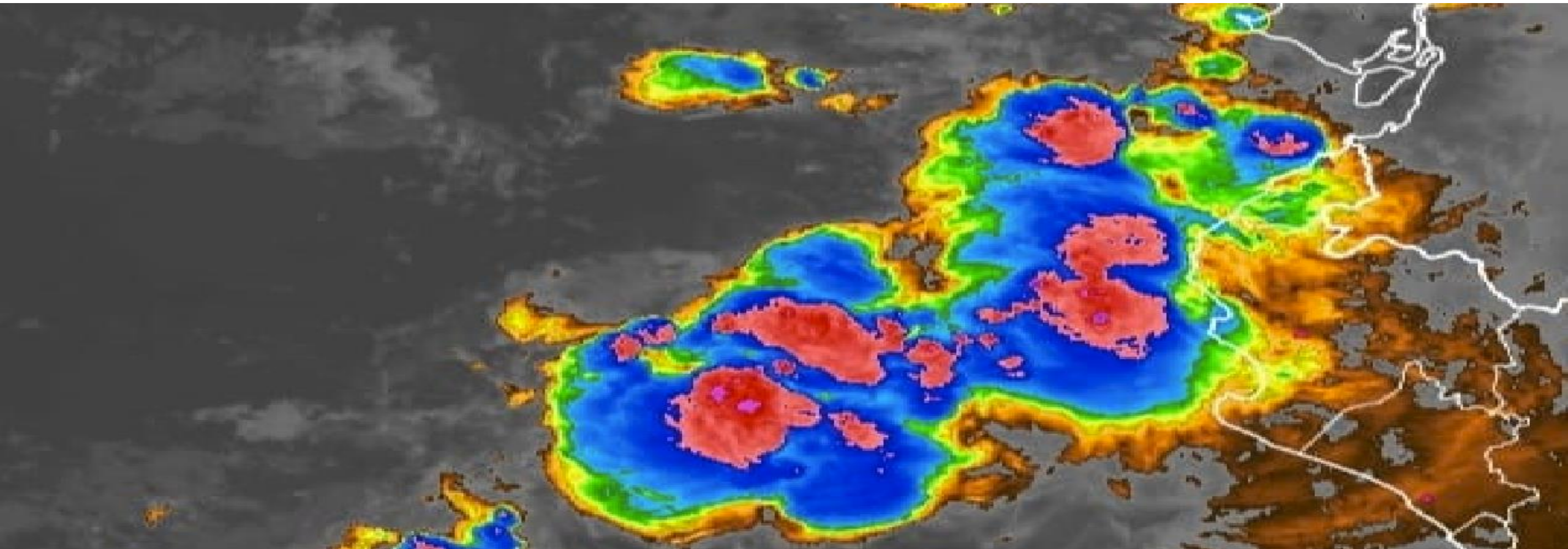


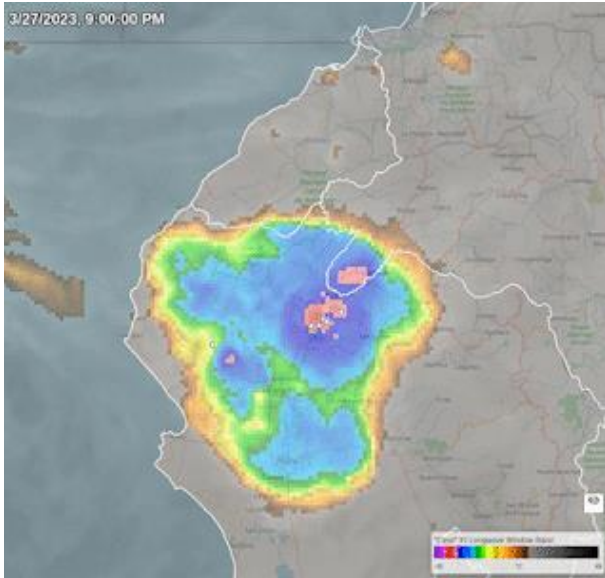
Índice de pronóstico de precipitaciones para la costa norte del Perú realizado con regresión logística.



M.S. Piero Rodrigo Rivas Quispe

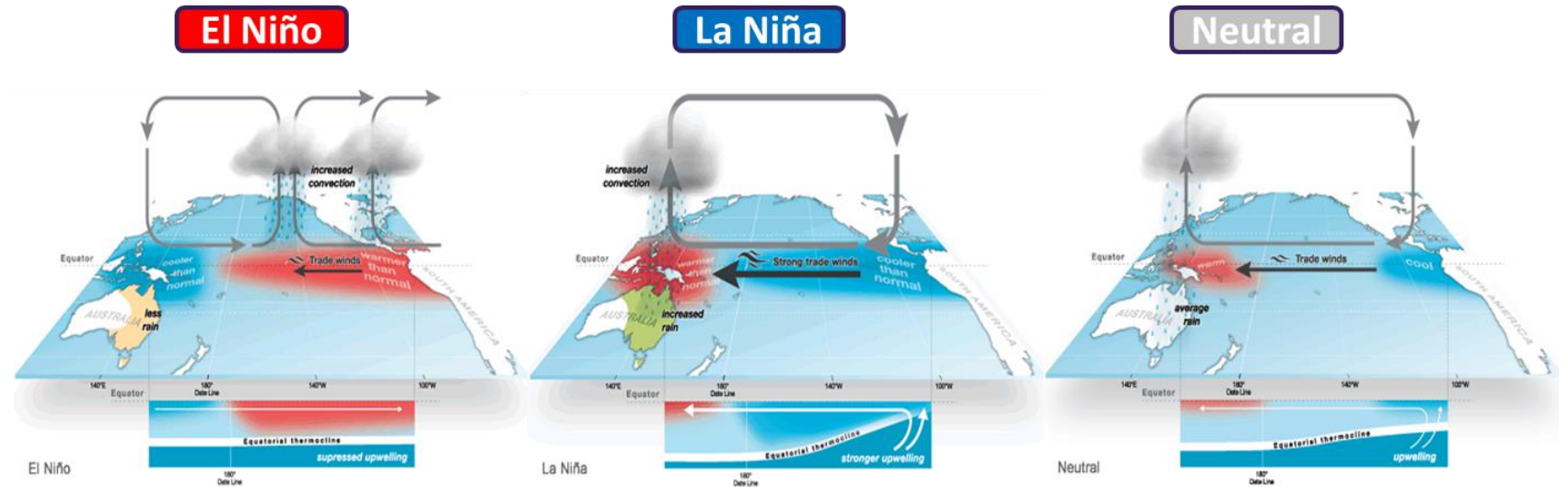


Tormentas en la costa norte del Peru

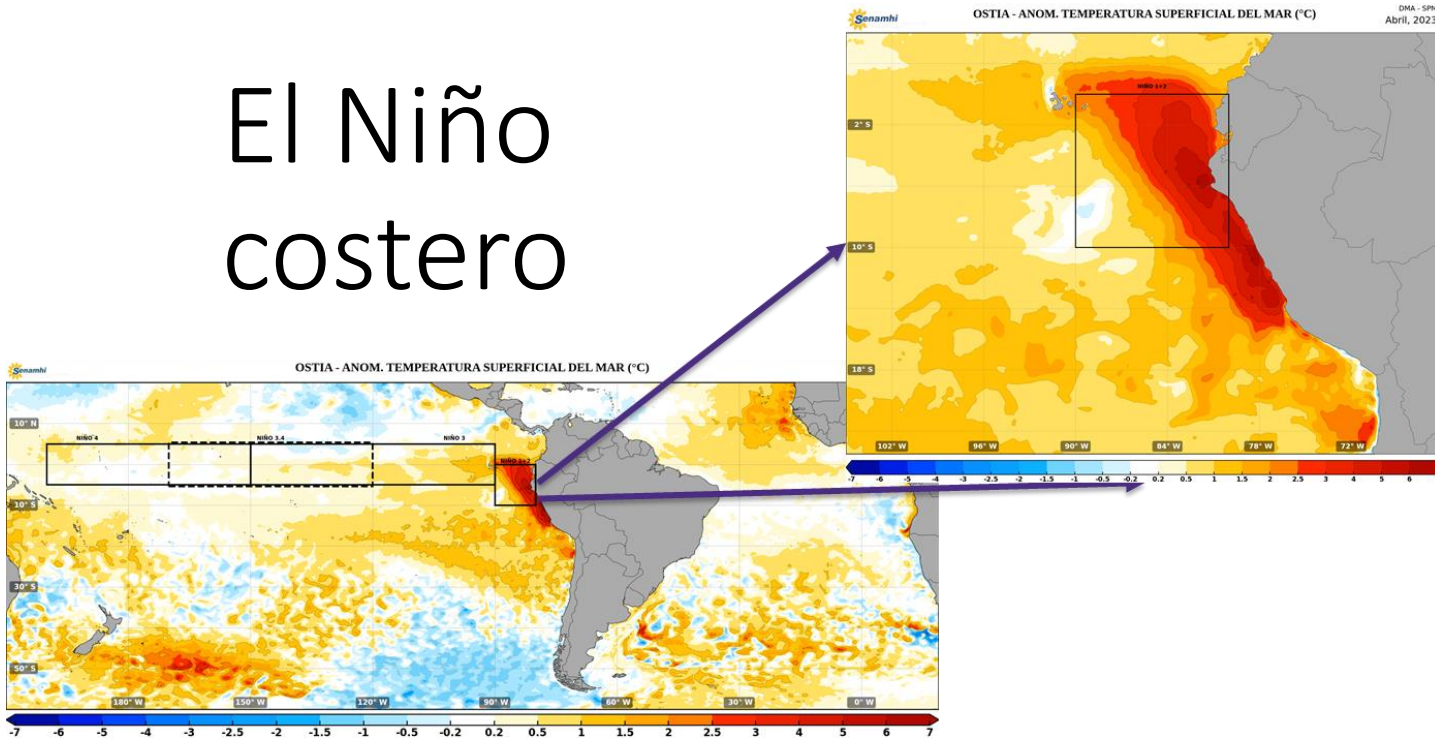


- La costa peruana es conocida por su clima desértico. Sin embargo, la llegada de El Niño lo transforma en una región tormentosa tropical.
- Estas tormentas pueden generar precipitaciones por encima de 150 mm en 24 horas.
- Eventos el Niño más Fuertes fueron en el 82-83, 97-98 y 2017

ENSO



El Niño costero



Patrones de circulación asociados a tormentas en la costa norte

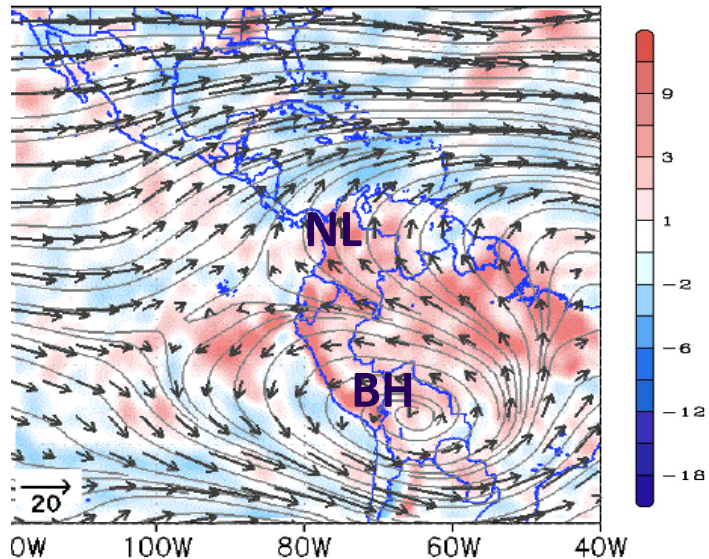
- El aumento de la precipitación en la Costa Norte del Perú está relacionado con los siguientes patrones de circulación (Quispe, 2018; y Aliaga et al, 2022) en niveles alto, medio y bajo:

