

CONFERENCIA



TALLER DE AVANCE  
***CLIMATOLOGÍA Y PRONÓSTICO DE VIENTOS EN  
LA MESÓSFERA Y BAJA TERMÓSFERA SOBRE LA  
COSTA CENTRAL Y NORTE DEL PERÚ***

CHRISTIAN MAURICIO T.

TESISTA DE POSTGRADO





# **PREDICTIBILIDAD DE LOS VIENTOS EN LA REGIÓN MLT**

## Sobre Jicamarca y Piura

# CONTENIDO

- Objetivo y descripción de investigación
- Datos
- Metodología
  - Estructura de modelo
  - Random Forest
  - Hiperparámetros
  - Elección de modelo
- Resultados
  - Comparación de serie de tiempo
- Conclusiones



# Objetivo y descripción de investigación

## **Objetivo:**

Construir modelo de predicción de vientos en la región MLT usando Machine learning.

## **Datos a usar:**

Vientos zonales, vientos meridionales, temperaturas, ratios de ozono e índices globales (kp, Dst, ap e f10.7).

## **Técnica propuesta:**

Random Forest para predicción.

## **Fuentes de datos:**

Radar SIMONE Jicamarca

Paginas web: GES DISC y OMNIWeb

## **Software a usar**

Python y R

# Datos

Todos los datos son tomados en un intervalo de 3 horas. Teniendo como etiqueta el valor medio del intervalo.

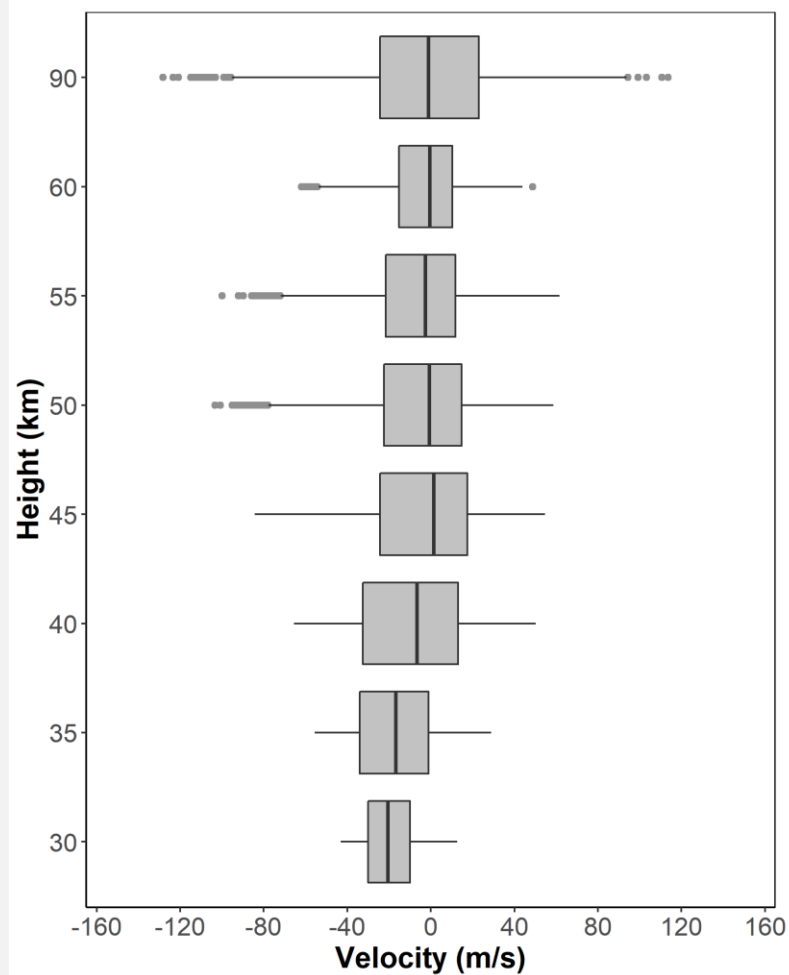
Intervalo	Etiqueta
[00:00, 03:00>	01:30
[03:00, 06:00>	04:30
[06:00, 09:00>	07:30
[09:00, 12:00>	10:30
[12:00, 15:00>	13:30
[15:00, 18:00>	16:30
[18:00, 21:00>	19:30
[21:00, 24:00>	22:30

Fecha Inicio	Fecha Fin
19/09/2019	30/06/2022

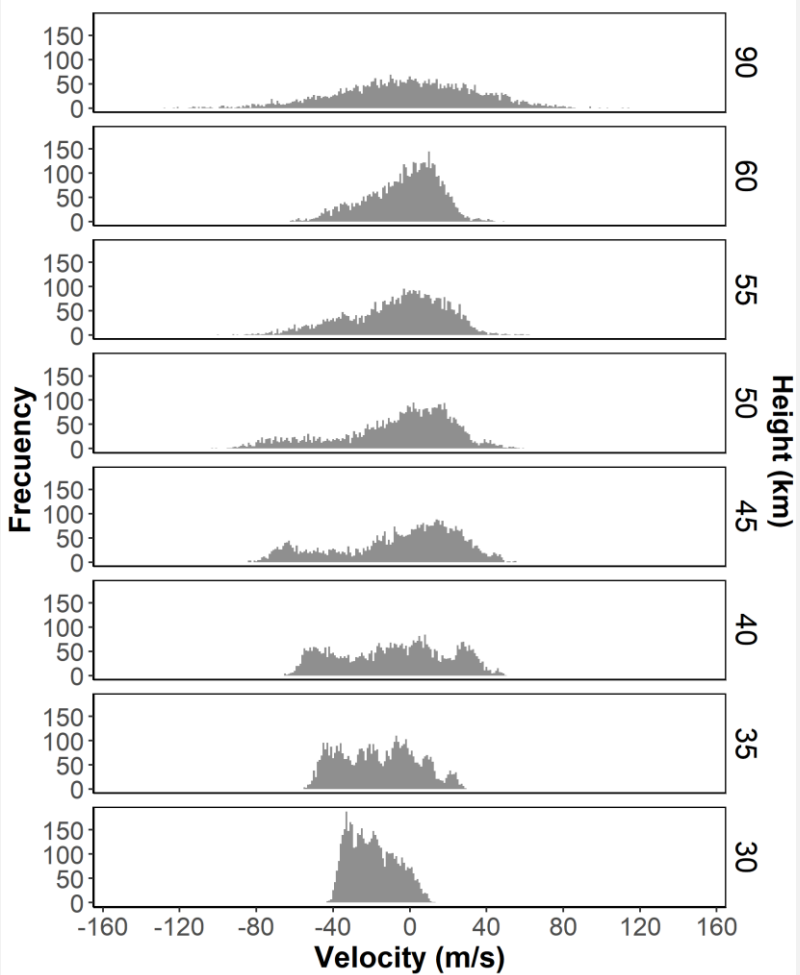
Día	Hora	X1	X2	...	Xn
19/09	01:30				
19/09	04:30				
19/09	07:30				
19/09	10:30				
19/09	13:30				
19/09	16:30				
19/09	19:30				
19/09	22:30				

