

Estimación de la velocidad del viento con el método de cuadrados mínimos

Alan Pierri, Luciana Reznik, Matias Domingues

10 de junio de 2014

Resumen

El objetivo del siguiente trabajo fue el cálculo de una función de estimación de la velocidad del viento en Irlanda mediante la técnica de cuadrados mínimos.

Palabras Clave: Cuadrados mínimos, Cholesky, Método QR, Viento en Irlanda

Índice

1. Introducción	3
2. Metodología	3
2.1. Conversión de datos	4
2.2. Cálculo de los coeficientes de la función de aproximación	4
3. Resultados obtenidos	5
3.1. Velocidad media por estación	5
3.2. Coeficientes de la función de aproximación	5
3.3. Error cuadrático medio del ajuste.	5
4. Conclusiones	7
4.1. Histogramas de los errores cometidos.	7
5. Anexo	20
5.1. Método main	20
5.2. Método QR con Gram Shmidt	21
5.3. Método de Cholesky	22
5.4. Cálculo de la velocidad estimada	22
5.5. Sustitución hacia adelante	22

1. Introducción

La meteorología es la ciencia que estudia la atmósfera y su evolución y una de sus principales aplicaciones es la predicción del tiempo futuro, de enorme utilidad en muchas actividades humanas, desde la seguridad en los medios de transporte hasta su empleo para planear nuestras vacaciones, pasando por muchísimas otras aplicaciones.

Para conocer el futuro estado de la atmósfera, a partir de un estado inicial de la misma, se utilizan actualmente como herramienta básica los llamados modelos numéricos de predicción del tiempo. Estos modelos físico-matemáticos simulan el comportamiento de la atmósfera a partir de unos datos iniciales constituidos por las observaciones de las variables atmosféricas (presión, temperatura, humedad, viento, etc.). Sobre estos datos iniciales se aplican (se integran) las ecuaciones físico-matemáticas que definen su evolución temporal. El resultado será el valor de esas variables en un instante futuro. [1]

En el siguiente trabajo, se intenta obtener una función de aproximación con el método de cuadrados mínimos para predecir la velocidad del viento en Irlanda, a partir de los datos obtenidos en 12 estaciones meteorológicas entre los años 1961 y 1978. En la sección de metodología se explicarán los pasos realizados para ello, y luego se presentarán los resultados obtenidos y las conclusiones del trabajo.

2. Metodología

Para obtener la estimación de la velocidad del viento en función del tiempo en días, se utilizó el método de cuadrados mínimos, que trata de aproximar los valores obtenidos empíricamente a una curva determinada. Para ello, se utilizaron dos implementaciones diferentes. Por un lado el método de QR utilizando Gram-Schmidt y por el otro, el método de Cholesky. Para la comparación de los datos, también se utilizó la función propia que provee Matlab.

La función a partir de la cual se ajustaron los datos, fue:

$$V(t) = A_0 + A_1 \cos(2\pi f_1 t) + B_1 \sin(2\pi f_1 t) \quad (1)$$

siendo

$$f_1 = 1/365,25 * \text{día}^{-1} \quad (2)$$

El período elegido de 365.25 días corresponde a la duración media de un año y fue la unidad elegida ya que año tras año, el clima se supone periódico. Tomar un período menor, incurriría en el error de pensar que por ejemplo, las estaciones del año no modifican los comportamientos de las velocidades de los vientos. En el siguiente gráfico, podemos ver la periodicidad del viento y su variación a lo largo de los meses para los primeros tres años de muestreo a modo de ejemplificación.

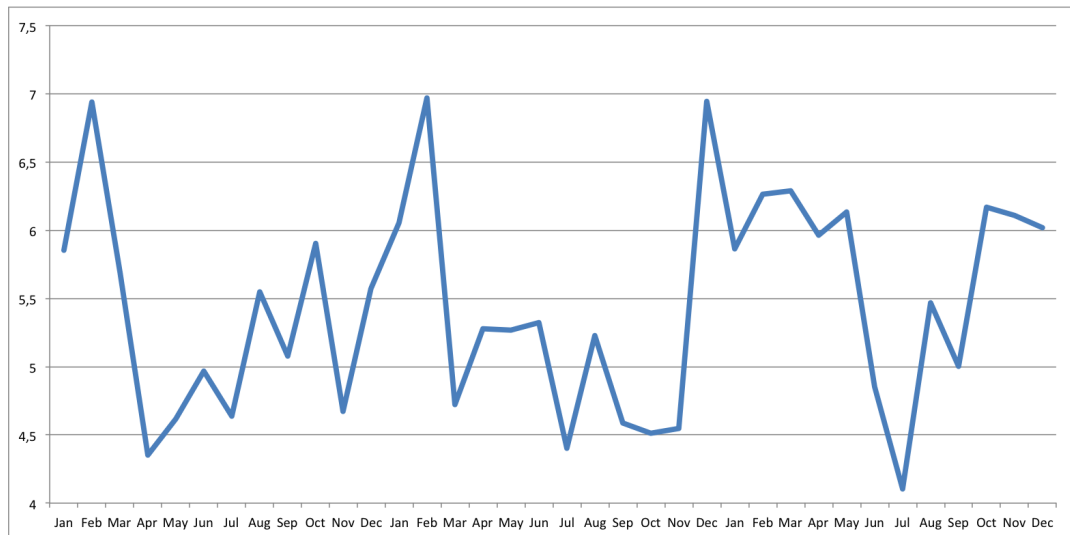


Figura 1: Velocidad promedio de todas las estaciones por mes

2.1. Conversión de datos

Los datos obtenidos por las estaciones meteorológicas, se encontraban en la unidad nudos. Para el mejor manejo de los mismos, se comenzó con la conversión de todos ellos a m/s.