Alta

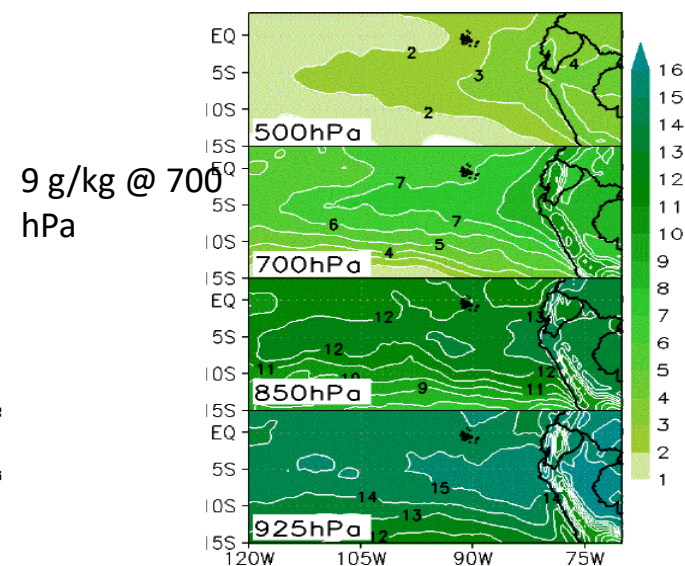
Media

Baja

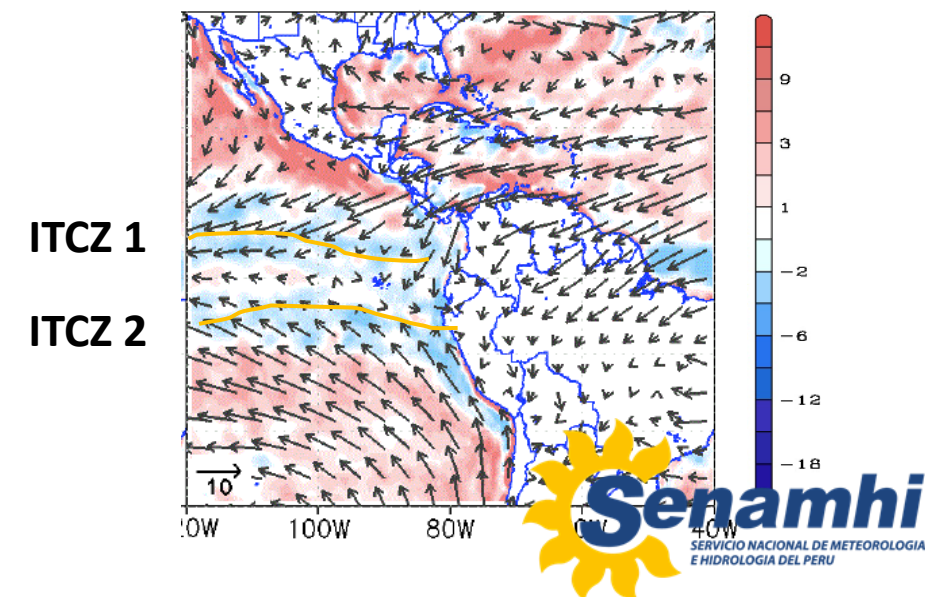
Lineas de corriente y divergencia en 250 hPa



Relacion de Mezcla



Vectores y divergencia en 950 hPa

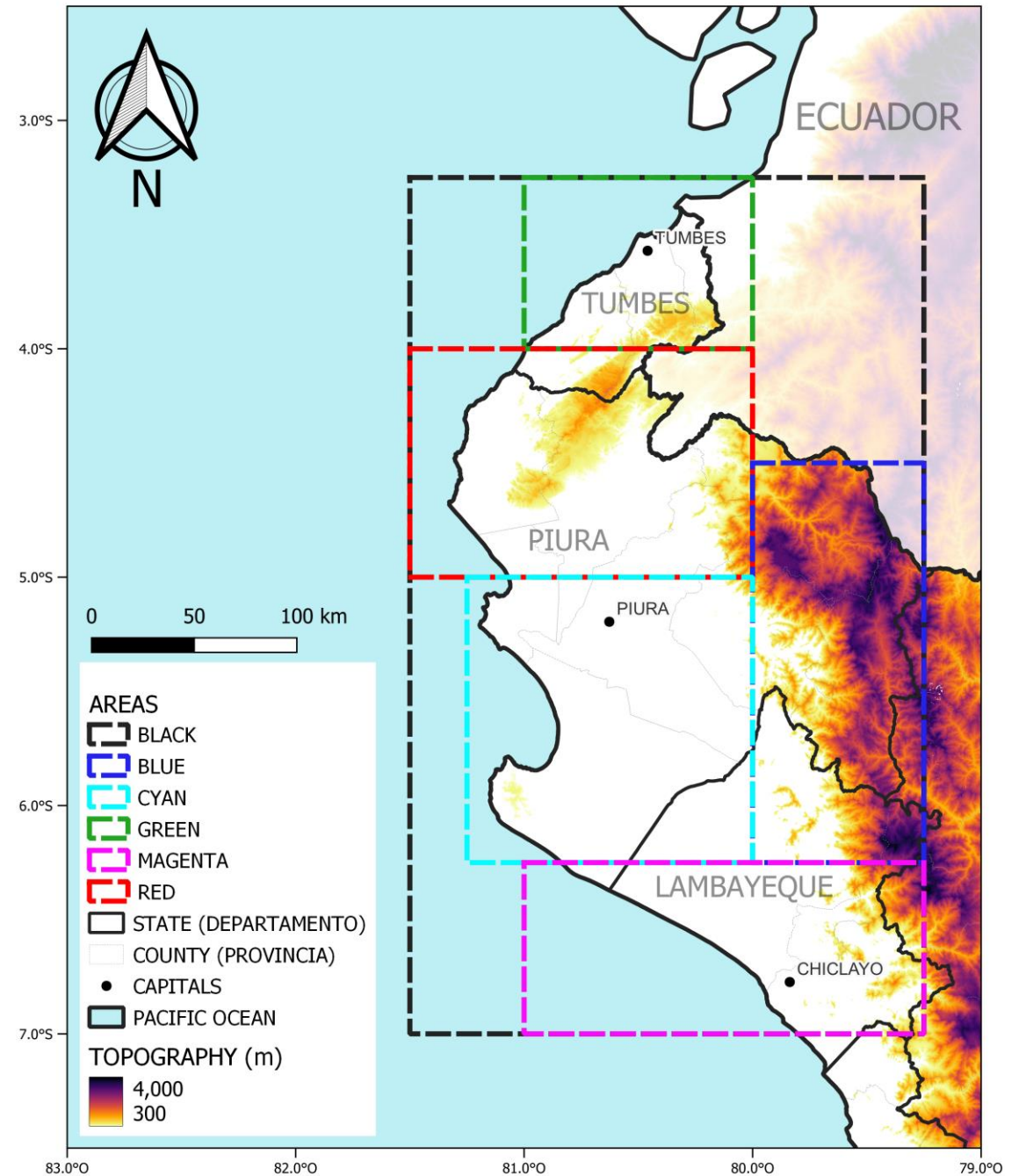


Objetivos

- Documentar la distribución e intensidad de las precipitaciones en la Costa Norte del Perú utilizando estaciones meteorológicas y datos satelitales.
- Desarrollar un índice para predecir la probabilidad de precipitación en esta región utilizando variables meteorológicas comúnmente disponibles, datos de precipitación satelitales y reanálisis con regresión logística.
- Pruebe la validez del índice utilizando datos que no se utilizaron para desarrollarlo.

Área de estudio

Región	Latitud (S°)	Longitud (W°)	Color
Área total	-3.25; -7	-81.5; -79.5	Negro
Andes de la costa norte	-4.5; -6.25	-80; -79.25	Azul
Desierto de Sechura	-5; -6.25	-81.25; -80	Celeste
Tumbes	-3.25; -4	-81; -80	Verde
Lambayeque	-6.25; -7	-81; -79.25	Fucsia
Cerros de Amotape	-4; -5	-81.5; -80	Rojo

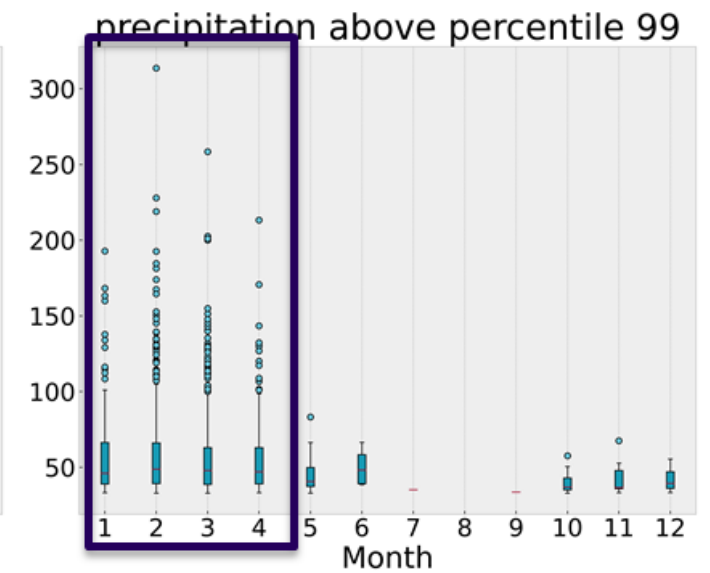
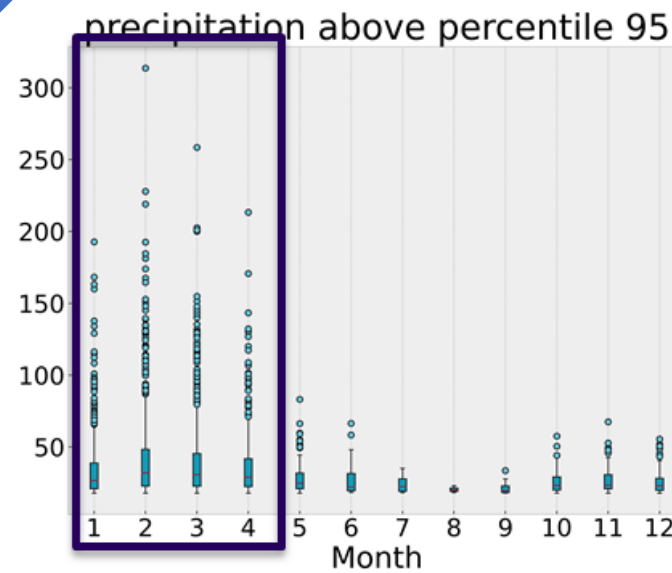
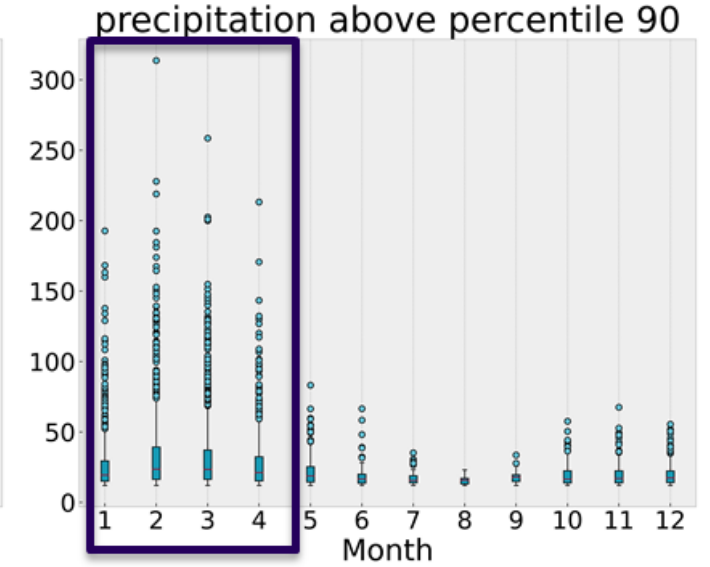
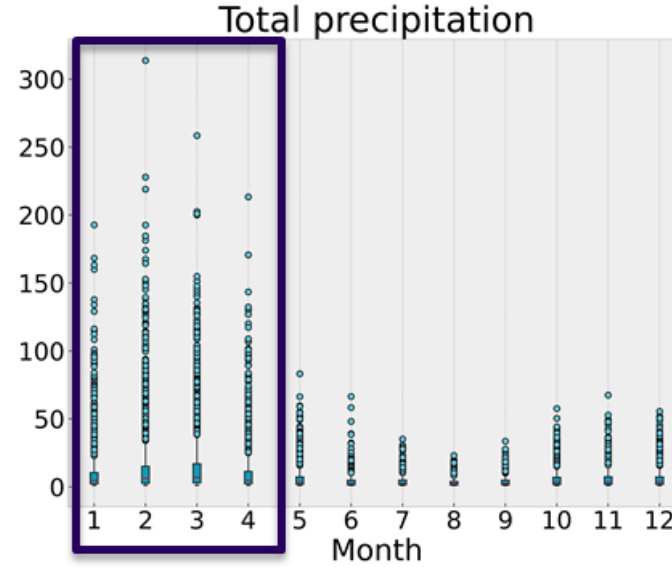
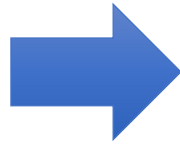
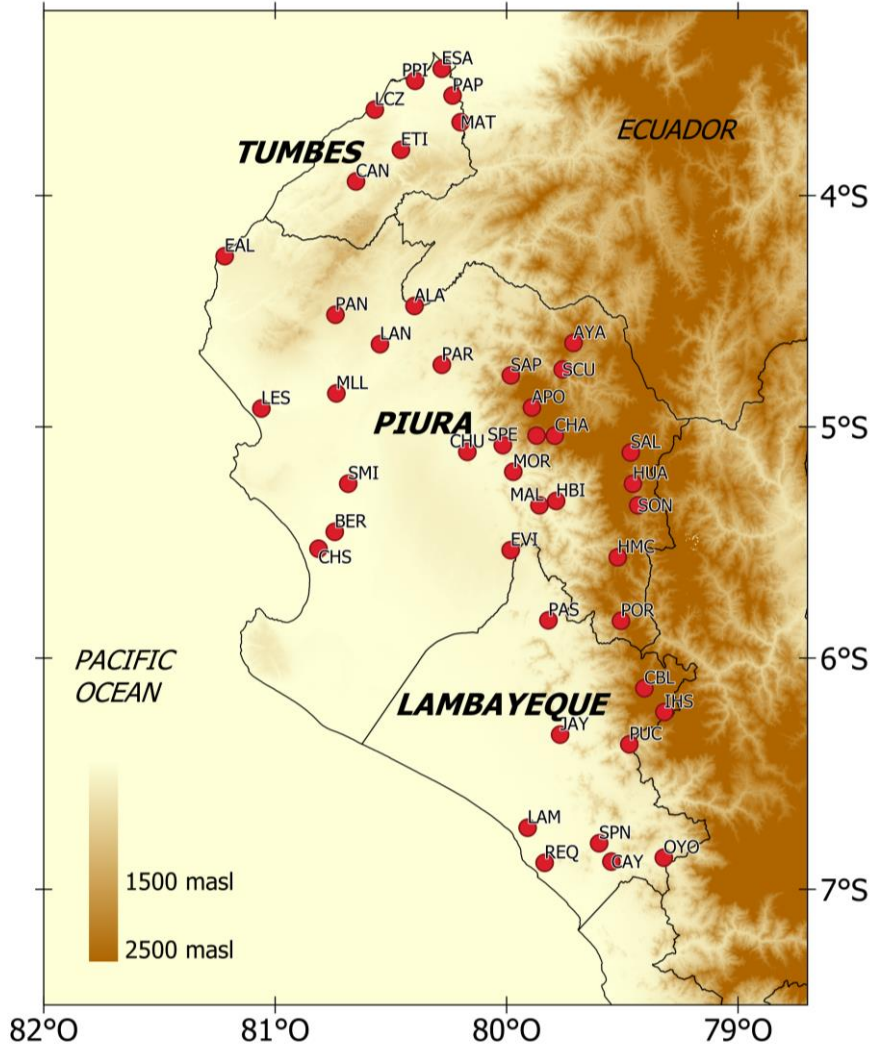