# Datos

**Box plots of zonal winds  
over Jicamarca**



**Histograms of zonal winds  
over Jicamarca**

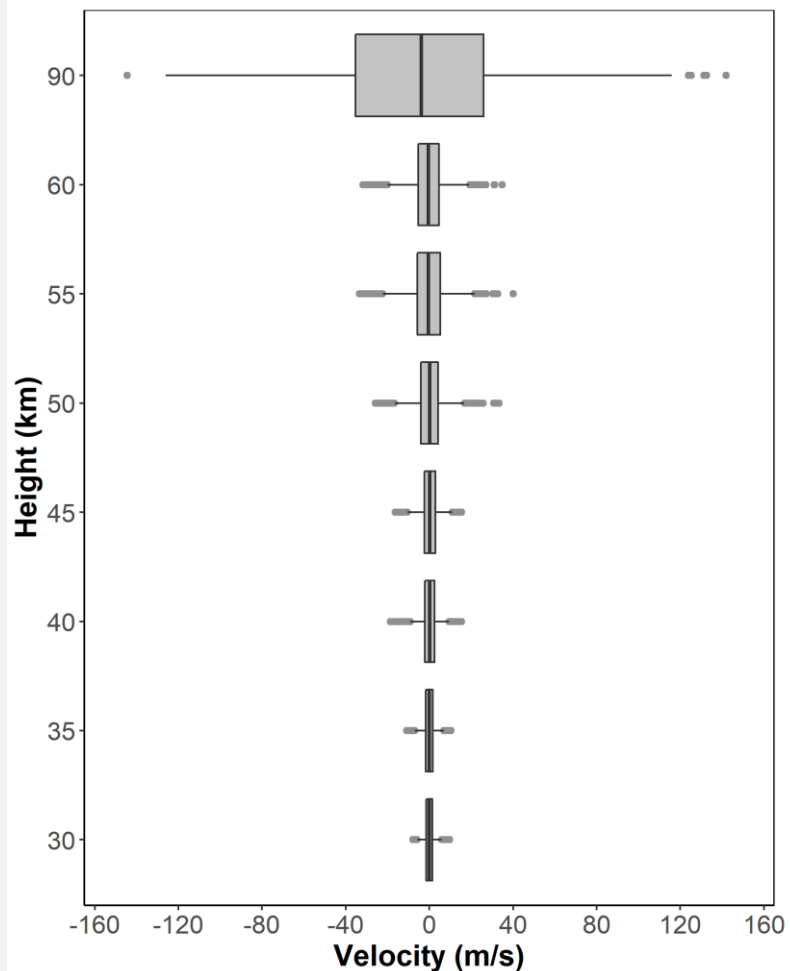


	Mean	Std.Dev	CV	Skewness	Kurtosis
u0_30	-19.35	12.30	-0.64	0.35	-0.87
u0_35	-16.25	19.96	-1.23	0.11	-0.97
u0_40	-8.34	28.18	-3.38	-0.09	-1.06
u0_45	-5.95	31.26	-5.26	-0.62	-0.58
u0_50	-7.39	30.26	-4.09	-0.79	0.00
u0_55	-6.46	25.10	-3.89	-0.55	-0.09
u0_60	-3.37	18.60	-5.52	-0.54	-0.16
u0_90	-1.92	35.03	-18.23	-0.22	-0.04

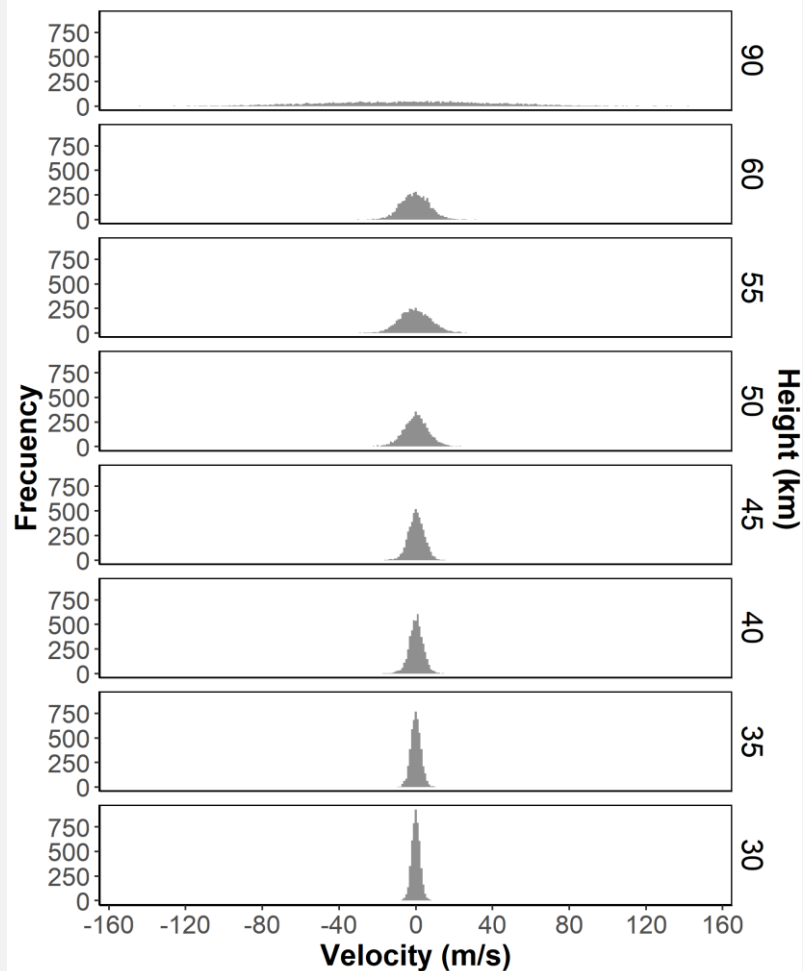
	Min	Q1	Median	Q3	Max
u0_30	-43.09	-29.97	-20.61	-9.87	12.62
u0_35	-55.50	-34.09	-16.77	-1.14	28.97
u0_40	-65.42	-32.64	-6.51	13.10	50.34
u0_45	-84.25	-24.36	1.46	17.58	54.71
u0_50	-103.41	-22.34	-0.63	14.67	58.75
u0_55	-99.92	-21.67	-2.62	11.87	61.71
u0_60	-61.89	-15.31	-0.41	10.33	48.82
u0_90	-128.23	-24.42	-1.05	23.02	113.61