2.2. Cálculo de los coeficientes de la función de aproximación

Para el cálculo de los coeficientes A_0 , A_1 y B_1 de la función (1), lo primero que se realizó fue la separación de los datos por estación y se iteró por cada una de las estaciones para obtener la función de aproximación de acuerdo al método elegido (QR, Cholesky o Matlab). Por último, se calculó el error cuadrático medio con cada método para comparar su eficacia.

3. Resultados obtenidos

3.1. Velocidad media por estación

Estación	$V(t)(m/s)$
1	6.35707
2	5.47711
3	5.99067
4	3.23929
5	5.37646
6	3.64286
7	5.03213
8	4.36409
9	4.36889
10	4.47136
11	6.74920
12	8.03229

Cuadro 1: Valores de las velocidades medias obtenidas

3.2. Coeficientes de la función de aproximación

Estación	A_0	A_1	B_1
1	6.3604	-0.0169	0.0367
2	5.4770	-0.0443	-0.0006
3	5.9985	0.0347	0.0843
4	3.2442	-0.0088	0.0536
5	5.3795	-0.0577	0.0334
6	3.6486	-0.0208	0.0621
7	5.0399	0.0318	0.0842
8	4.3699	-0.0181	0.0632
9	4.3706	-0.0109	0.0189
10	4.4794	-0.0310	0.0876
11	6.7500	0.0051	0.0089
12	8.0251	-0.0218	-0.0781

Cuadro 2: Coeficientes obtenidos con la implementación de QR

3.3. Error cuadrático medio del ajuste.

Estación	A_0	A_1	B_1
1	6.3604	-0.0169	0.0367
2	5.4770	-0.0443	-0.0006
3	5.9985	0.0347	0.0843
4	3.2442	-0.0088	0.0536
5	5.3795	-0.0577	0.0334
6	3.6486	-0.0208	0.0621
7	5.0399	0.0318	0.0842
8	4.3699	-0.0181	0.0632
9	4.3706	-0.0109	0.0189
10	4.4794	-0.0310	0.0876
11	6.7500	0.0051	0.0089
12	8.0251	-0.0218	-0.0781

Cuadro 3: Coeficientes obtenidos con la implementación de Cholesky

Estación	A_0	A_1	B_1
1	6.35707081	-0.01820940	0.03674981
2	5.47710517	-0.04429044	-0.00060650
3	5.99066747	0.03156476	0.08427996
4	3.23929484	-0.01074494	0.05357270
5	5.37646141	-0.05888658	0.03337499
6	3.64286185	-0.02311715	0.06211073
7	5.03212609	0.02872138	0.08423441
8	4.36409277	-0.02040780	0.06322703
9	4.36889099	-0.01160855	0.01893036
10	4.47135695	-0.03417731	0.08758244
11	6.74920424	0.00476971	0.00888903
12	8.03228508	-0.01890634	-0.07813697

Cuadro 4: Coeficientes obtenidos con la implementación de Matlab

Estación	ECM Matlab	ECM propio
1	5.49E+04	5.7230E+04
2	4.82E+04	4.9984E+04
3	4.35E+05	4.5648E+04
4	2.26E+04	2.3200E+04
5	4.23E+04	4.4001E+04
6	2.73E+04	2.8138E+04
7	4.31E+04	4.4515E+04
8	3.52E+04	3.6279E+04
9	3.02E+04	3.6279E+04
10	3.53E+04	3.6388E+04
11	5.92E+04	6.1826E+04
12	7.80E+04	8.1695E+04

Cuadro 5: Comparación del ECM por estación con la función de Matlab y las propias

4. Conclusiones

Podemos observar que con los ambos métodos propios utilizados, los valores de los coeficientes A_0 , A_1 y B_1 son prácticamente iguales. Eso significa, que se puede utilizar cualquiera de ellos indistintamente. También tardan un tiempo muy similar, por lo que no habría diferencia entre uno y el otro.

Sin embargo, a la hora de comparar los resultados obtenidos con los que proporciona la función propia de Matlab, pudimos ver que hay una pequeña diferencia a favor de la de Matlab, que como se puede ver en la tabla de ECM comete un error cuadrático medio un poco menor en todas las estaciones.

Por otro lado, el valor de A_0 tiende siempre al valor medio de la velocidad de las estaciones meteorológicas, ya que la suma de los coeficientes de la serie de Fourier

$$A_1 \cos(2\pi f_1 t) + B_1 \sin(2\pi f_1 t) \quad (3)$$

tienden a anularse.

4.1. Histogramas de los errores cometidos.

En las siguientes figuras, se graficó un histograma del valor absoluto de los errores cometidos divididos en 10 intervalos. Se tomó el valor absoluto para intentar ver cuál fue la desviación respecto del valor esperado tanto en valores positivos como negativos.

Podemos observar que para todas las estaciones la distribución de los errores es muy similar y la mayoría de los mismos se encuentran en el rango de valores de 0 a 5 en valor absoluto.