Datos & Metodología

Dataset	Region	Periodo	Resolución espacial	Resolución temporal	fuelle
Estaciones	Costa norte Peru	1998 - 2021	40 estaciones	diario	SENAMHI
Satélite (TRMM &GPM)	Vo6 SAM (Sur America)	1998 – 2013 (TRMM) 2014 – 2021 (GPM)	0.05° (5.5 km)	1 to 2 barridos por día	Universidad de Washington
ONI & ICEN	3.4 and 1+2 region	1998 - 2021	Niño 3.4 y 1+2 región	Mensual (promedio trimestral)	NOAA, IGP
ERA5	Sur América	1998 - 2021	0.25° (27.5 km)	03h	ECMWF

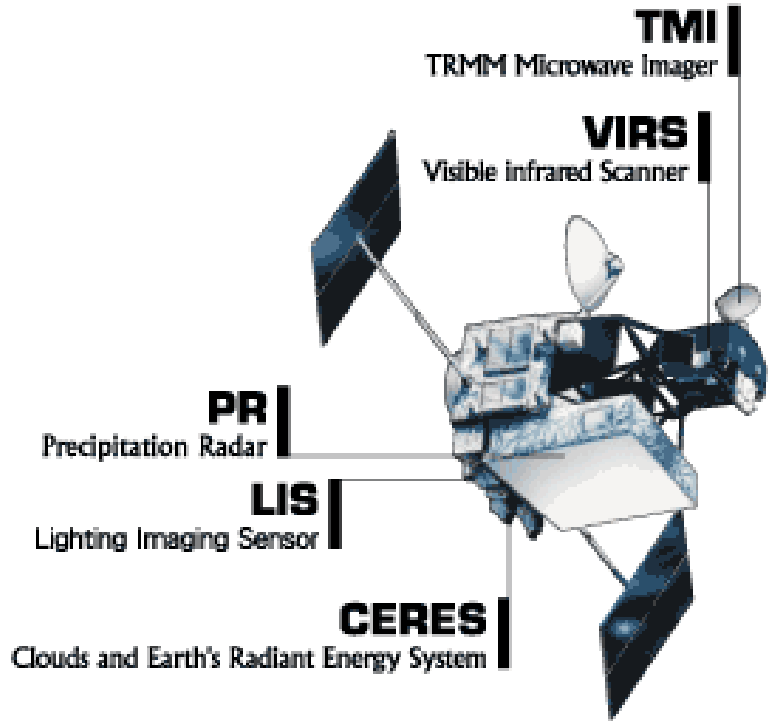


Datos de estaciones

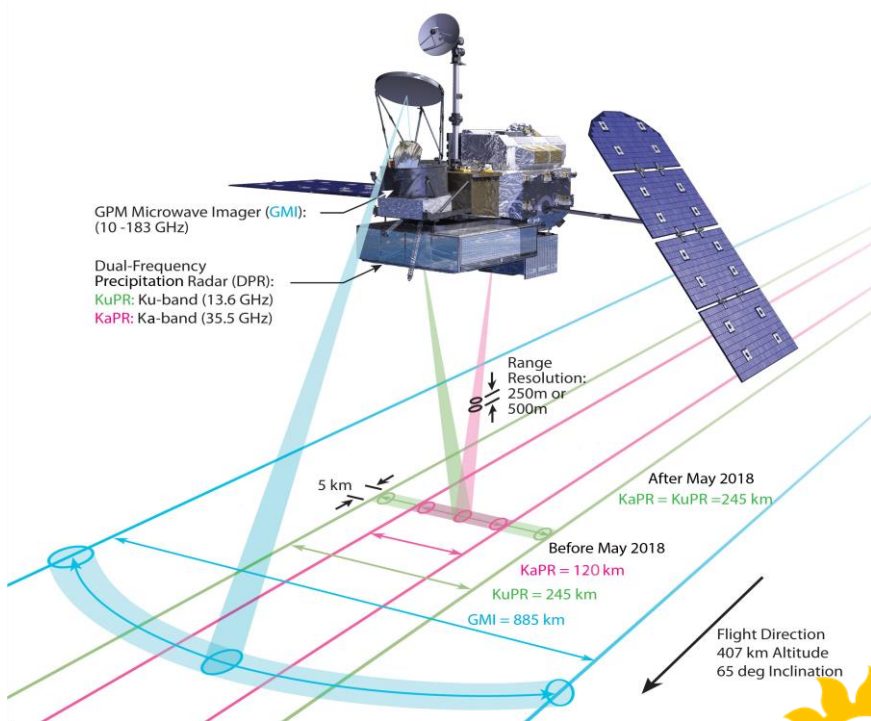


Data Satelital

TRMM: Tropical Rainfall Measurement Mission (1998 - 2013)



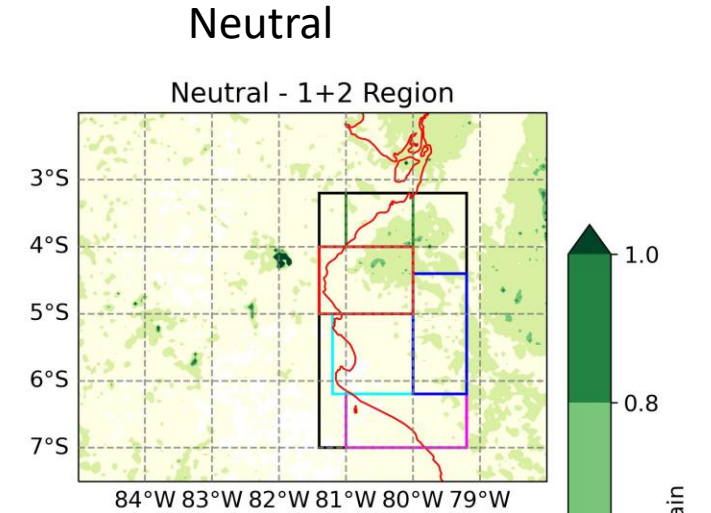
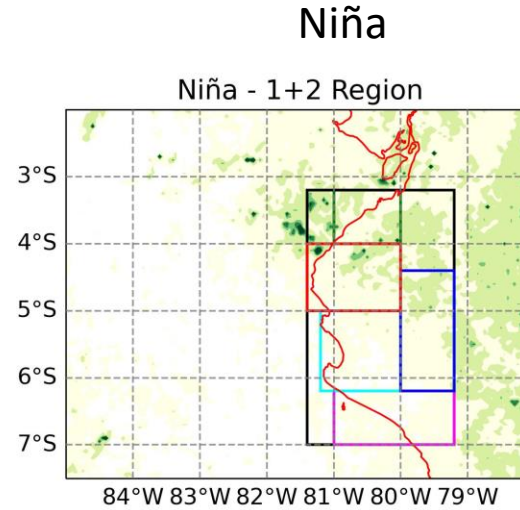
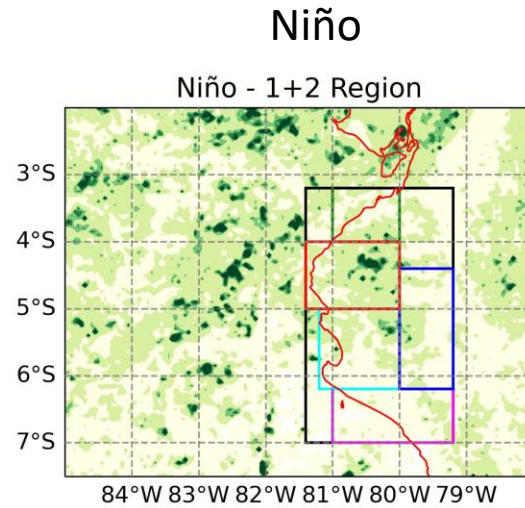
GPM: Global Precipitation Measurement (2014 - presente)



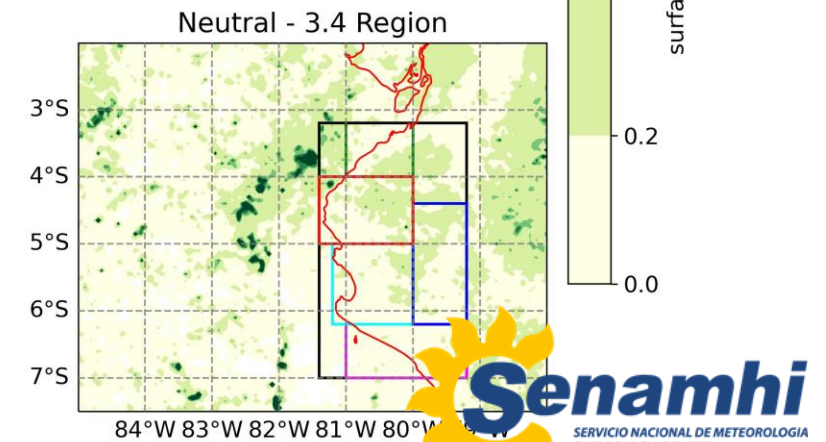
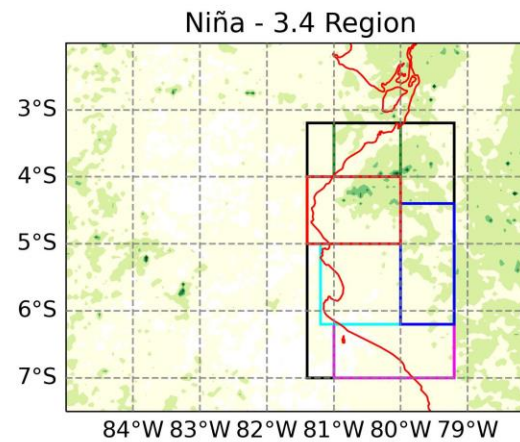
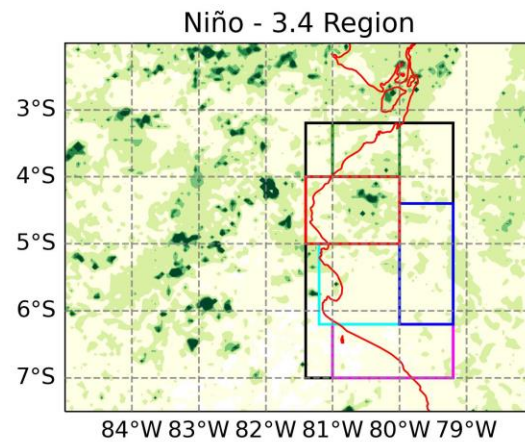
TRMM & GPM