# Datos

Box plots of meridional winds over Jicamarca



Histograms of meridional winds over Jicamarca

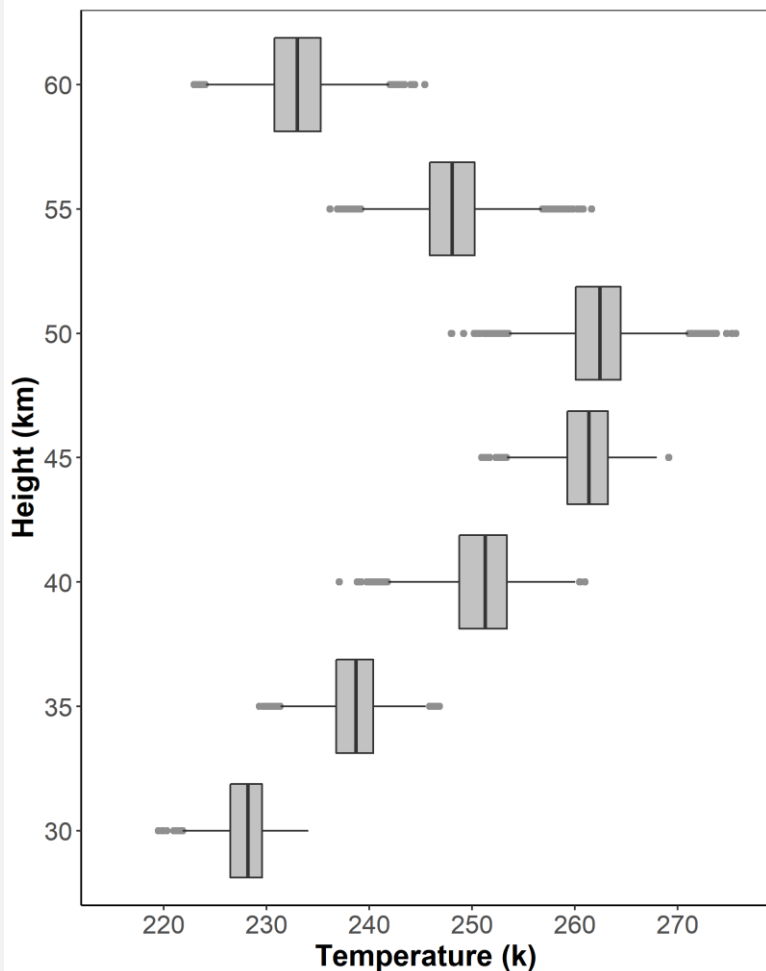


	Mean	Std.Dev	CV	Skewness	Kurtosis
v0_30	0.03	2.21	80.75	0.05	0.33
v0_35	0.02	2.65	114.96	0.06	0.37
v0_40	0.20	3.76	18.44	-0.20	1.08
v0_45	0.26	4.12	15.98	-0.13	0.35
v0_50	0.06	6.69	105.16	0.01	0.89
v0_55	-0.29	8.40	-28.91	0.10	0.48
v0_60	-0.42	7.42	-17.73	0.03	0.51
v0_90	-4.51	42.74	-9.47	0.00	-0.43

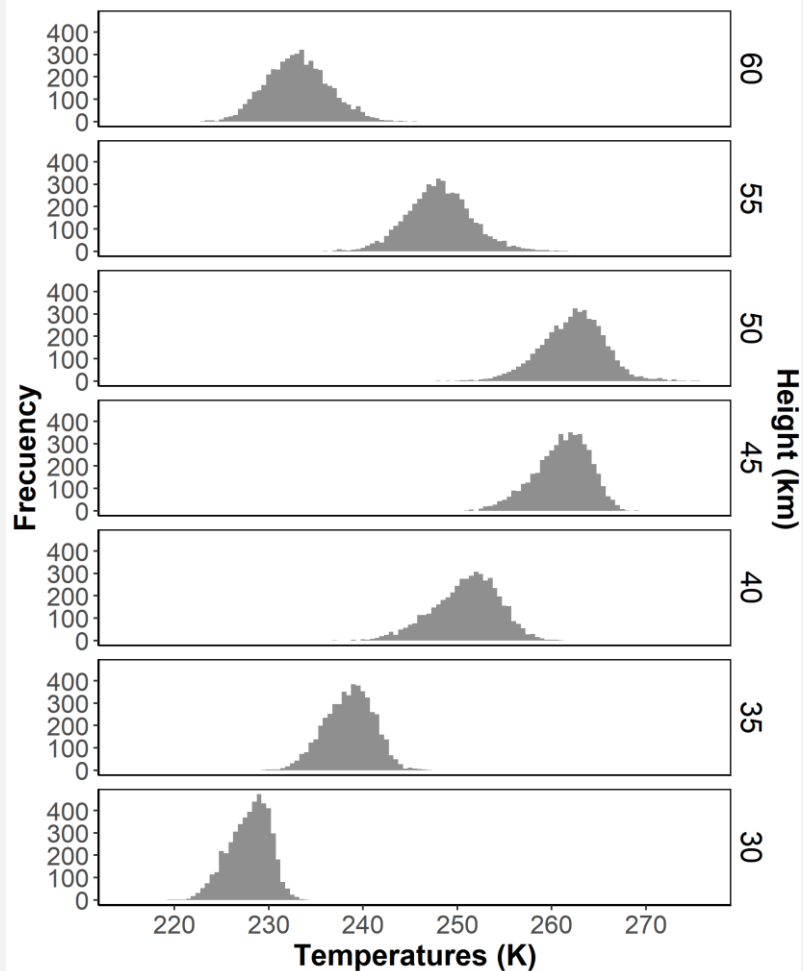
	Min	Q1	Median	Q3	Max
v0_30	-7.83	-1.46	0.03	1.48	9.66
v0_35	-10.86	-1.76	-0.03	1.74	10.29
v0_40	-18.72	-2.09	0.25	2.55	15.36
v0_45	-16.27	-2.42	0.25	2.95	15.30
v0_50	-25.81	-4.12	0.11	4.22	33.43
v0_55	-33.32	-5.84	-0.49	5.16	40.09
v0_60	-31.79	-5.33	-0.46	4.51	34.72
v0_90	-144.48	-35.30	-3.90	25.95	142.02

