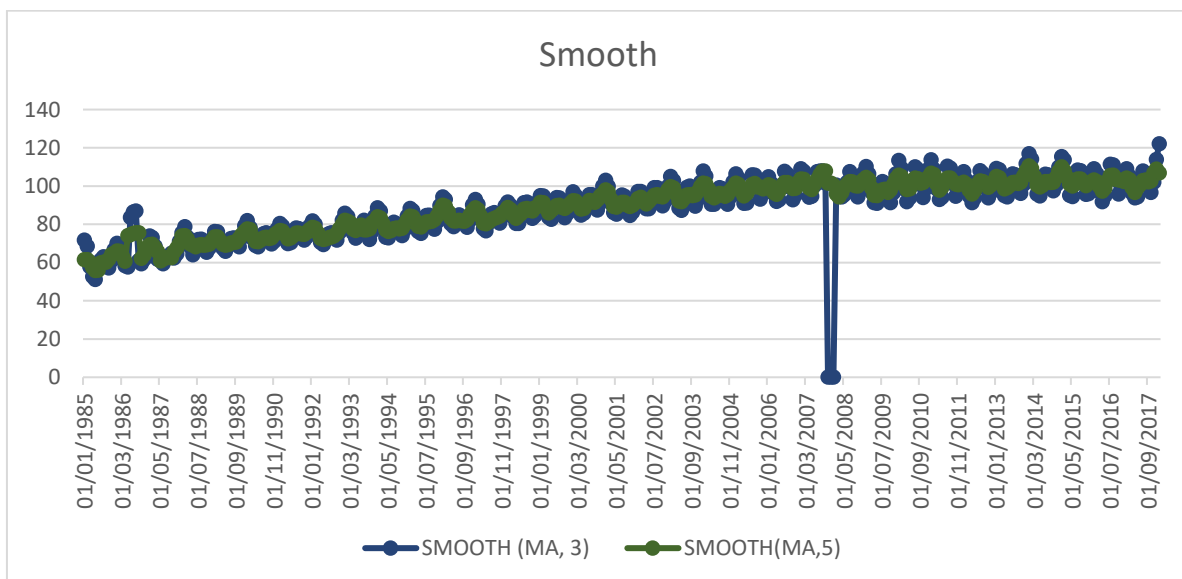


Figura 10. Serie de Tiempo con Smooth (3 y 5).

Como se observa en la tabla 1, la técnica cuyos valores son más parecidos a los originales es la de Smooth con 5 valores, por lo que, esta técnica podría ser la más viable para esta serie de tiempo.

Instituto Politécnico Nacional
Centro de Investigación en Computación
Semestre B22

Tabla 1. Medidas calculadas de las técnicas aplicadas a la serie de tiempo.

	IPG2211A2N	OUTLIERS	VALORES PERDIDOS Y OUTLIERS
min	55.3151	40	40
max	129.4048	135	135
avg	88.8472176	88.9916481	88.97192751
std	15.3878337	15.6482132	15.57363453

	CORTAR	REMUESTREAR (MUESTRAS IMPARES)	RECORRER	SMOOTH(MA,5)
min	86.3	55	40	55.69518
max	129.4048	135	135	110.53152
avg	101.990731	89.57208711	88.97192751	88.94050947
std	9.71999861	15.81752372	15.57363453	13.28063416

IMPUTAR (LOCF)	IMPUTAR (NOCF)	IMPUTAR(MEAN)	SMOOTH (MA, 3)
55.3151	55.3151	40	0
129.4048	129.4048	129.4048	122.063
88.8678199	88.76954887	88.76690527	88.21621788
15.29634999	15.39556983	15.45315352	16.09714218