Precipitación promedio sobre diferentes fases del ENOS

Region Niño 1+2
ICEN



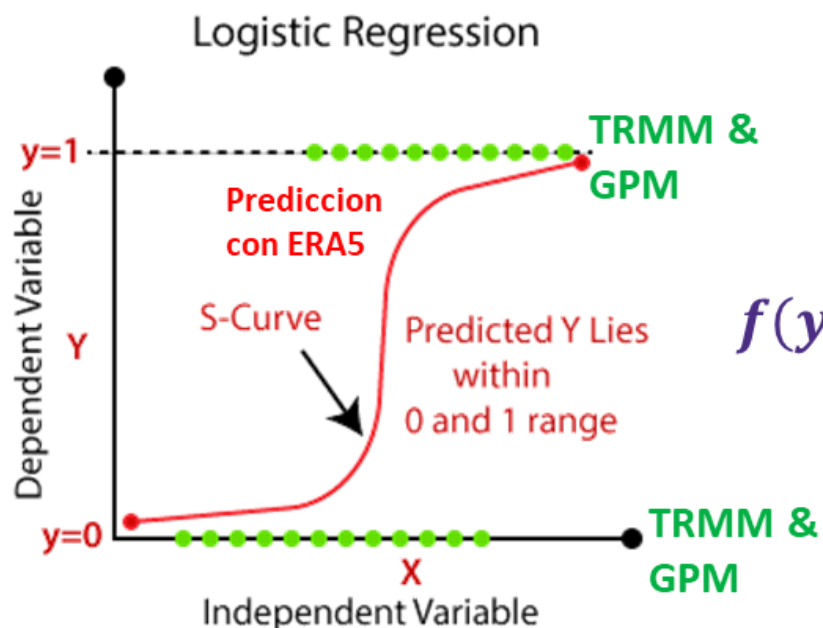
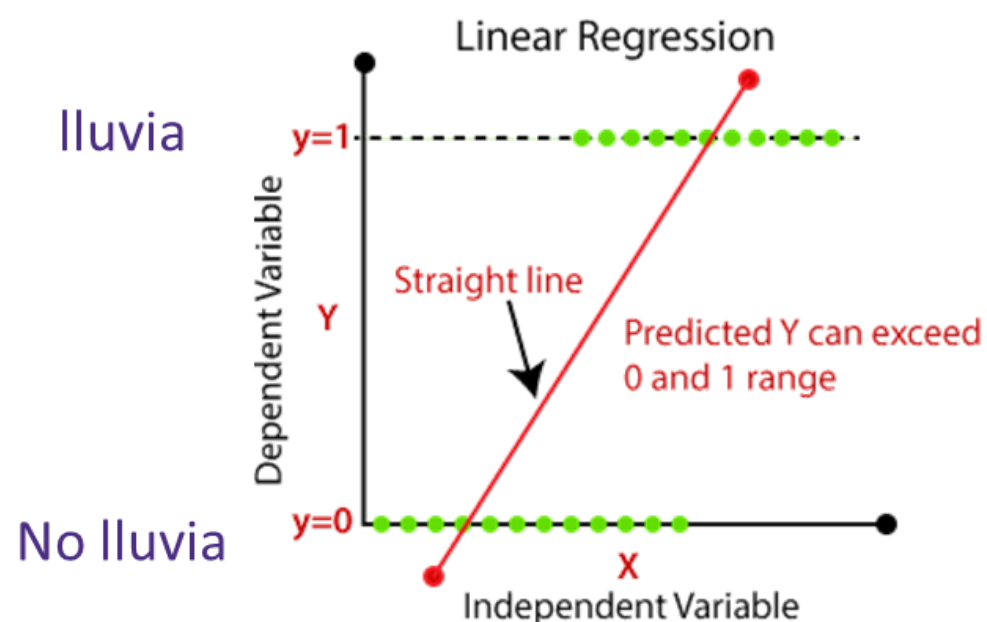
Region Niño 3.4
ONI



Metodología para Regresión Logística (RL)

Regresión Logística (RL)

Es un método de clasificación para variables binarias (0,1) que modela la probabilidad de un resultado (predicción) dadas variables de entrada (ERA5).

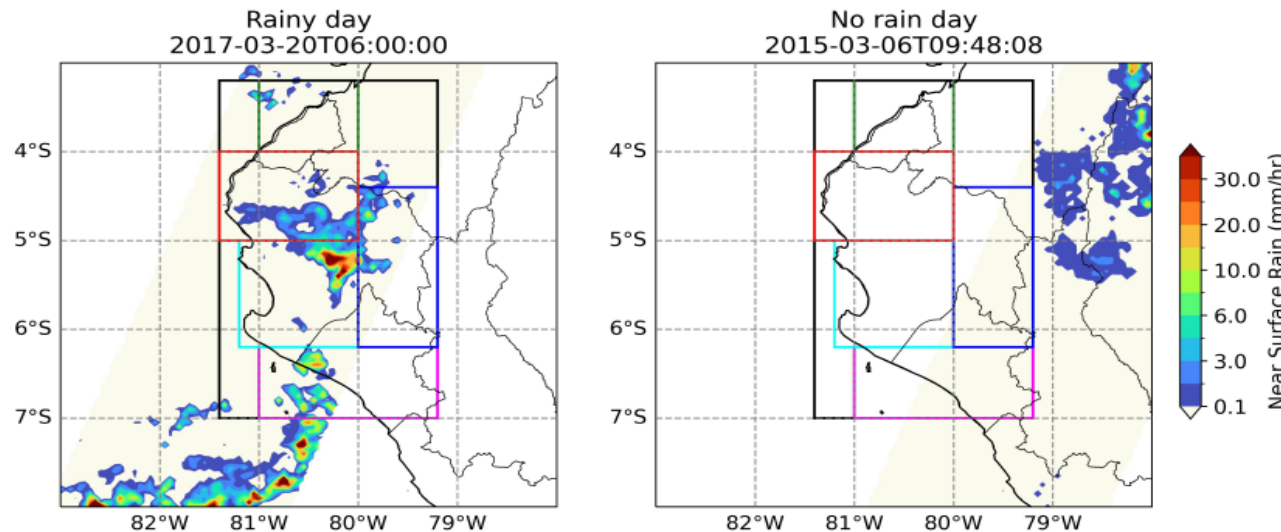


$$f(y) = \frac{1}{1 + e^{-y}}$$

$$y(x) = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$

TRMM & GPM (Procesamiento RL)

- Los archivos NetCDF de TRMM y GPM deben procesarse para que sean el predictor en LR. Días lluviosos (1) y días sin lluvia (0) utilizando las variables de lluvia en franja y superficie.
- Para que se considere un día lluvioso, la suma de todos los píxeles de lluvia debe ser igual o superior a estos umbrales de precipitación (PT): 25,50,100,150 y 200 mm/región



Swath and surface rain present in the area of study

Only swath is present inside the area of study, surface rain is outside.

Numero de dias en cada lista (1998 - 2021)

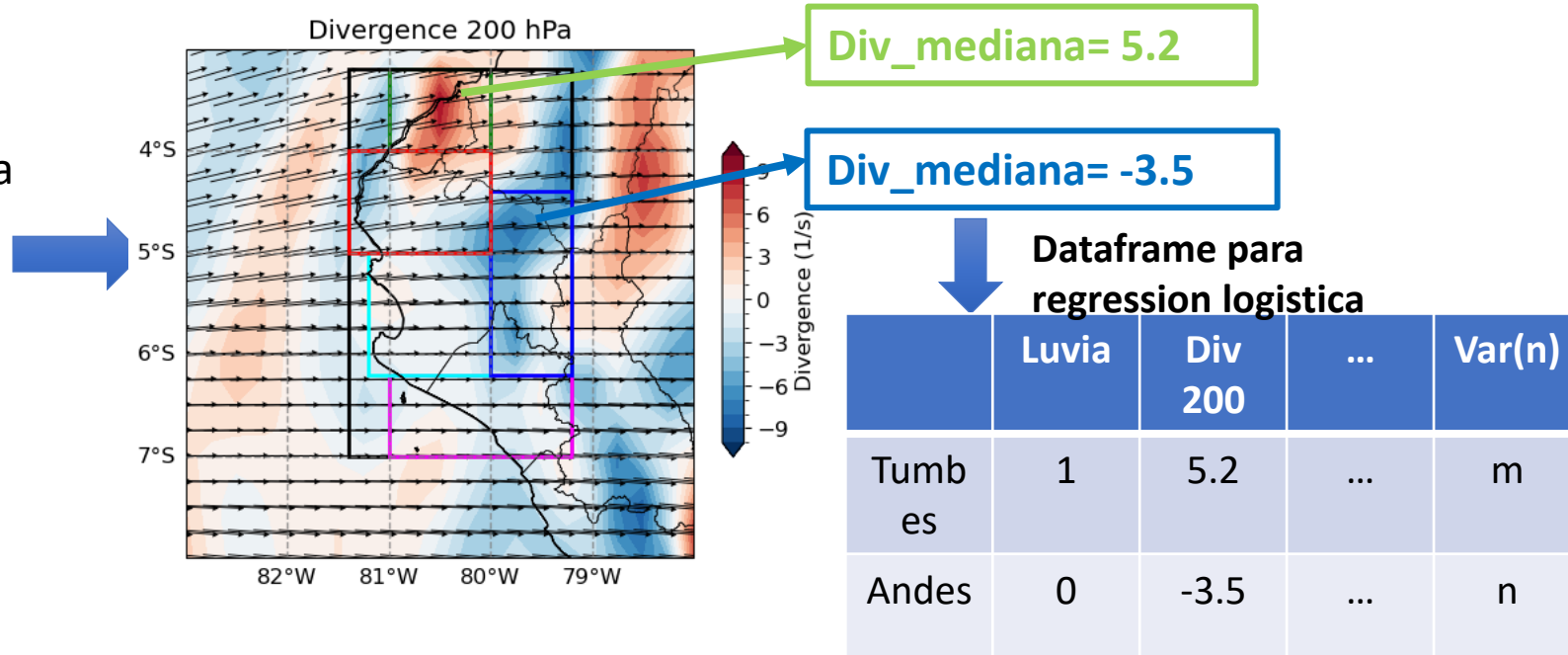
Color box	Precipitation Threshold (PT)				
	25	50	100	150	200
red	522	459	416	386	370
blue	629	555	481	444	421
cyan	446	414	390	382	372
green	519	448	390	372	353
magenta	456	423	398	387	380
black	1015	883	743	669	615

ERA5

- Con base en los listados de días lluviosos y sin lluvia, se calcularon las siguientes variables (22 en total) para dichos días. Posteriormente, la mediana de cada variable en cada cuadro de color fue su valor representativo en el marco de datos para el LR.

- ONI and ICEN
- Temperatura potencial equivalente en: 950, 850, 700, 500 hPa
- Relación de Mezcla en: 700, 600, 500 hPa
- Divergencia: 950, 850, 250, 200 hPa
- Vorticidad at 500 hPa
- Viento cortante en U y V 1000 – 400 hPa
- Agua Precipitable
- CAPE
- Gálvez Davison Index

Por Ejemplo:



Ejemplo de dataframe

date	time	sum_rain	rain	ICEN	ONI	EPT_950	EPT_850	EPT_700	EPT_500	MR_700	MR_600	MR_500	Div_200	Div_250	Div_850	Div_950	Vor_500	U_WS_1000_400	V_WS_1000_400	PWAT	CAPE	Lifted_Index	K_Index	Total_Totals	GDI
20090218	105343	0	0	-0.5	-0.8	339.6057	339.4839	335.2428	335.2428	8.097013	5.536668	2.925583	0.587617	0.587617	-0.60995	-0.54376	-0.74691	-0.18309	-5.19162	50.30872	230.9174	-1.59455	33.83253	43.25966	23.50887
20170221	41916	2164.493	1	1.37	-0.2	347.8669	342.8797	336.9108	336.9108	8.373425	4.456059	3.032503	0.816434	0.816434	0.985973	-0.31616	-0.51602	-10.7678	-0.60533	54.2646	1075.886	-3.00015	34.48869	43.23373	27.32958
20020202	55904	0	0	0.03	0	338.4777	343.2789	336.7593	336.7593	9.253501	6.337186	2.335266	-1.05455	-1.05455	0.324751	-0.51232	0.041432	5.186409	8.334239	51.97472	0	4.842742	34.35359	40.01634	7.528259
20030316	42155	1205.724	1	-0.18	0.4	337.3218	340.7756	341.1934	341.1934	7.37759	6.229274	4.620201	-0.87197	-0.87197	0.337063	-0.61246	-0.14102	-12.9412	7.082695	50.85138	43.09318	-0.93303	32.44507	44.43022	32.00088

... (Cientos de filas)