# Datos

Box plots of temperatures over Jicamarca



Histograms of temperatures over Jicamarca

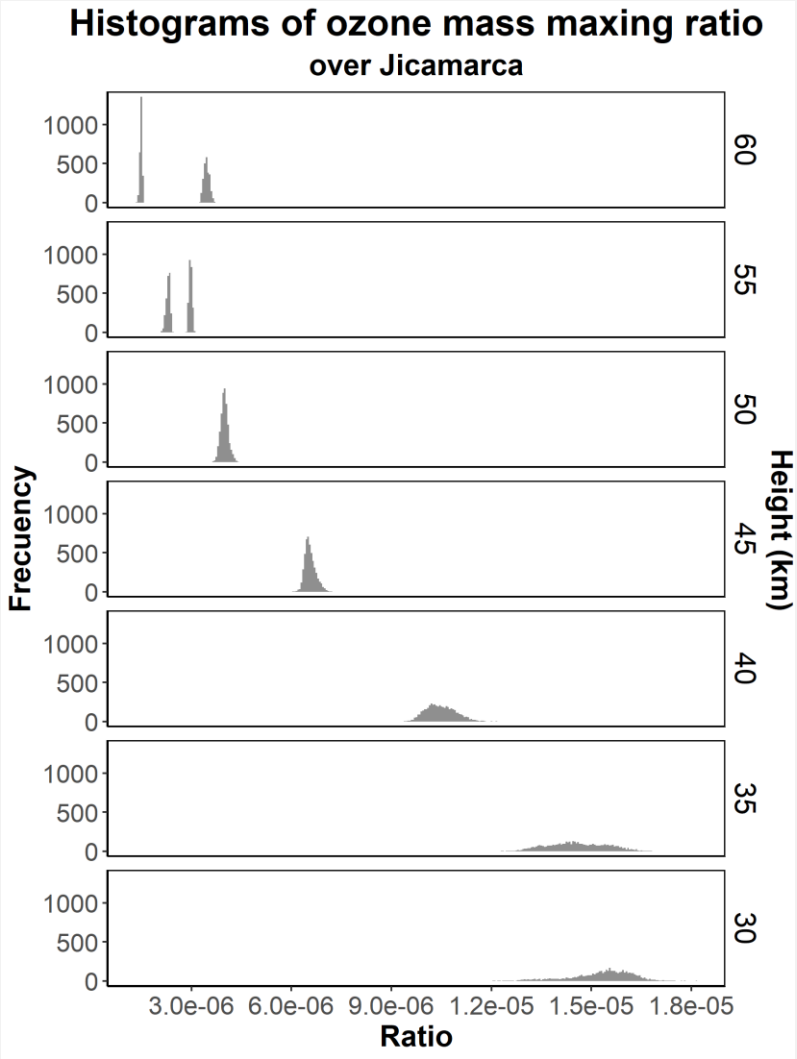
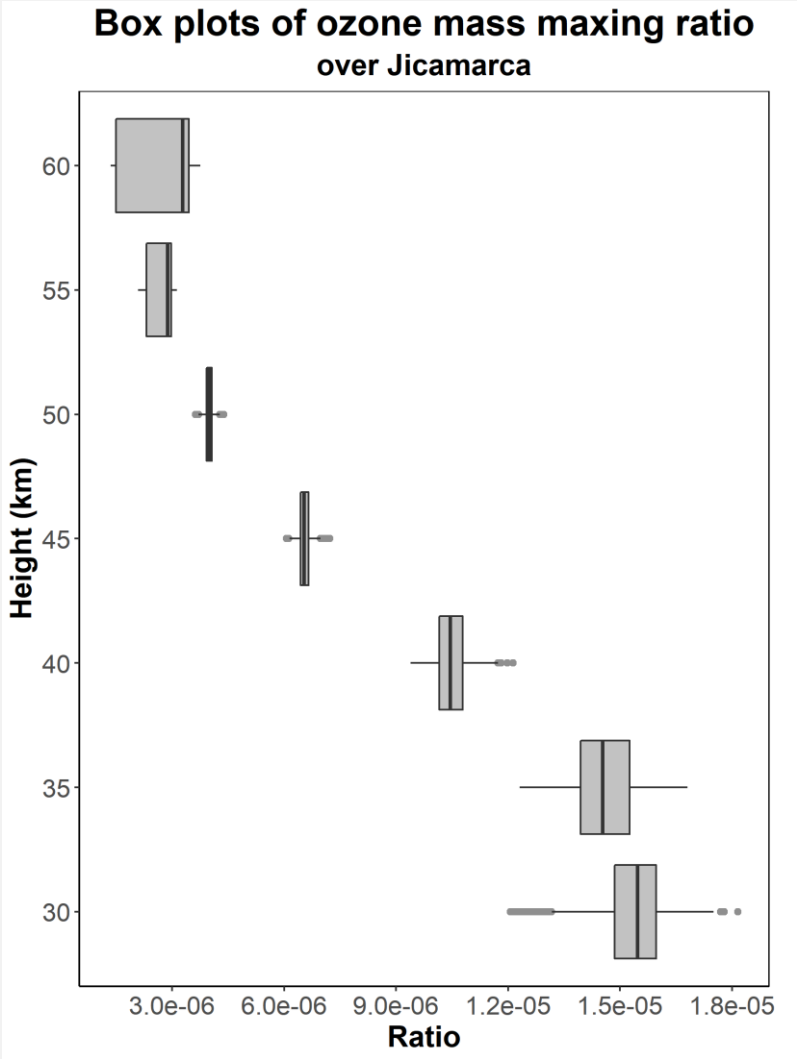


	Mean	Std.Dev	CV	Skewness	Kurtosis
temp_30	227.92	2.16	0.01	-0.41	-0.23
temp_35	238.54	2.57	0.01	-0.19	-0.16
temp_40	250.96	3.45	0.01	-0.40	0.00
temp_45	261.10	2.88	0.01	-0.47	-0.10
temp_50	262.26	3.37	0.01	-0.18	0.53
temp_55	248.11	3.46	0.01	0.15	0.48
temp_60	233.09	3.30	0.01	0.15	-0.07

	Min	Q1	Median	Q3	Max
temp_30	219.47	226.46	228.19	229.54	234.06
temp_35	229.28	236.77	238.71	240.39	246.81
temp_40	237.06	248.76	251.30	253.40	261.00
temp_45	250.94	259.27	261.38	263.22	269.17
temp_50	247.99	260.11	262.46	264.49	275.67
temp_55	236.18	245.88	248.05	250.26	261.63
temp_60	222.96	230.77	233.00	235.25	245.42



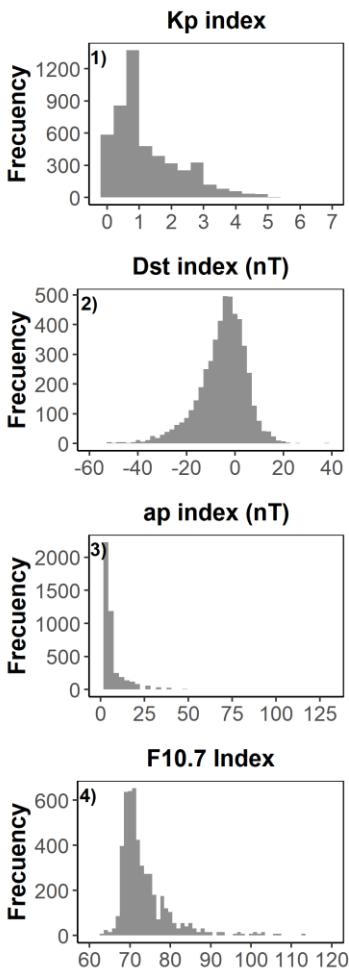
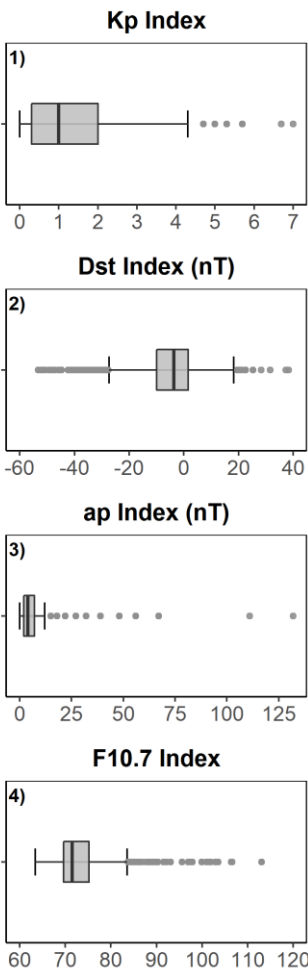
# Datos