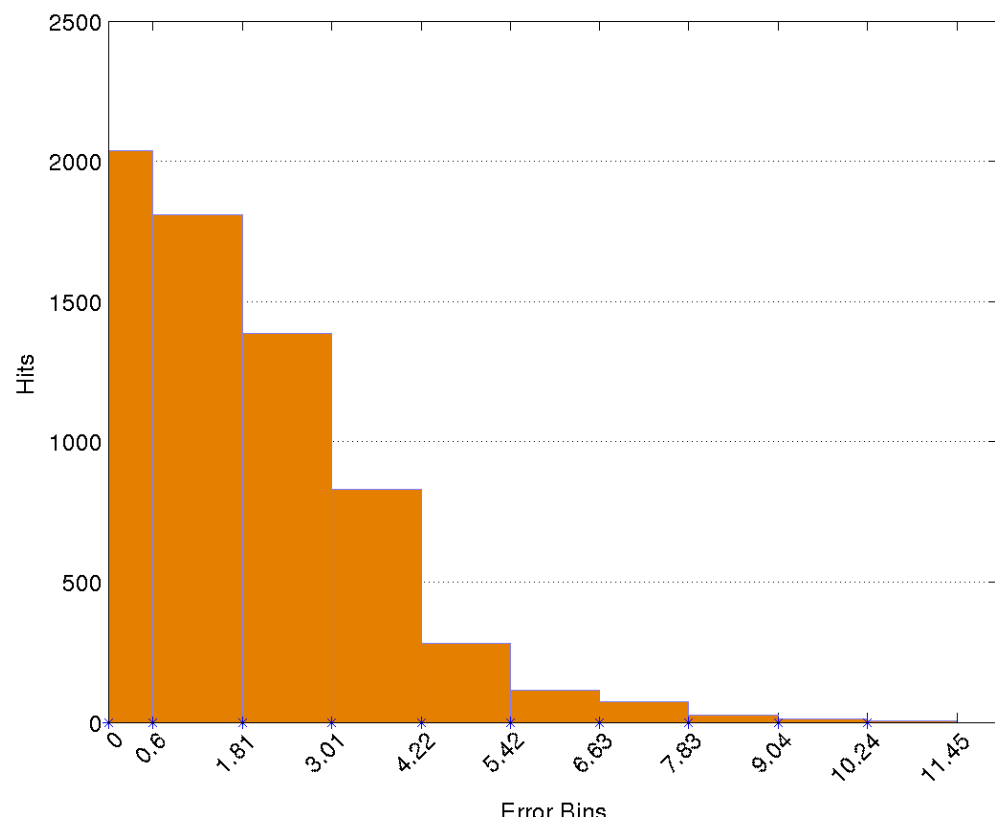


Figura 2: Error para los datos obtenidos en la estación 1

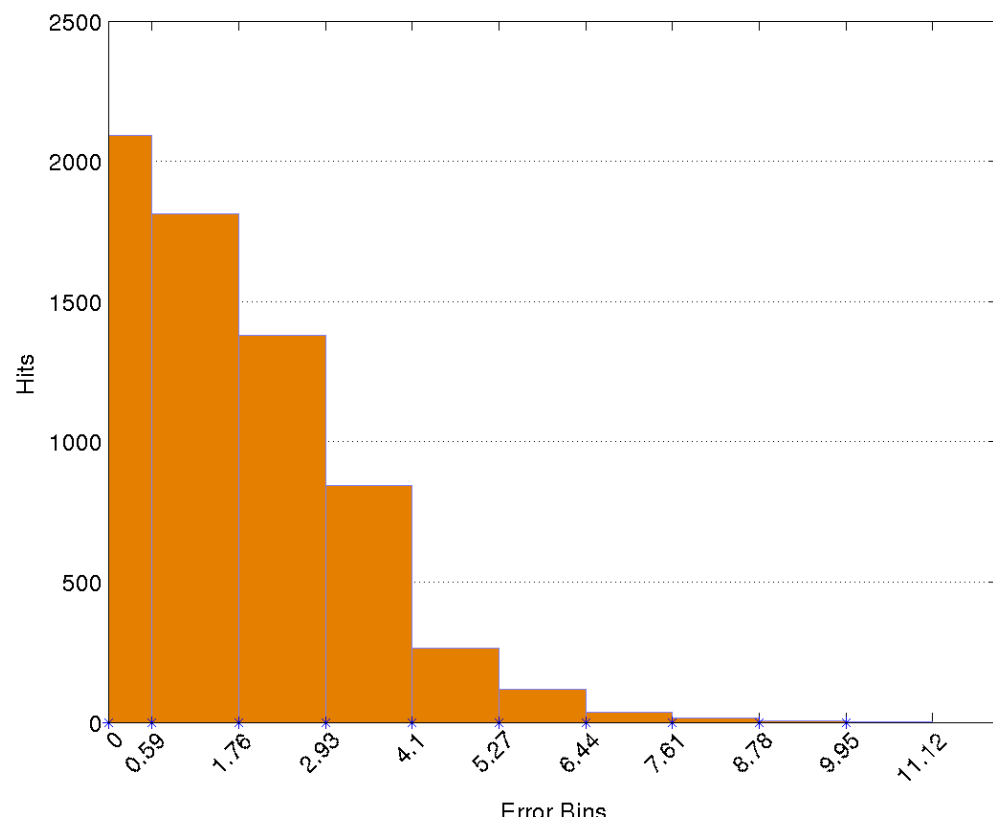


Figura 3: Error para los datos obtenidos en la estación 2

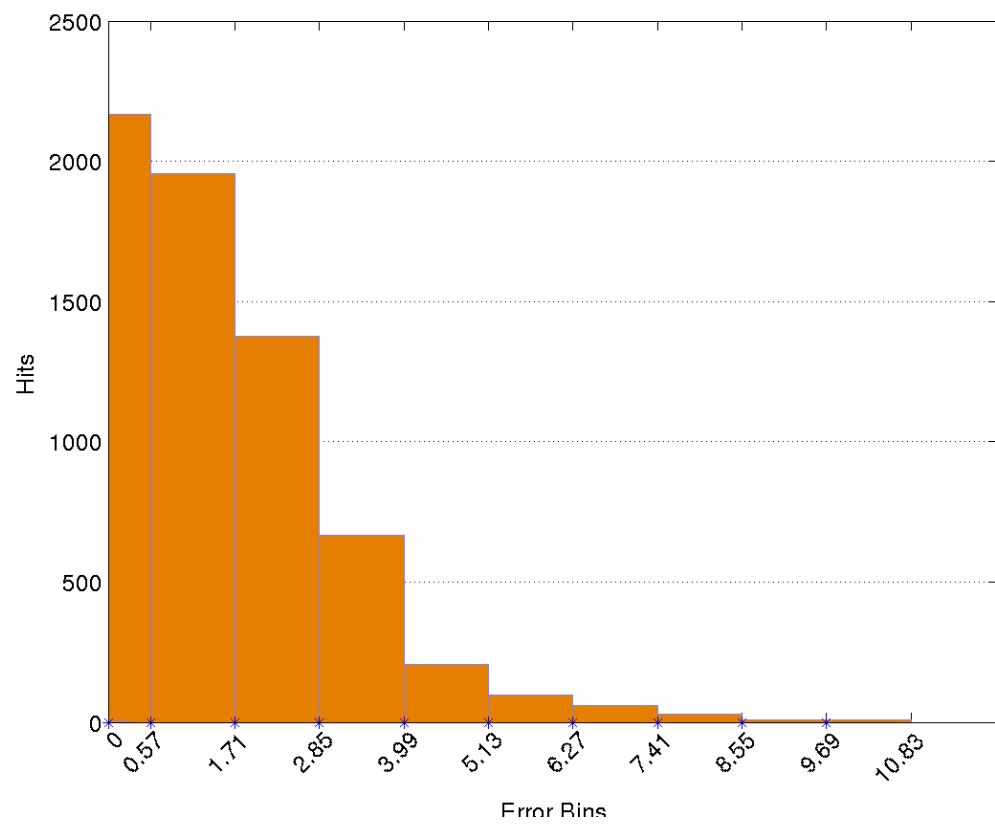


Figura 4: Error para los datos obtenidos en la estación 3

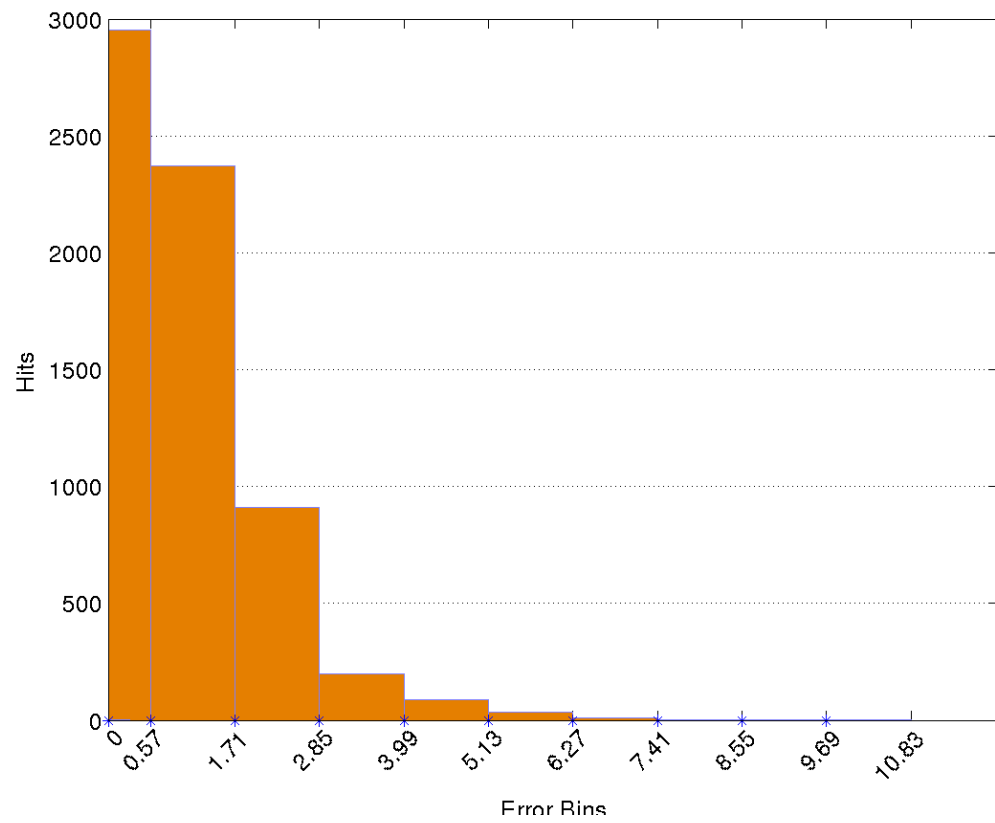


Figura 5: Error para los datos obtenidos en la estación 4

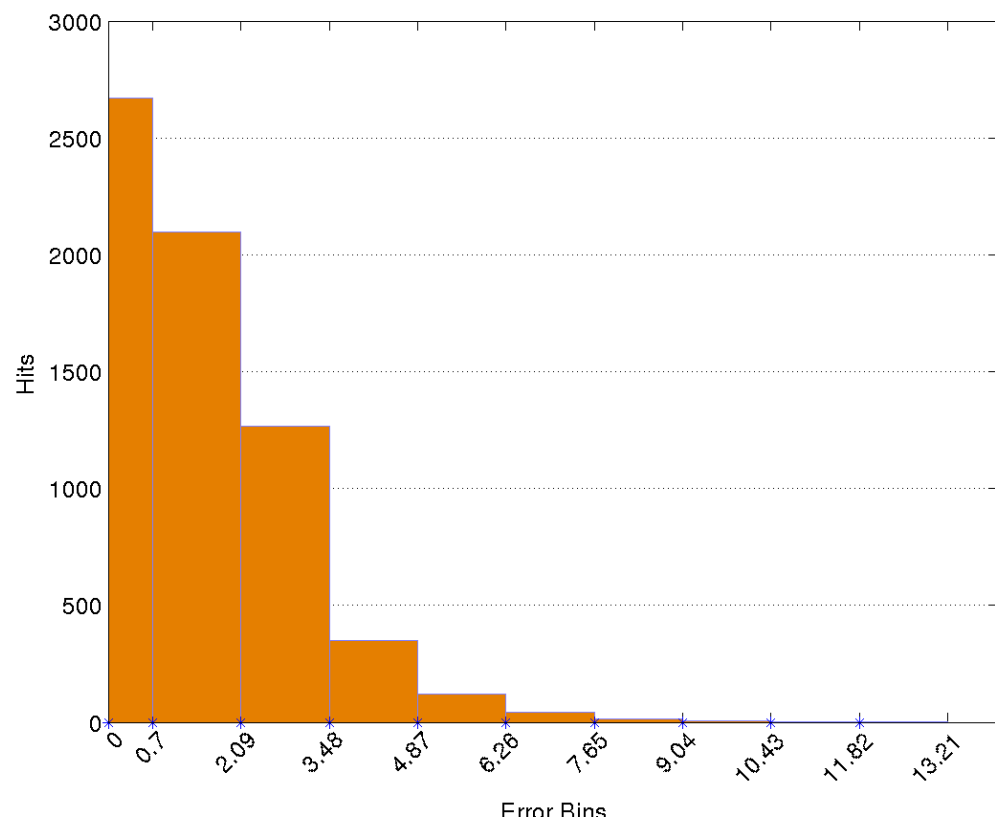


Figura 6: Error para los datos obtenidos en la estación 5

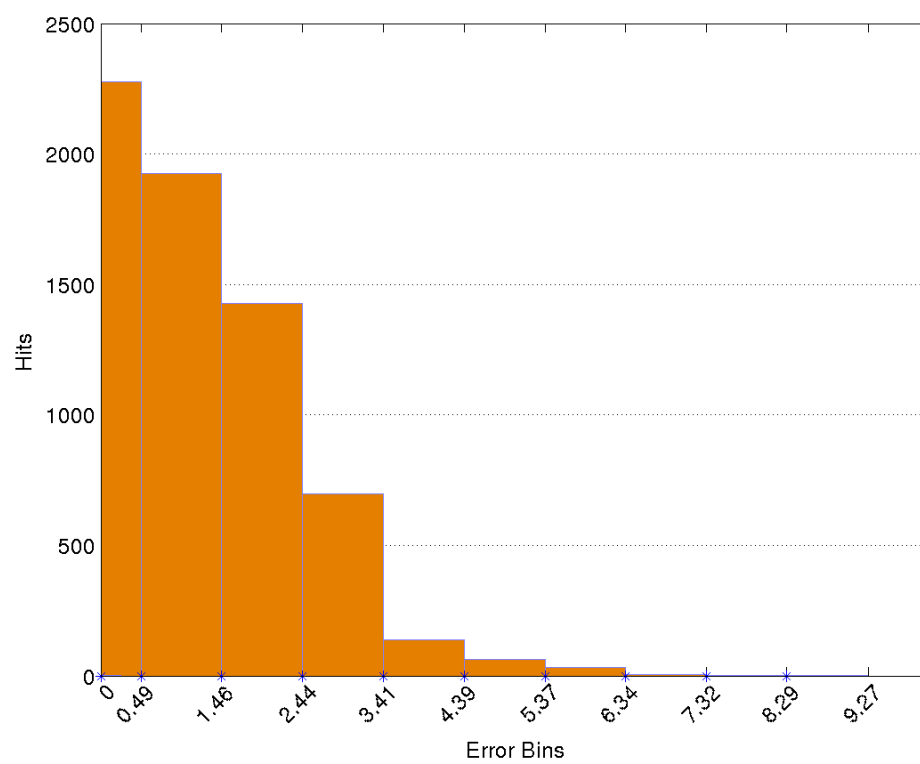


Figura 7: Error para los datos obtenidos en la estación 6

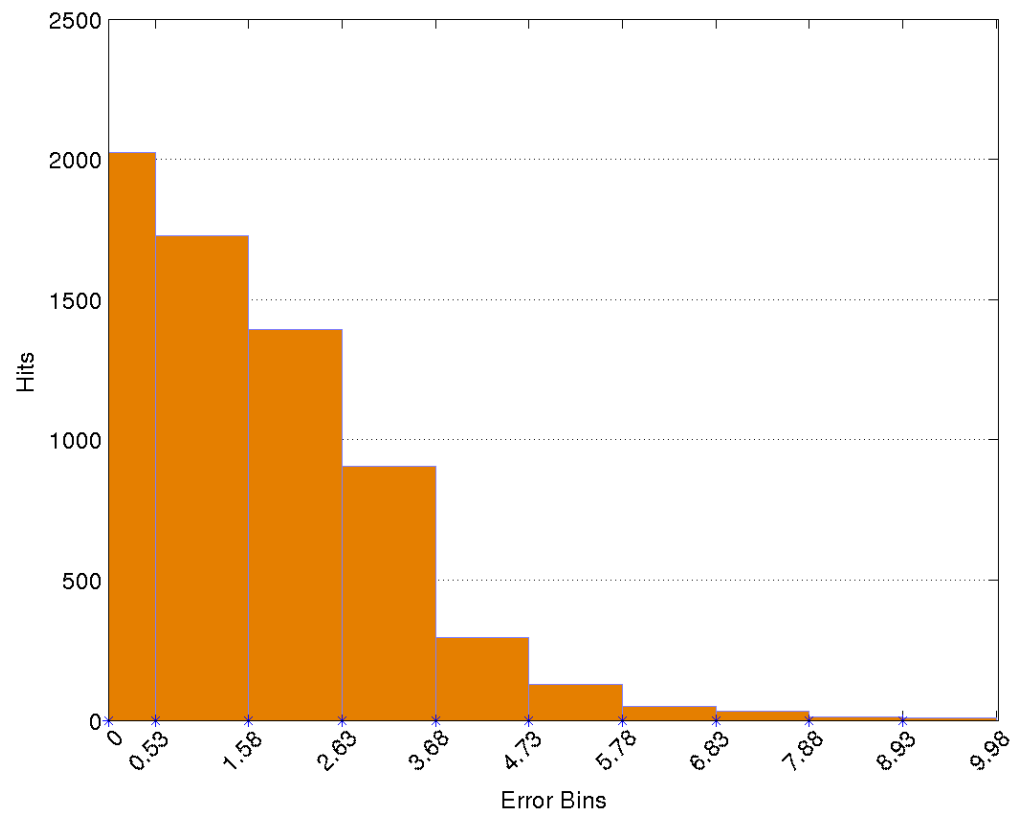


Figura 8: Error para los datos obtenidos en la estación 7