20010119	202350	0	0	-0.68	-0.7	336.294	335.4059	339.4644	339.4644	8.21093	6.270047	4.312819	-0.41879	-0.41879	0.207748	-1.79026	-0.0604	-22.6226	-7.70708	51.87083	0	1.357216	33.25291	41.57431	30.47244
----------	--------	---	---	-------	------	---------	----------	----------	----------	---------	----------	----------	----------	----------	----------	----------	---------	----------	----------	----------	---	----------	----------	----------	----------

Tabla de contingencia

Utilizando el recuento de a, b, c, d de la tabla de contingencia, se calculan las siguientes métricas:

Probability of Detection (POD):

$$POD = \frac{a}{a + c}$$

False Alarm Ratio (FAR):

$$FAR = \frac{b}{b + a}$$

Critical Success Index (CSI):

$$CSI = \frac{a}{a + b + c}$$

Tabla de contingencia		Lluvia observada	
		Si	No
Lluvia pronosticada	Si	Positivo verdadero (a)	Falso positivo (b)
	No	Falso negativo (c)	Negativo verdadero (d)

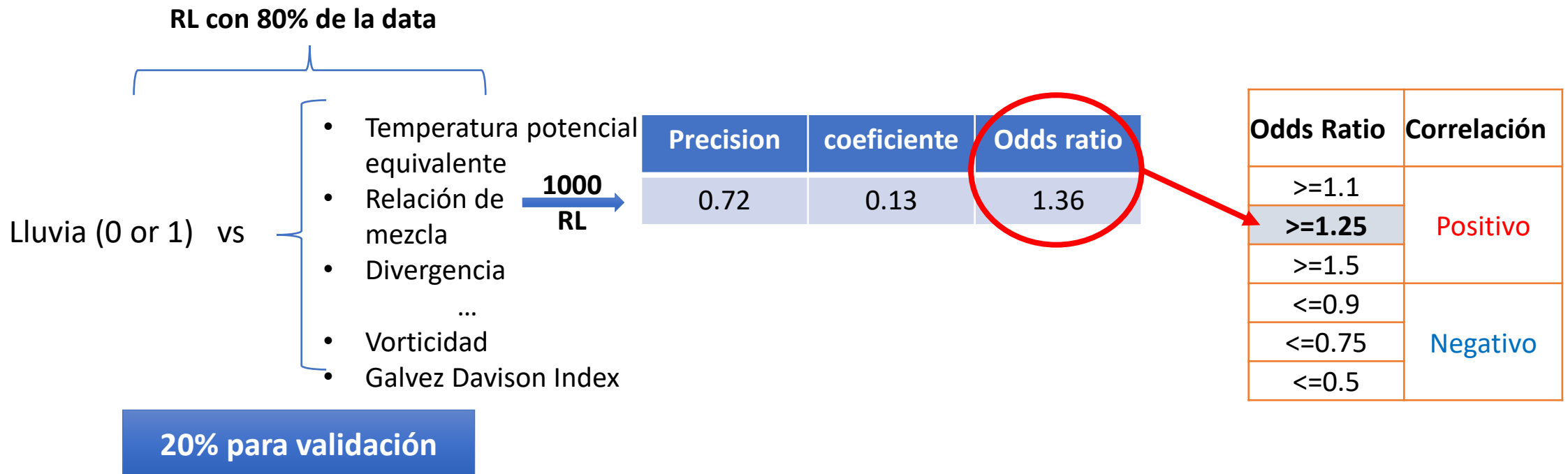
Resultados

Regresion logistica individual por variable

date	time	sum_rain	rain	ICEN	ONI	EPT_95_0	EPT_85_0	EPT_70_0	EPT_50_0	MR_70_0	MR_60_0	MR_50_0	Div_20_0	Div_25_0	Div_85_0	Div_95_0	Vor_50_0	U_WS_1000_4_00	V_WS_1000_4_00	PWAT	CAPE	Lifted_Index	K_Index	Total_Totals	GDI		
20090218	105343	0	0	-0.5	-0.8	339.6057	339.4839	335.2428	335.2428	8.097013	5.536668	2.925583	0.587617	0.587617	0.609955	0.54376	0.74691	0.18309	5.19162	50.30872	230.9174	1.59455	33.83253	43.25966	23.50887		
20170221	41916	2164.493	1	1.37	-0.2	347.8669	342.8797	336.9108	336.9108	8.373425	4.456059	3.032503	0.816434	0.816434	0.985973	0.31616	0.51602	10.7678	0.60533	54.2646	1075.886	3.00015	34.48869	43.23373	27.32958		
20020202	55904	0	0	0.03	0	338.4777	343.2789	336.7593	336.7593	9.253501	6.337186	2.335266	1.05455	1.05455	0.324751	0.51232	0.041432	5.186409	8.334239	51.97472	0	4.842742	34.35359	40.01634	7.528259		
20030316	42155	1205.724	1	-0.18	0.4	337.3218	340.7756	341.1934	341.1934	7.37759	6.229274	4.620201	0.87197	0.87197	0.337063	0.61246	0.14102	12.9412	7.082695	50.85138	43.09318	0.93303	32.44507	44.43022	32.00088		
... (Cientos de filas)																											
20010119	202350	0	0	-0.68	-0.7	336.294	335.4059	339.4644	339.4644	8.21093	6.270047	4.312819	0.41879	0.41879	0.207748	1.79026	-0.0604	22.6226	7.70708	51.87083	0	1.357216	33.25291	41.57431	30.47244		



Regresion logistica individual por variable



Resultado de 1000 RL

Aceptado	Variable	Conteo de Odds Ratio					
		>=1.1	>=1.25	>=1.5	<=0.9	<=0.75	<=0.5
Si	Div_200	70	30.7	3.2	2.4	0.2	0
	Div_250	69.8	30.9	3.2	2.5	0.2	0
	Div_850	0.4	0.1	0	93.3	52.3	3
	Div_950	0.5	0.1	0	79.7	53.7	2.4
	GDI	50.7	0	0	0	0	0
	MR_500	100	99.9	95.8	0	0	0
	MR_600	100	100	98.1	0	0	0
	MR_700	100	99.8	83.4	0	0	0
	PWAT	100	41.8	0	0	0	0
no	CAPE	0	0	0	0	0	0
	EPT_500	25.2	14.3	2.3	0	0	0
	EPT_700	25.1	14	2.3	0	0	0
	EPT_850	22.6	5	0	0	0	0
	EPT_950	43.5	10.7	0	0	0	0
	ICEN	87.4	79	46.7	0.9	0	0
	ONI	54.4	23.8	7.7	13.5	1.2	0
	U_WS_1000_400	0	0	0	2.4	0	0
	V_WS_1000_400	0.3	0	0	0	0	0
	Vor_500	17.7	1.9	0.1	17	0.4	0

Las variables aceptadas tienen mas de 50% de apariciones por encima o debajo del umbral de Odds Ratio

Bajo porcentaje de Odds Ratio, ICEN y ONI no fueron tomados en cuenta debido a que no se adecuan a un pronostico a nivel operative.

Todas las combinaciones

- Trabajar con las 12 variables seleccionadas permite evaluar cada combinación posible de ellas en grupos de 2, 3, 4,... 12.
- Otro aspecto en las combinaciones será el PT utilizado para determinar un día lluvioso (25, 50, 100, 150, 200 mm)
- El umbral de validación (VT) de 60, 70 y 80 (utilizado para determinar si una predicción se considera lluvia o no lluvia)
- Regiones en la Costa norte

	Variables seleccionadas
1	Div_200
2	Div_250
3	Div_850
4	Div_950
5	GDI
6	K_Index
7	Lifted_Index
8	MR_500
9	MR_600
10	MR_700
11	PWAT
12	Total_Totals

→ $C_2^{12}, C_3^{12}, C_4^{12}, \dots, C_{12}^{12}$

Por ejemplo
 Div 850 and MR 500 (2)
 GDI, PWAT, Div 200 (3)
 Div 850, MR 600, K index, GDI (4)

**4095
combinaciones
posibles!**



PT
25
50
100
150
200



VT
60
70
80



Regiones
Total
Tumbes
Amotape
Andes
Sechura
Lambayeque



**Todas las
combinacion
es posibles**

**368550
RL's**

80%  Calculado con 80% of data

20%  Calculado con 20% of data

region	PT	VT	Variables elegidas
Amotape	100	70	('Div_200', 'GDI', 'PWAT')

Variable	Coeficiente
Div 200	0.18
GDI	0.09
PWAT	1.15

Intercepto	Precision	Validacion precision	Diferencia precision
-9.72071	0.678571	0.638095	0.040476

Tabla de contingencia

Verd. Positivo (a)	Falso positivo (b)	Falso negativo (c)	Verd. negativo (d)
18	10	28	49