	Mean	Std.Dev	CV	Skewness	Kurtosis
o3_30	15.3E-6	900.6E-9	0.06	-0.76	0.41
o3_35	14.6E-6	859.5E-9	0.06	0.05	-0.75
o3_40	10.5E-6	435.4E-9	0.04	0.30	-0.37
o3_45	6.6E-6	163.5E-9	0.02	0.66	0.42
o3_50	4.0E-6	112.1E-9	0.03	0.27	0.25
o3_55	2.6E-6	338.7E-9	0.13	-0.03	-1.90
o3_60	2.5E-6	985.4E-9	0.40	0.00	-1.98

	Min	Q1	Median	Q3	Max
o3_30	12.1E-6	14.8E-6	15.5E-6	16.0E-6	18.2E-6
o3_35	12.3E-6	13.9E-6	14.5E-6	15.3E-6	16.8E-6
o3_40	9.4E-6	10.2E-6	10.5E-6	10.8E-6	12.1E-6
o3_45	6.1E-6	6.4E-6	6.5E-6	6.7E-6	7.2E-6
o3_50	3.6E-6	3.9E-6	4.0E-6	4.1E-6	4.4E-6
o3_55	2.1E-6	2.3E-6	2.9E-6	3.0E-6	3.1E-6
o3_60	1.4E-6	1.5E-6	3.3E-6	3.4E-6	3.8E-6

# Datos



	Mean	Std.Dev	CV	Skewness	Kurtosis
ap_index_nT	5.84	7.04	1.20	4.46	41.74
Dst.index_nT	-4.95	9.99	-2.02	-0.85	1.90
f10.7_index	73.51	6.44	0.09	2.43	7.89
Kp_index	1.25	1.07	0.86	1.09	0.93

	Min	Q1	Median	Q3	Max
ap_index_nT	0.00	2.00	4.00	7.00	132.00
Dst.index_nT	-53.33	-10.00	-3.67	1.67	38.33
f10.7_index	63.40	69.60	71.50	75.20	113.10
Kp_index	0.00	0.30	1.00	2.00	7.00

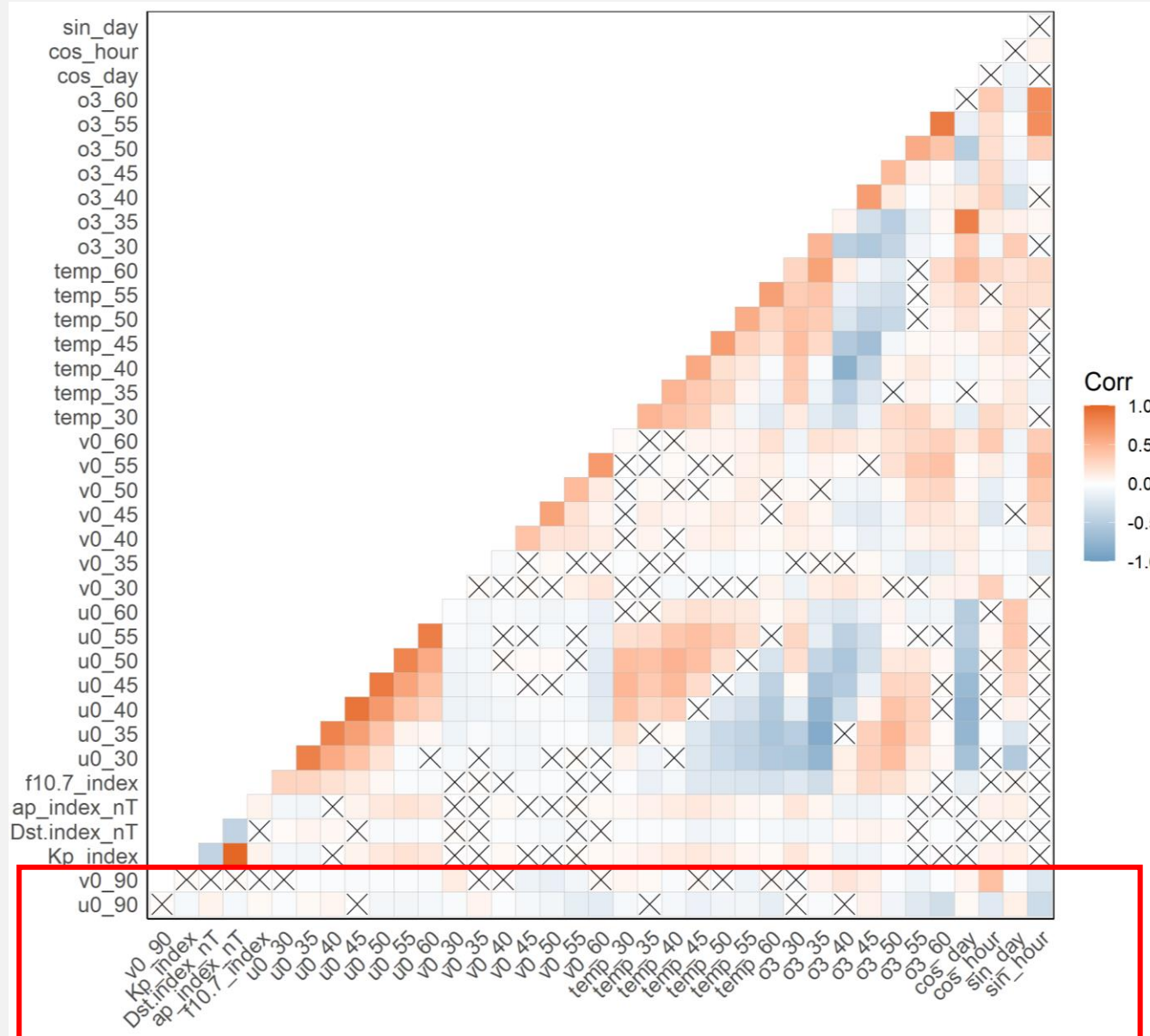
# Datos

Variables dependientes	Variables independientes	
<b>SIMONe Jicamarca</b>	<b>GES DISC</b>	<b>OMNIWeb</b>
Vientos zonales (u) (Entre 75 y 105 km de altura)	Vientos zonales (u) (Entre 30 y 60 km de altura)	Índice Kp
Vientos meridionales (v) (Entre 75 y 105 km de altura)	vientos meridionales (v) (Entre 30 y 60 km de altura)	Índice Dst
	Temperaturas (temp) (Entre 30 y 60 km de altura)	Índice ap
	Ratio máximo de ozono (o3) (Entre 30 y 60 km de altura)	Índice solar f10.7
		<b>Variables de tiempo</b>
		Cos_hour
		Sin_hour
		Cos day
		Sin day