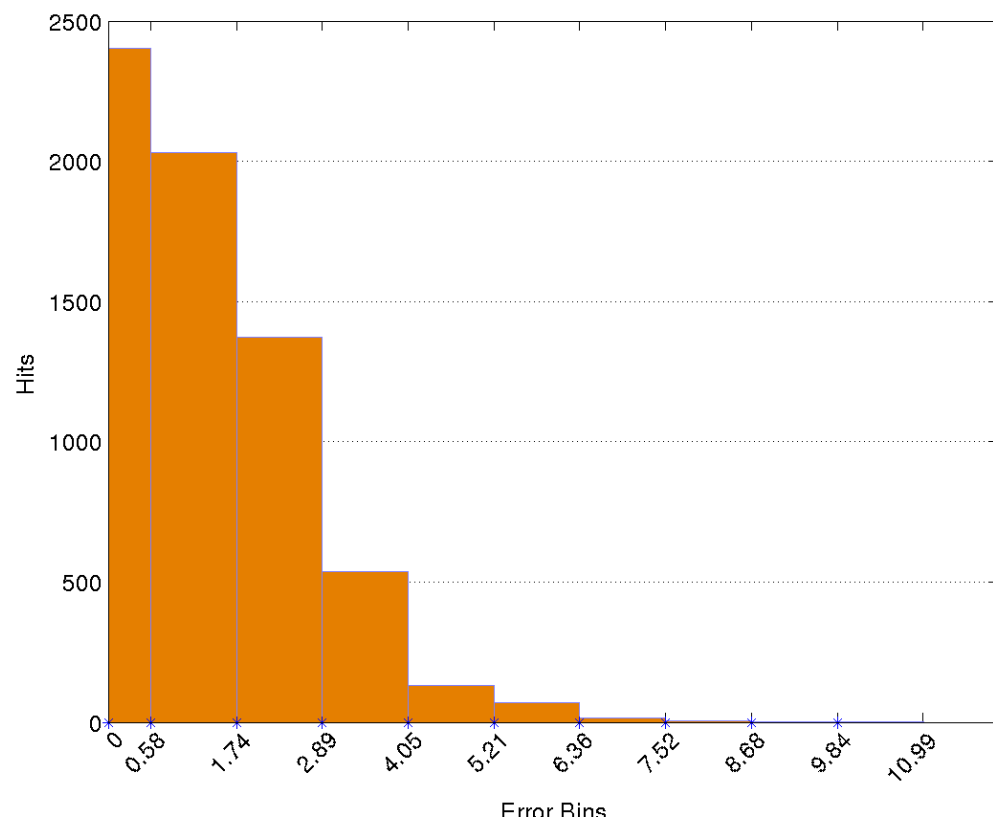


Figura 9: Error para los datos obtenidos en la estación 8

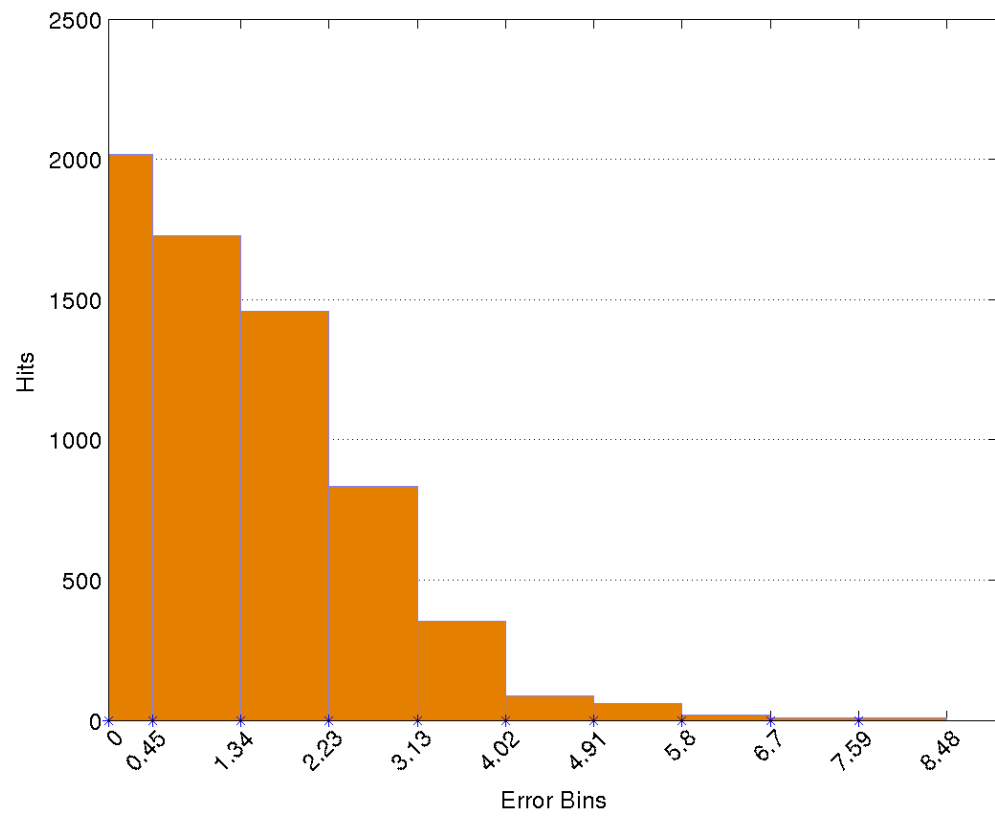


Figura 10: Error para los datos obtenidos en la estación 9

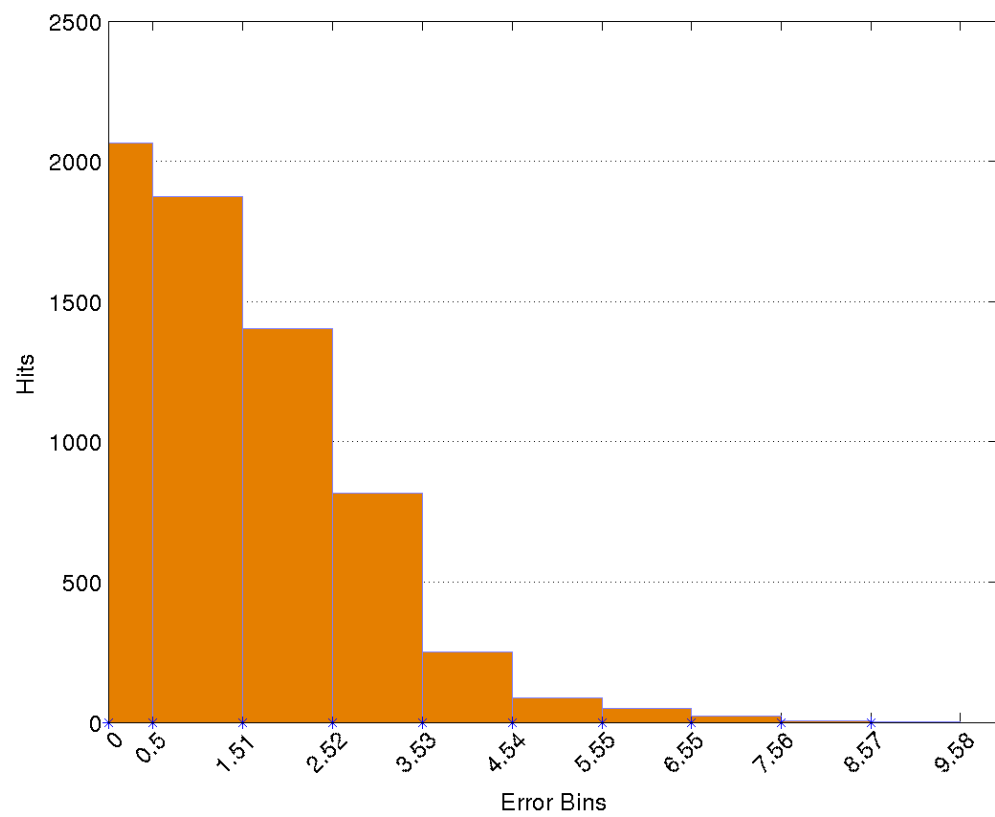


Figura 11: Error para los datos obtenidos en la estación 10

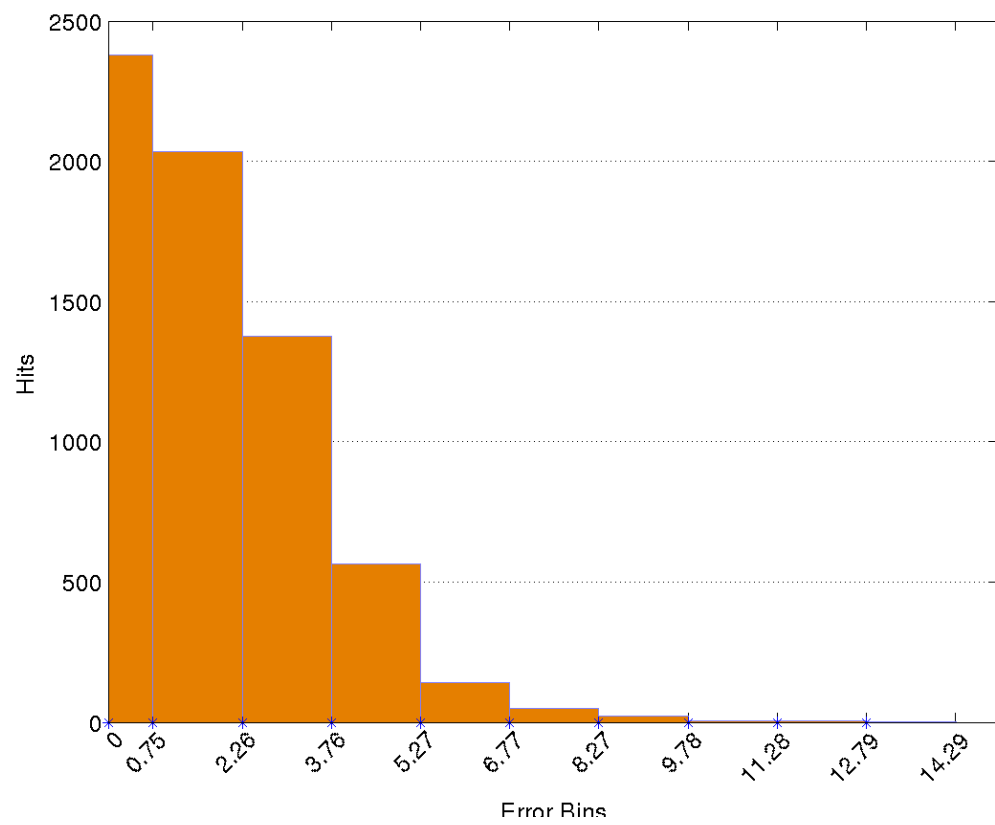


Figura 12: Error para los datos obtenidos en la estación 11

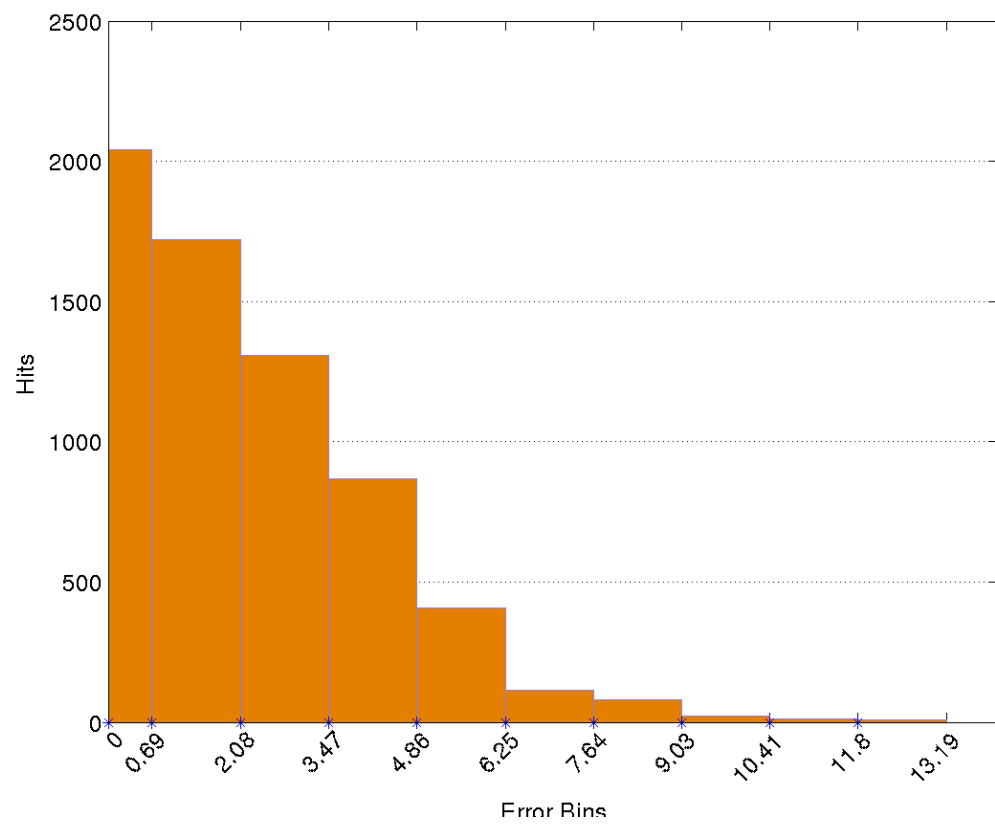


Figura 13: Error para los datos obtenidos en la estación 12

5. Anexo

Los principales códigos utilizados fueron:

5.1. Método main

execute.m

```
1 function ans = execute(data,f)
2     ans = struct ();
3     column = data(:,1:3);
4
5     for i=4:size(data(1,:),2);