POD	FAR	CSI
0.39	0.36	0.32



POD

Mas alto es mejor

FAR

Mas bajo es mejor

VT=60

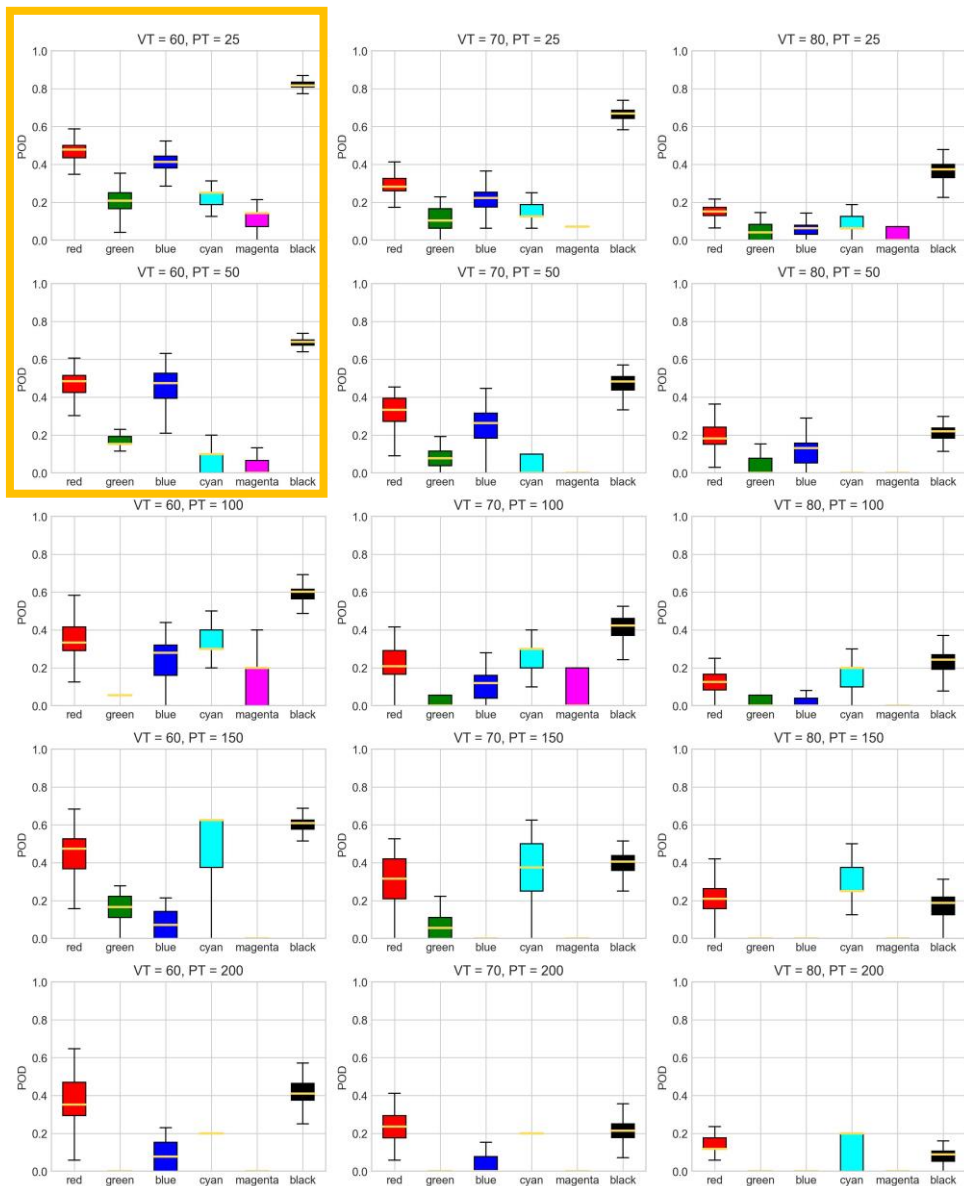
VT=70

VT=80

VT=60

VT=70

VT=80



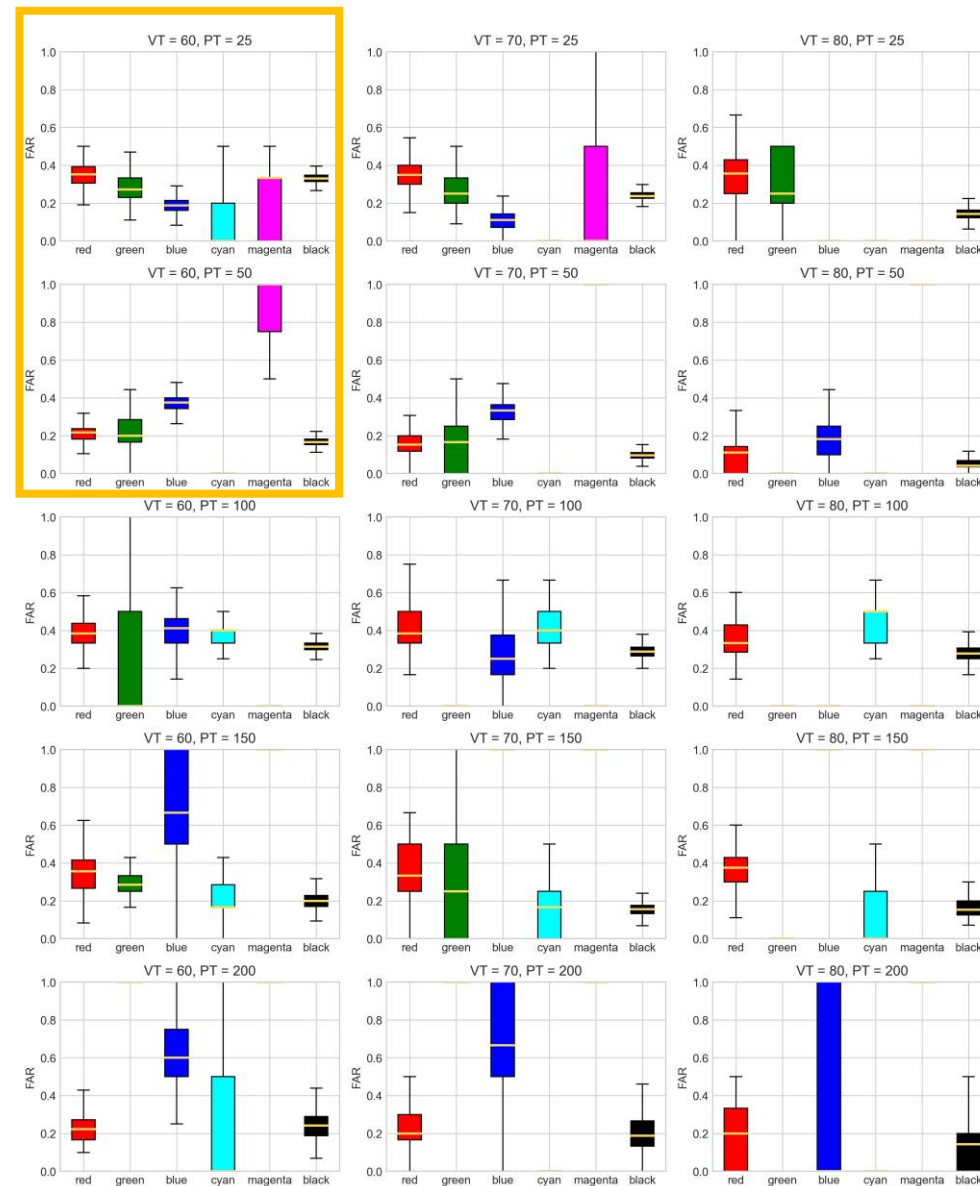
PT=25

PT=50

PT=100

PT=150

PT=200



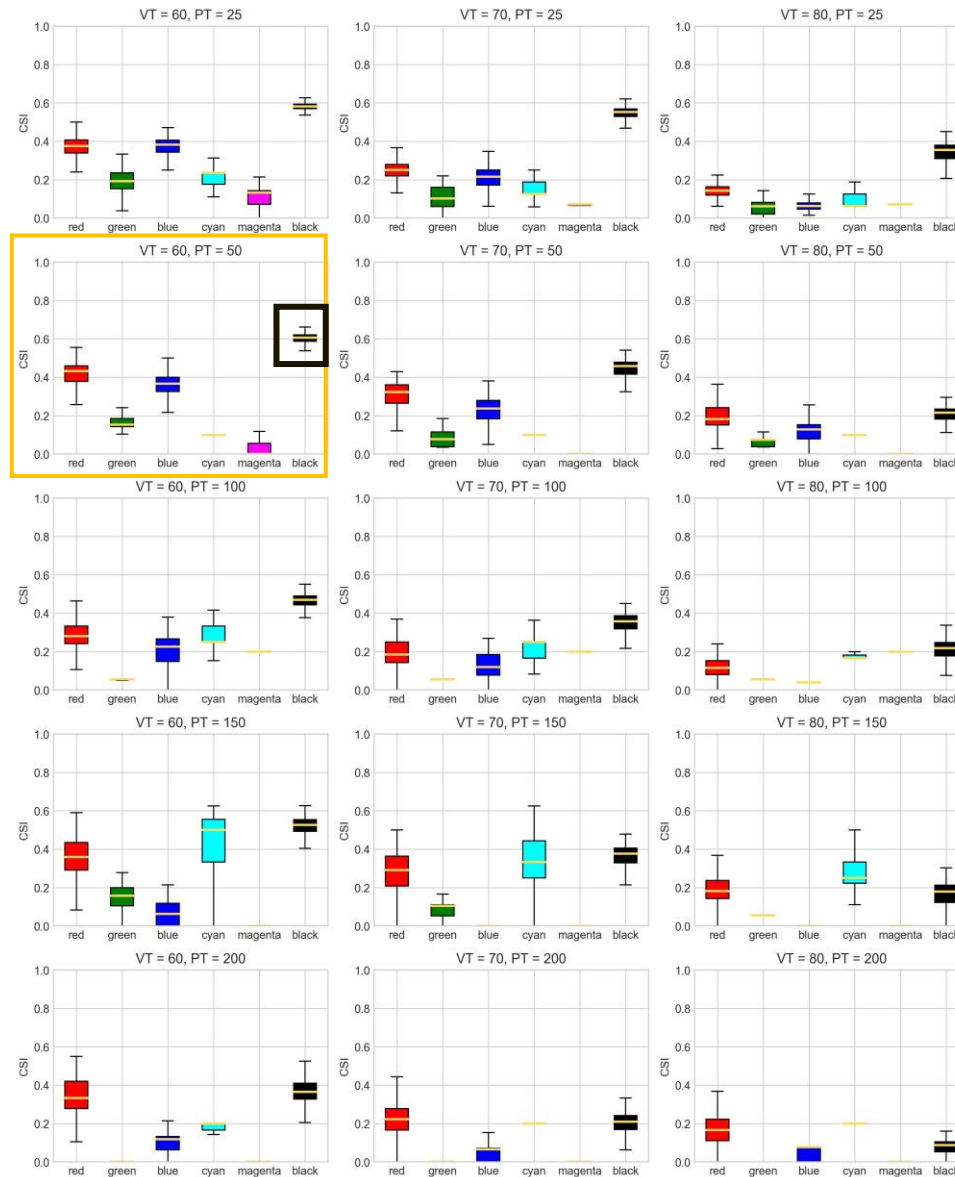
CSI

Mas alto es mejor

VT=60

VT=70

VT=80



PT=25

PT=50

PT=100

PT=150

PT=200

Mejor combinación

- Según los diagramas de caja, la mejor combinación parcial resultó ser PT=50, VT=60 y color box = negro. Ahora es el momento de encontrar la mejor combinación de variables dentro de estas características.
- Al ordenar todas las combinaciones de variables por el CSI más alto se obtienen las siguientes seis combinaciones:

Region	PT	VT	Variables	CSI	POD	FAR
Total	50	60	('MR_700', 'Div_950', 'Div_250', 'GDI')	0.66	0.74	0.13
			('MR_700', 'Div_950', 'Div_200', 'GDI')			
			('MR_700', 'Div_950', 'Div_250', 'Div_200', 'GDI')			
			('Div_950', 'Div_250', 'GDI', 'PWAT')			
			('Div_950', 'Div_200', 'GDI', 'PWAT')			
			('Div_950', 'Div_250', 'Div_200', 'GDI', 'PWAT')			

Mejor combinación

Fueron elegidos porque estas variables están presentes en los patrones de circulación que incrementan las precipitaciones en la costa norte peruana.

- **Relación de mezcla a 700 hPa**: valores superiores a 9 g/Kg favorecen la precipitación. (Quispe, 2018)
- **Divergencia a 950 hPa**: segunda banda de la ZCIT proporciona convergencia (divergencia negativa), (Quispe, 2018)
- **Divergencia a 250 hPa**: el sistema acoplado BH y NL proporciona divergencia en niveles altos. (Quispe, 2018) y (Aliaga et al, 2022)
- **GDI**: Presenta un perfil termodinámico de la atmósfera baja y media (Gálvez y Davison, 2016)

Region	PT	VT	Variables	CSI	POD	FAR
Total	50	60	('MR_700', 'Div_950', 'Div_250', 'GDI')	0.66	0.74	0.13
			('MR_700', 'Div_950', 'Div_200', 'GDI')			
			('MR_700', 'Div_950', 'Div_250', 'Div_200', 'GDI')			
			('Div_950', 'Div_250', 'GDI', 'PWAT')			
			('Div_950', 'Div_200', 'GDI', 'PWAT')			
			('Div_950', 'Div_250', 'Div_200', 'GDI', 'PWAT')			

RAMI (Rivas, Anderson-Frey, McMurdie Index)

Algoritmo de Cálculo:

Variable	Relación de Mezcla 700 hPa	Divergencia 950 hPa	Divergencia 250 hPa	GDI	Intercepto
Coficiente	0.16	-0.27	0.15	0.06	-2.53

1^{ro}

Variable por coeficiente (ERL):

$$p = 0.16 * MR_{700} - 0.27 * D_{950} + 0.15 * D_{250} + 0.06 * GDI - 2.53$$

2^{do}

Función logística: $LF = \frac{1}{1 + e^{-p}}$

3^{ro}

Probabilidad de Lluvia (%):

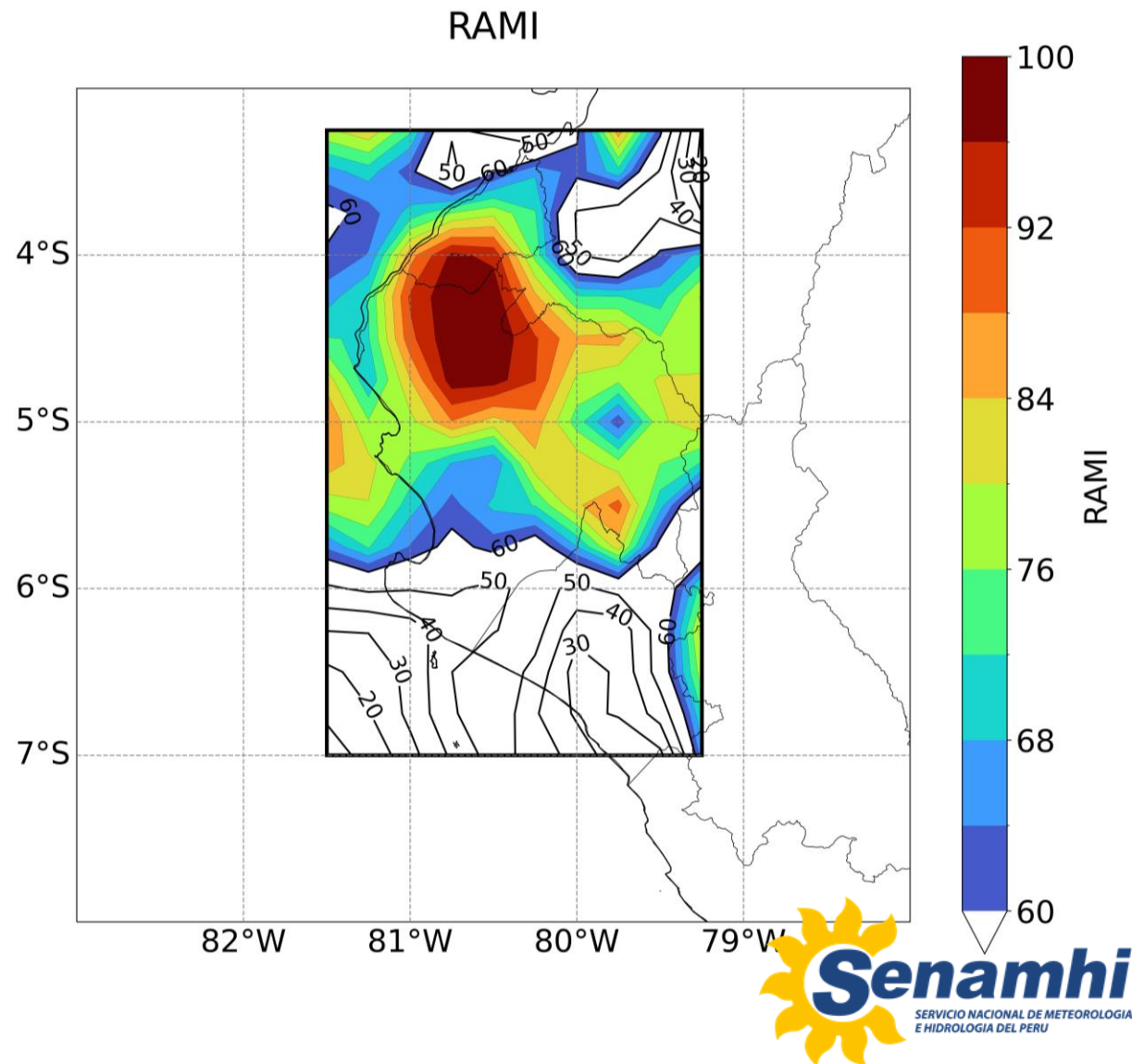
$$RAMI = LF * 100$$

RAMI

Después de calcular RAMI, graficarla requiere los siguientes pasos:

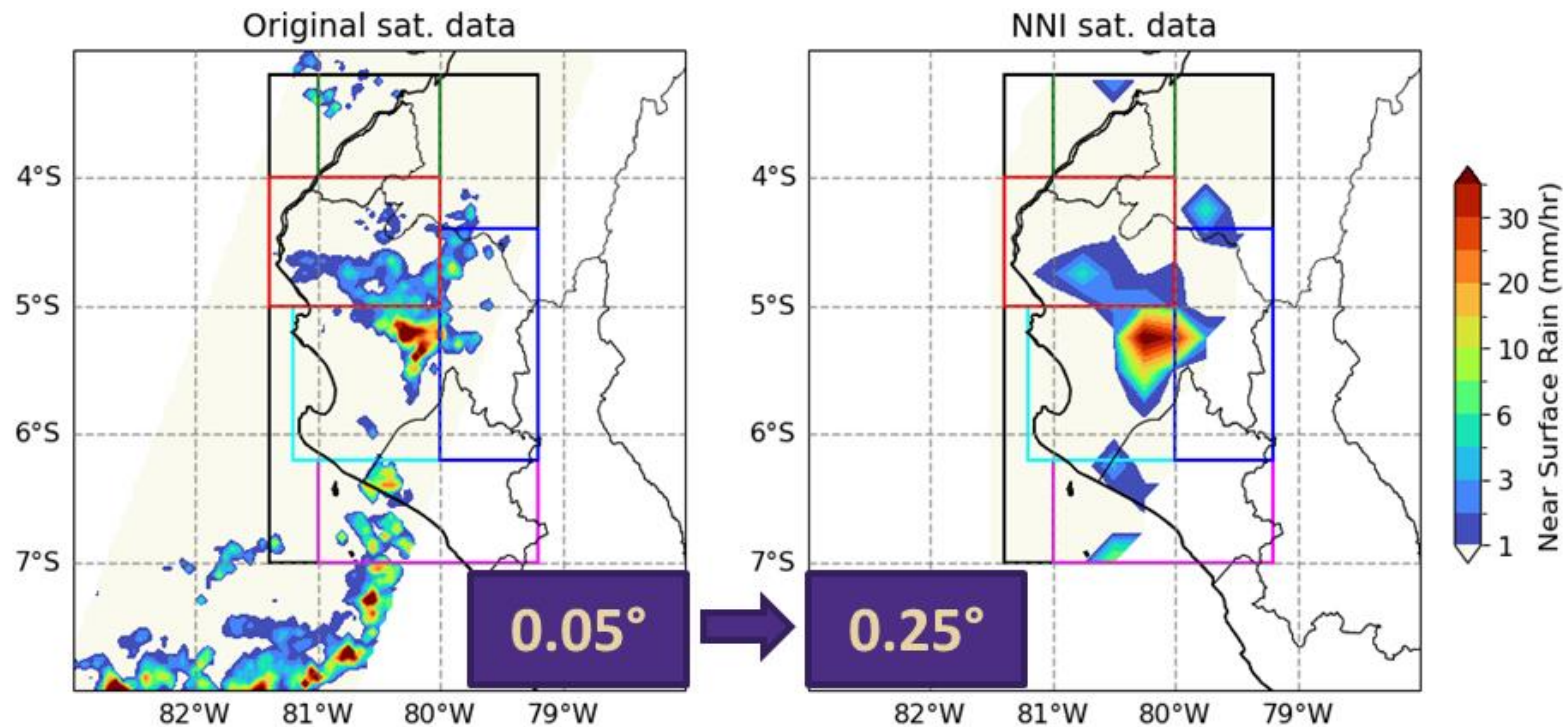
- Utilice el mapa de colores “Turbo” de Matplotlib para los valores entre 60 y 100, con contornos rellenos de color cada 4 valores (60, 64, 68,...96, 100)
- Utilice líneas negras continuas para los valores entre 0 y 60, con contornos cada 10 valores (0, 10, 20... 50, 60)
- Trazar únicamente dentro del área del cuadro negro (latitud: -3,25; -7°S y longitud: -81,5; -79,5°W)
- Debe usarse únicamente en los meses de: enero, febrero, marzo, abril.

Forma final de RAMI



Validación pixel por pixel

Dado que se trata de una comparación de datos satelitales (resolución de $0,05^\circ$) con ERA5 (resolución de $0,25^\circ$), la resolución más pequeña debe ampliarse utilizando la interpolación del vecino más cercano (Nearest Neighbor Interpolation).

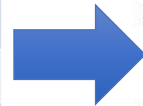


Validación pixel por pixel

- Una vez realizada la interpolación, es hora de comparar ambos conjuntos de datos.
- Se vuelve a utilizar la tabla de contingencia.
- El umbral de validación para el nuevo índice es 60.

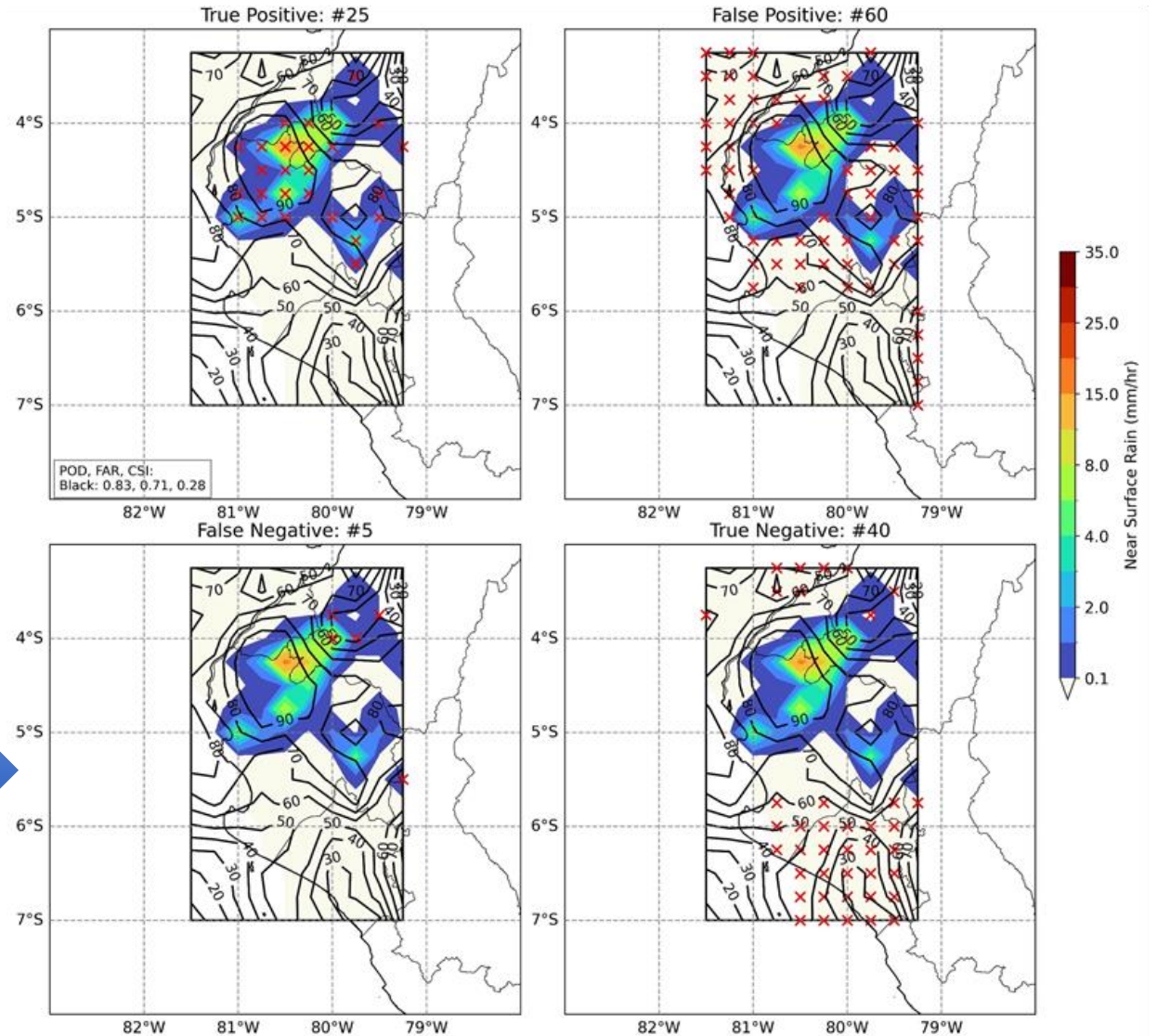
Por ejemplo:

Tabla de contingencia		Lluvia observada	
		Si	No
Lluvia pronosticada	Si	Positivo verdadero (a)	Falso positivo (b)
	No	Falso negativo (c)	Negativo verdadero (d)



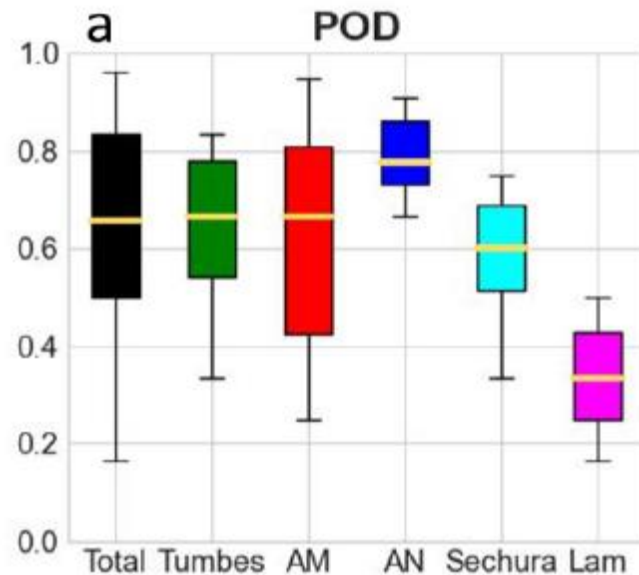
Lineas negras=RAMI
Colores=Lluvia observada

✗ Grilla en cada categoría

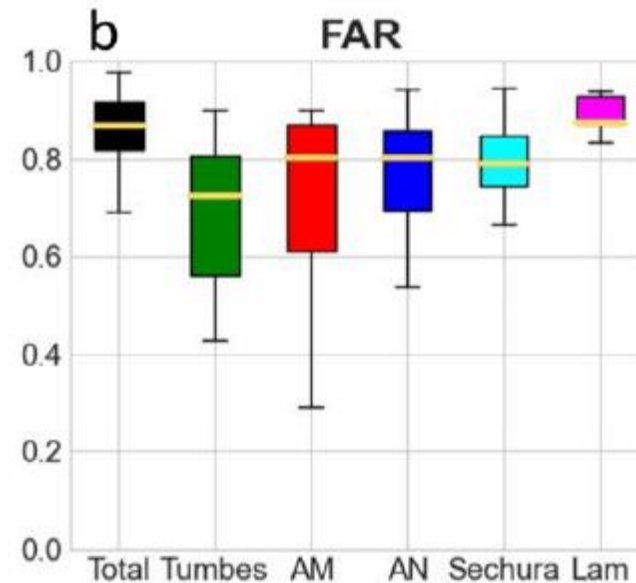


Validación pixel por pixel

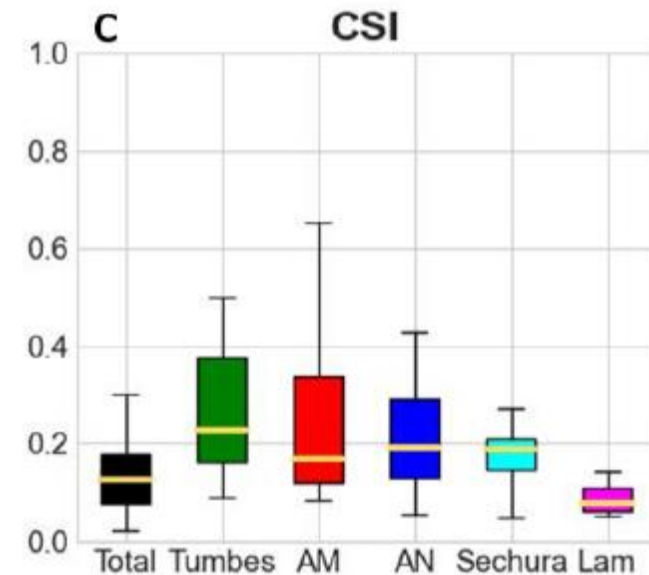
Altos valores de POD con alta varianza en el area Total