Instituto  
Geofísico  
del Perú

## Datos Correlaciones



	u0_90
sin_hour	-0.35320596
o3_60	-0.33089478
o3_55	-0.22318541
cos_hour	-0.20887724

	v0_90
cos_hour	0.41228043
sin_hour	-0.23135751
o3_40	0.16618338
v0_50	-0.15430684

# Metodología

## Estructura de modelo

19/09/2019

30/06/2020

01/07/2020

30/06/2022

Datos de entrenamiento y testeo

75 %

25 %

Comparación de 1 año de datos  
Entre predicción y datos reales

### Variables de tiempo

- Cosenos y senos de días del año (cos\_day, sin\_day)
- Cosenos y senos de horas del día (cos\_hour, sin\_hour)

### OMNIWeb

- Índice geomagnético planetario (Kp\_index)
- Índice geomagnético Dst (Disturbance Storm Time) (Dst-index nT)
- Índice ap (ap\_index\_nT)
- Índice f10.7 (f10.7\_index)

### GES DISC

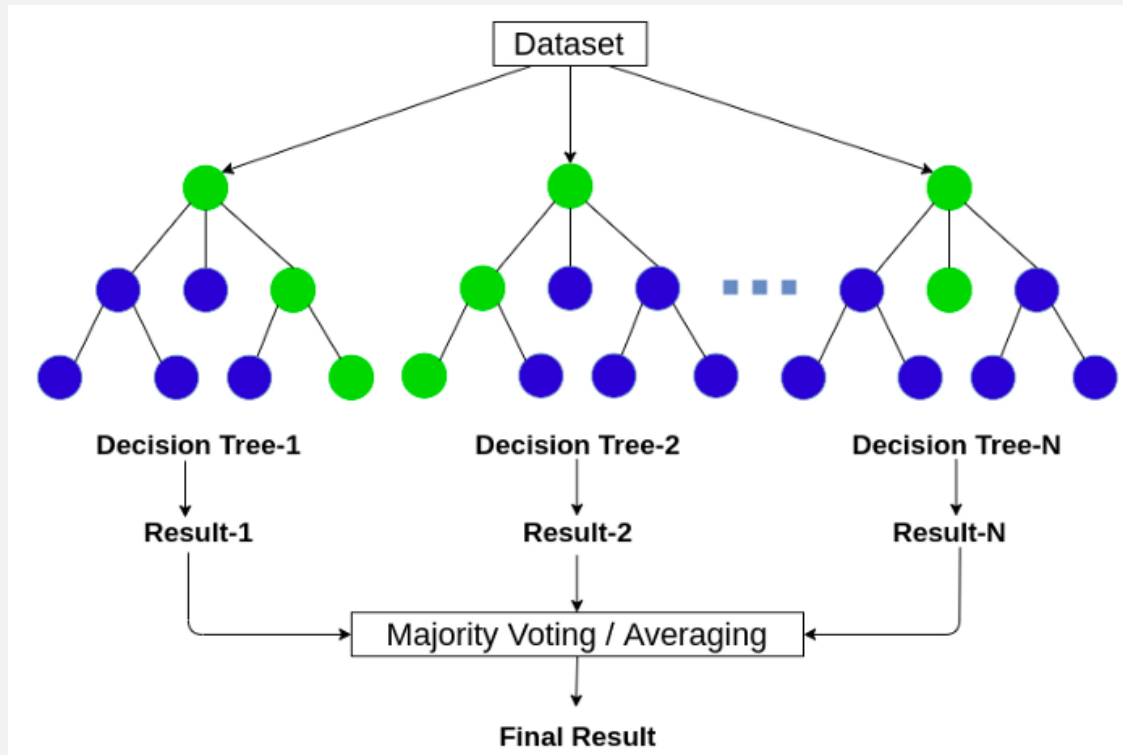
- Vientos zonales
- Vientos meridionales
- Temperaturas
- Ratio máximo de masa de ozono

Los datos están interpolados a alturas de 30, 35, 40, 45, 50, 55 y 60 km

Modelo con Random Forest

Predicciones de Vientos a 90 km de altura

# Metodología Random Forest



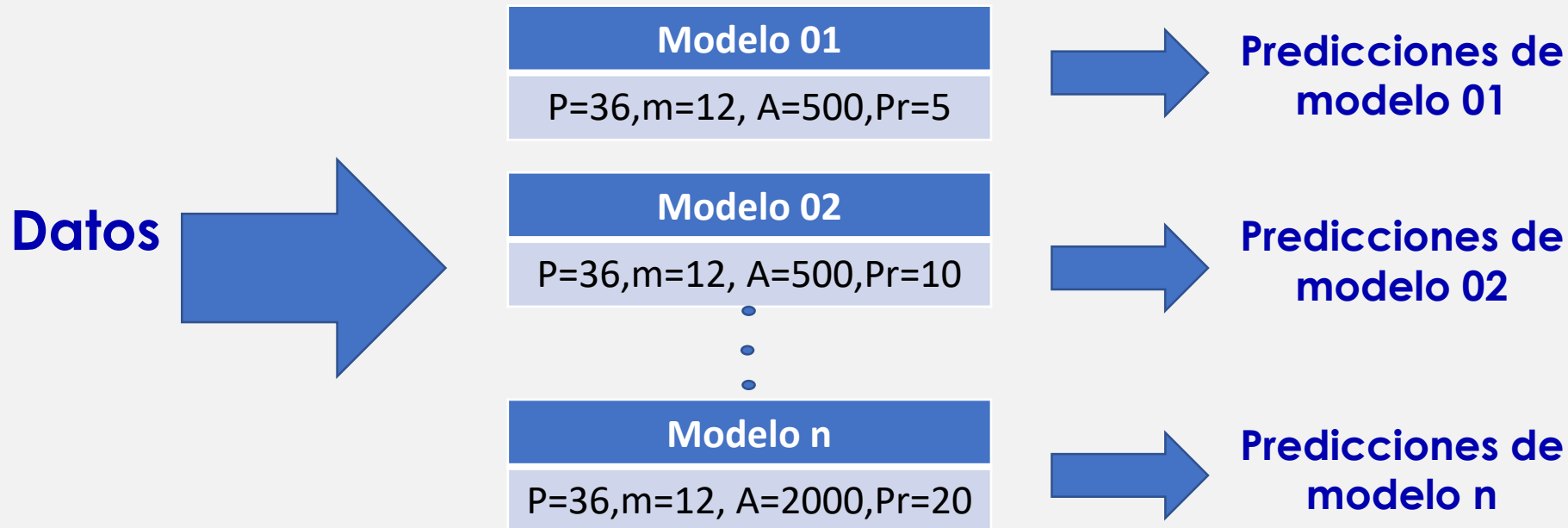
## Hiperparámetros

- P predictores: cantidad total de predictores.
- Cantidad máxima de m predictores aleatorios: Para el caso de regresión se recomienda que  $m = p/3$ .
- Número de árboles: Cantidad aleatoria que la define el investigador.
- Profundidad del árbol: Es la cantidad de niveles que tendrá el árbol. Este es un parámetro crítico porque cuanto más niveles tiene el árbol, hay menos error en la predicción. Esto puede generar un sobre ajuste del modelo.