6         %para cada estacion armo una matriz con a o , mes, d a ,
           velocidad media
7         column(:,4) = data(:,i);
8         %calculo y guardo en A las constantes en funci n del
           tiempo que
9         %multiplican a A0, A1, B1
10        [A, y] = calculateMatrix(column);
11
12        %si se eligi metodo QR
13        if f == 0
14            [Q,R] = solveQRGS(A);
15
16            X(i-3, :) = fSubstitution(R,Q'*y);
17
18            [k,m] = histogram(y,A,X(i-3, :));
19            ans.(char('@' + i)).x = k;
20            ans.(char('@' + i)).m = m;
21
22        %si se eligi metodo Cholesky
23        elseif f == 1
24            B = A' * A;
25            G = cholesky(B);
26
27            p = fSubstitution(G,A'* y);
28            X(i-3, :) = fSubstitution(G,p);
29
30            [k,m] = histogram(y,A,X(i-3, :));
31            ans.(char('@' + i)).x = k;
32            ans.(char('@' + i)).m = m;
33
34        %si se eligi el metodo de Matlab
35        else
36            X(i-3, :) = A\y;
37            [k,m] = histogram(y,A,X(i-3, :));
38            ans.(char('@' + i)).x = k;
39            ans.(char('@' + i)).m = m;
40
41        end
42        %calculo el error absoluto obtenido
43        err(i)=calculateError(A,X(i-3, :),y);
44
45    end
46
47    err = err(:,4:end)';
```

```

48     ans.X = X;
49
50 end

```

5.2. Método QR con Gram Shmidt

solveQRGS.m

```

1 function [Q, R] = solveQRGS(A)