Cada árbol arroja una predicción. Pero el resultado final es el promedio del conjunto de resultados que da cada árbol de manera individual.

# Metodología Hiperparámetros

- N° de predictores:  $P=36$
- m predictores aleatorios:  $m = 36/3 = 12$
- Número de árboles (A): 500, 1000, 1500 y 2000
- Profundidades del árbol (Pr): 5, 10, 15 y 20



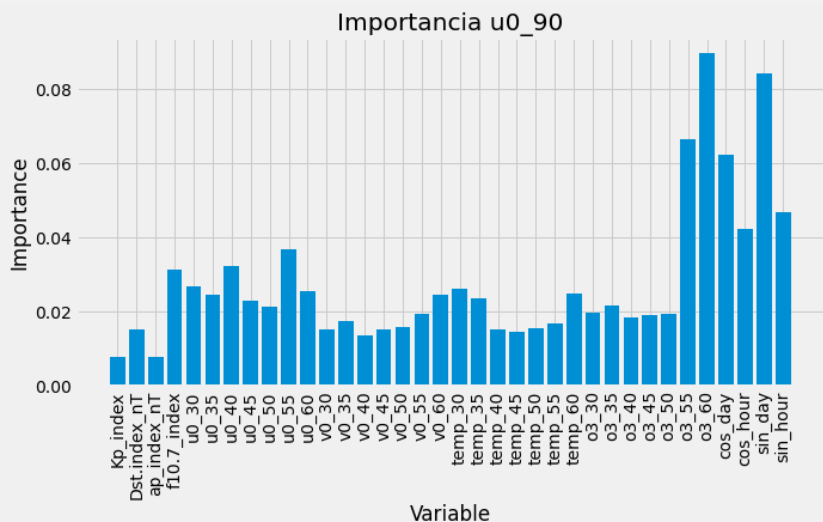
- N° de Modelos: 16 ( para estimar el viento zonal y meridional a 90 km de altura

# Metodología

## Elección de modelo

### Viento zonal a 90 Km de altura

Modelo
N° predictores = 36
N° árboles = 2000
N° máx predictores = 12
N° de niveles del árbol = 10



Training (75%)
R2 = 0.412
RMSE = 18.452
Testing (25%)
R2 = 0.401
RMSE = 27.075



Comparación con Nuevos datos (1 año)
Correlación = 0.440
RMSE = 31.845

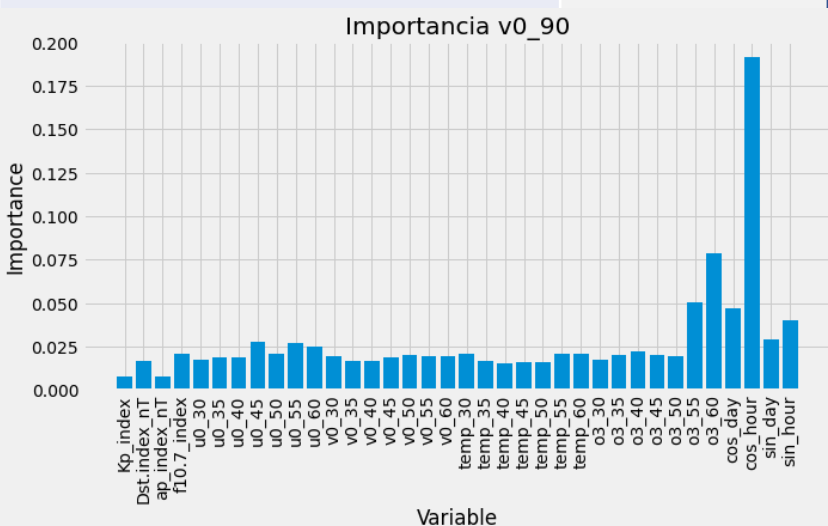


# Metodología

## Elección de modelo

### Viento meridional a 90 Km de altura

Modelo
N° predictores = 36
N° árboles = 2000
N° máx predictores = 12
N° de niveles del árbol = 10

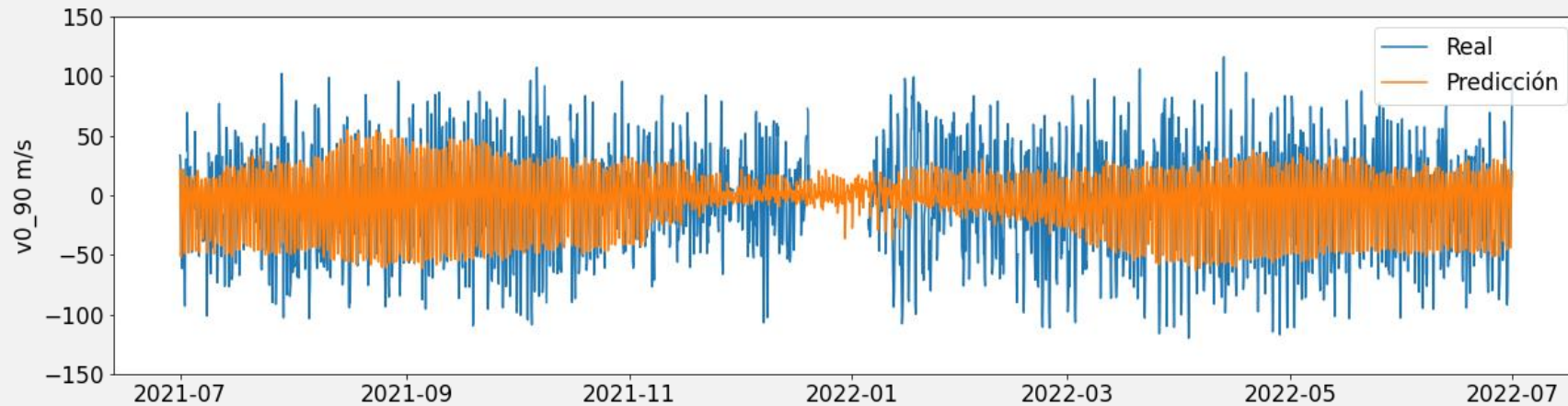
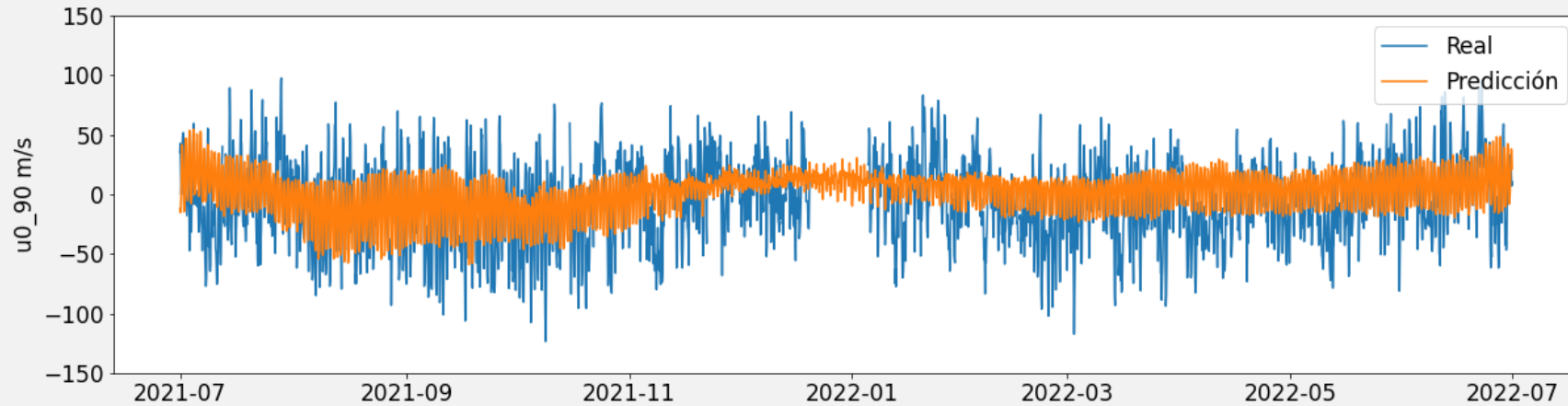