2     R = zeros(size(A,2),size(A,2));
3     Q = zeros(size(A));
4     for i = 1:size(A, 2)
5         v = A(:, i);
6         for j = 1:(i-1)
7             R(j,i) = Q(:, j)'*A(:, i);
8             v = v - R(j,i)* Q(:, j);
9         end
10        R(i, i) = norm(v);
11        Q(:, i) = v / R(i,i);
12    end
13
14 end

```

5.3. Método de Cholesky

cholesky.m

```
1 function chol = cholesky( matrix )
2
3
4 n = size(matrix,2);
5 chol = zeros(n);
6
7 for i=1:n
8     chol(i, i) = sqrt(matrix(i, i) - chol(i, :)*chol(i, :)' );
9     for j=(i + 1):n
10         chol(j, i) = (matrix(j, i) - chol(i,:)*chol(j, :)' )/chol(i, i
11         );
12     end
13 end
14 end
```

5.4. Cálculo de la velocidad estimada

getVelocity.m

```
1 function vel = getVelocity( vars )
2     fl = 1/365.25;
3     vel = [];
4     for i=1:size(vars,1)
5         a=vars(i,1);
6         b=vars(i,2);
7         c=vars(i,3);
8         vel(i) = a + b * cos(2*pi*fl*i)+ c * sin(2*pi*fl*i);
9     end
10     vel = vel';
11 end
```

5.5. Sustitución hacia adelante

fSubstitution.m

```
1 function x = fSubstitution(L,d)
2     n = size(d);
3     x=zeros(n);
4     for j=1:n
5         x(j)=d(j)/L(j,j);
6         d(j+1:n)=d(j+1:n)-L(j+1:n,j)*x(j);
7     end
8 end
```

Referencias

- [1] Manuel Palomares, *La meteorología y la predicción del tiempo ¿Cuál es actualmente su nivel de fiabilidad?, ¿por qué se equivocan los meteorólogos?* febrero 2008.