Training (75%)
R2 = 0.413
RMSE = 21.434
Testing (25%)
R2 = 0.401
RMSE = 32.034

Comparación con Nuevos datos (1 año)
Correlación = 0.535
RMSE = 35.939

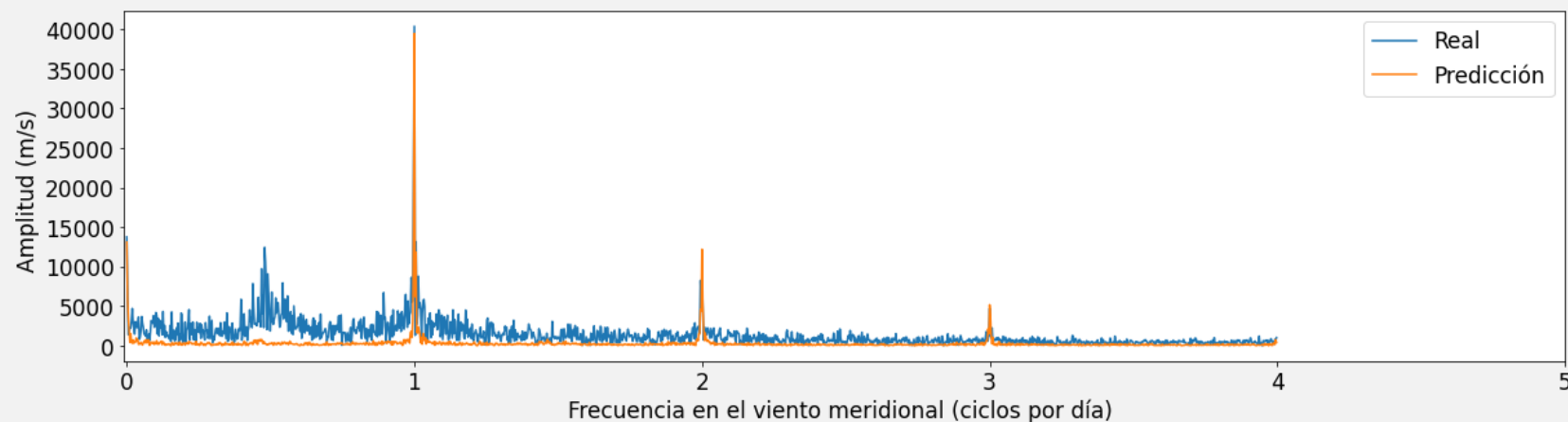
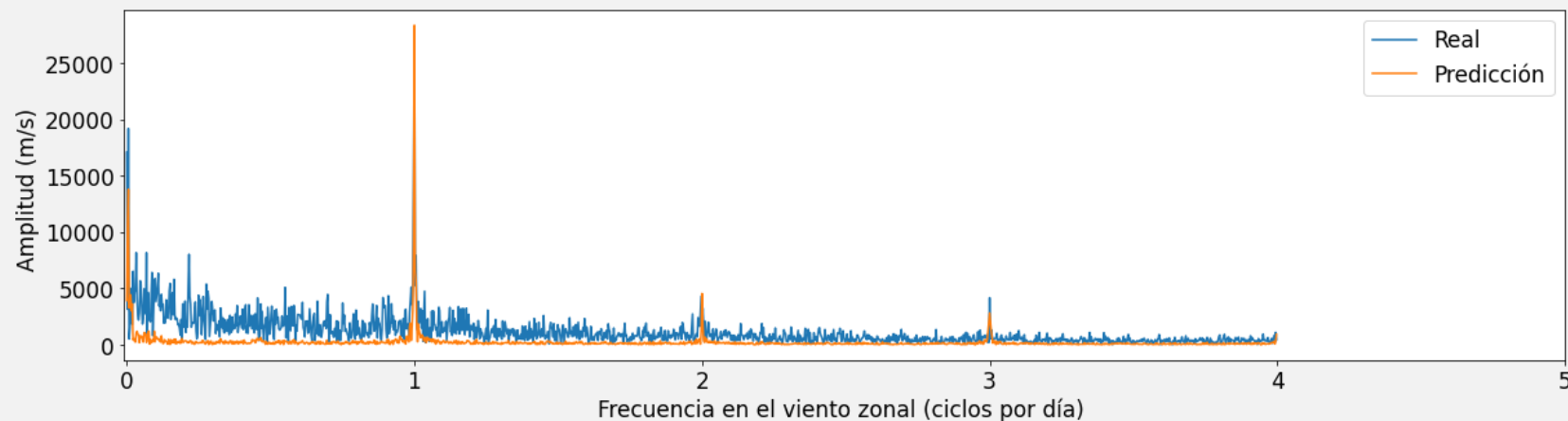
# Resultados

## Comparaciones de series de tiempo



# Resultados

## Comparaciones de series de tiempo



## Conclusiones y trabajo a Futuro

- Existen variables que no aportan información al modelo, como es el caso de los índices globales.
- El espectro muestra que las predicciones no reflejan la presencia de períodos de 2 días (0.5 ciclos por días).
- Las métricas indican que el modelo tiene una correlación moderada. Sin embargo las predicciones no siguen la variabilidad de los vientos, especialmente en la estación de verano.
- Como trabajo a futuro se pretende añadir variables que estén relacionadas a la atmósfera media. De igual manera se pretende aplicar redes neuronales para predicciones a corto plazo.

**¡Gracias por su atención!**