Altos valores de FAR indicant sobre estimacion en el pronostico



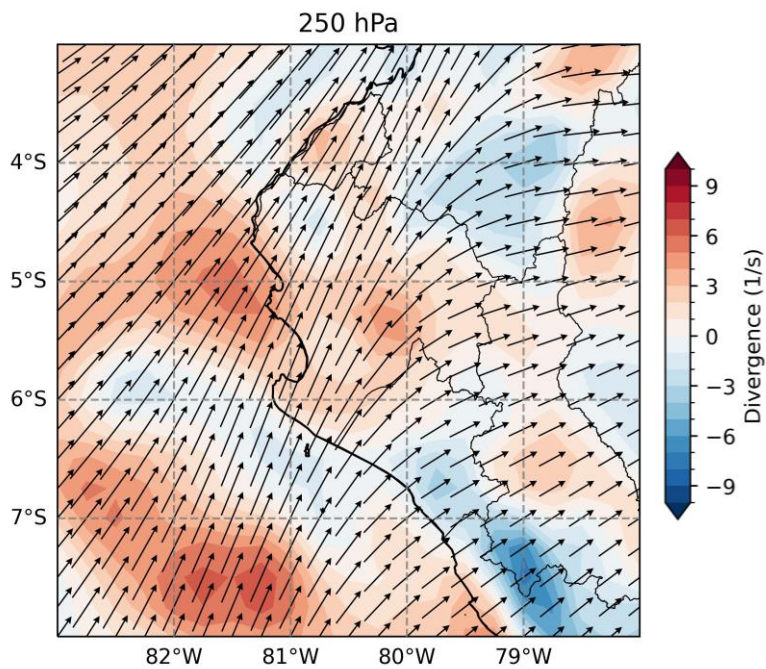
Bajos valores de CSI debido a los altos valores de FAR



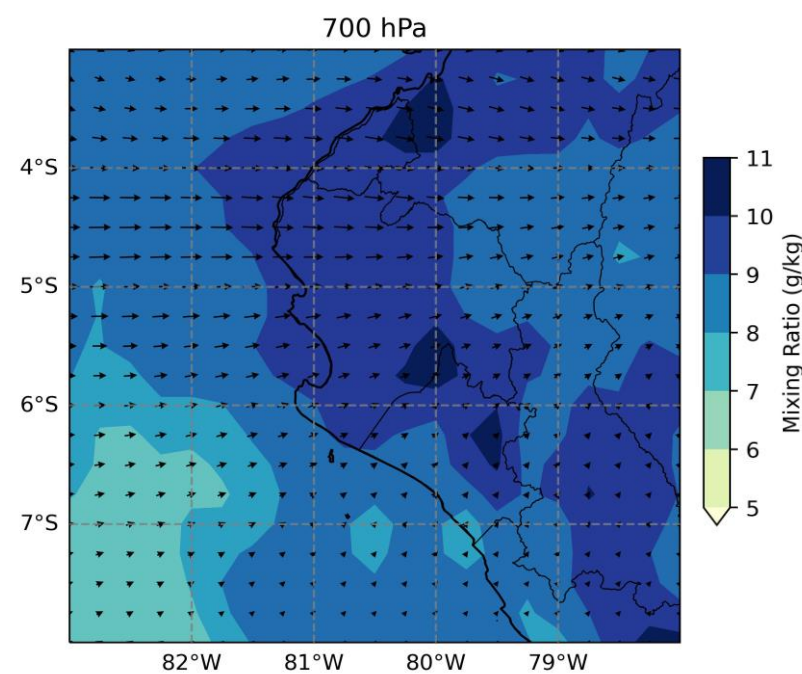
Casos de estudio (3 de Abril, 2023)

3 de Abril del 2023: 23h hora local

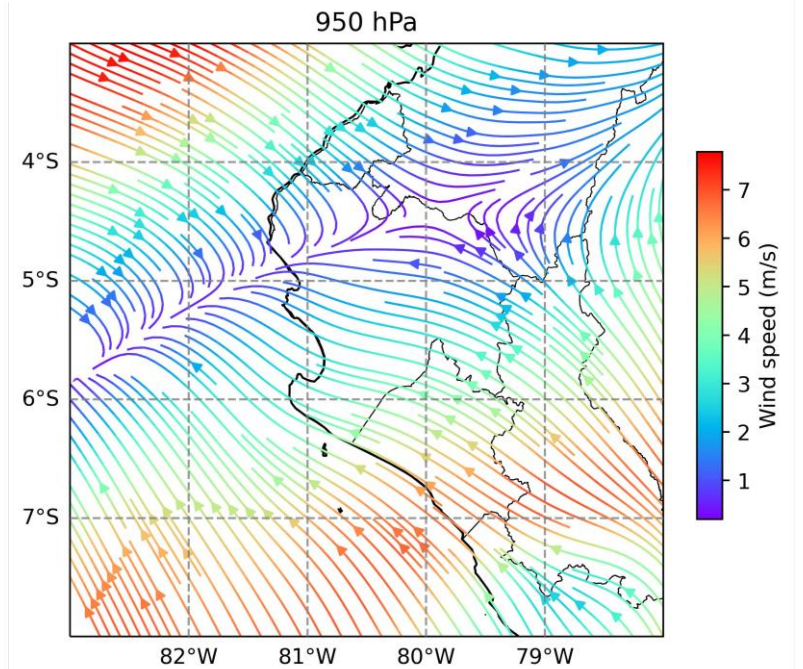
Niveles altos



Niveles medios



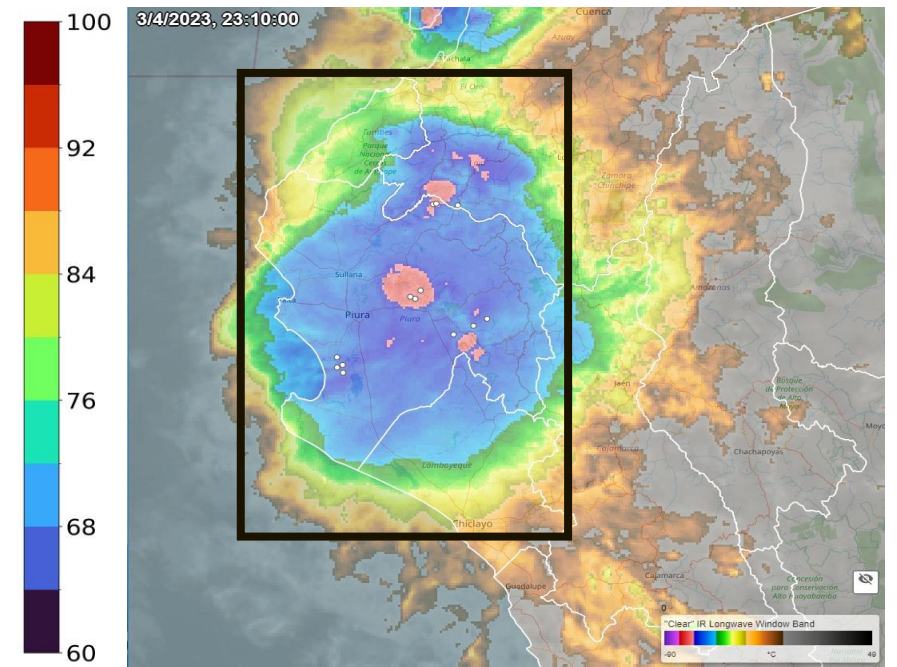
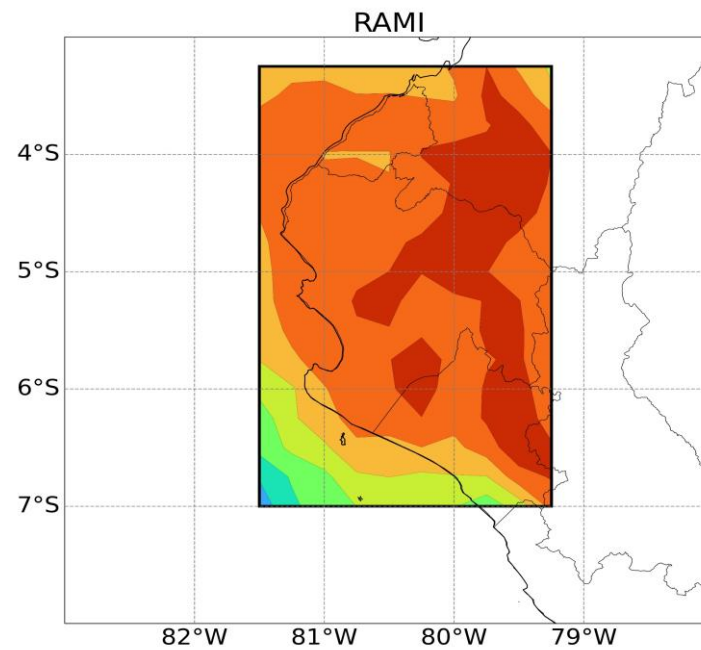
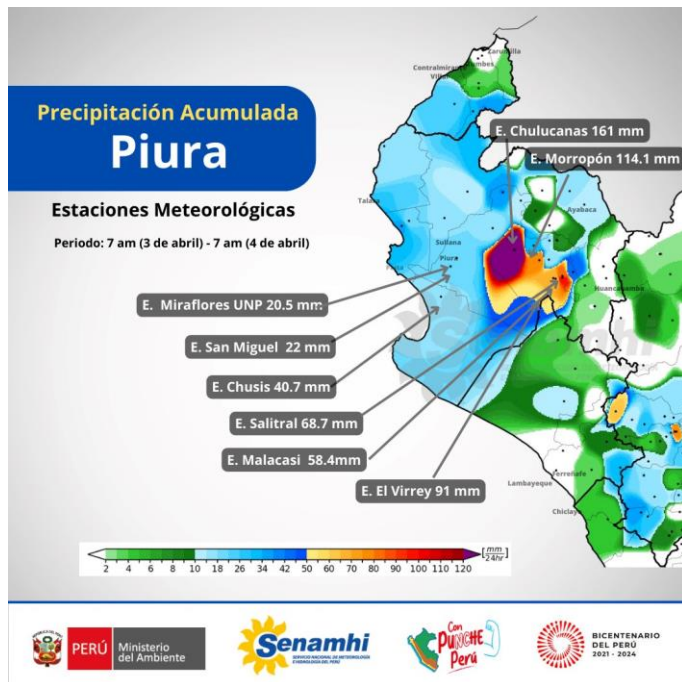
Niveles bajos



Casos de estudio (3 de Abril, 2023)

Precipitación total del 3 a las 7h al 4 a las 7h del 2023

3 de Abril del 2023: 23h Local time



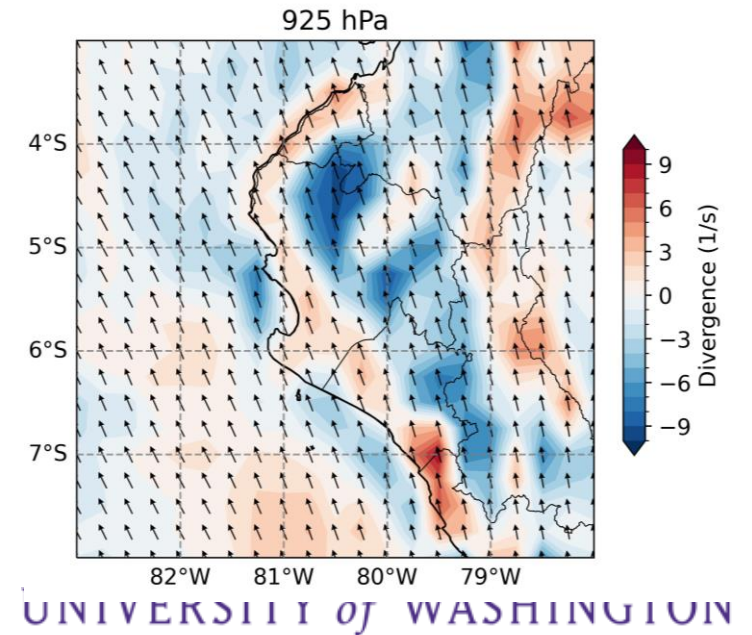
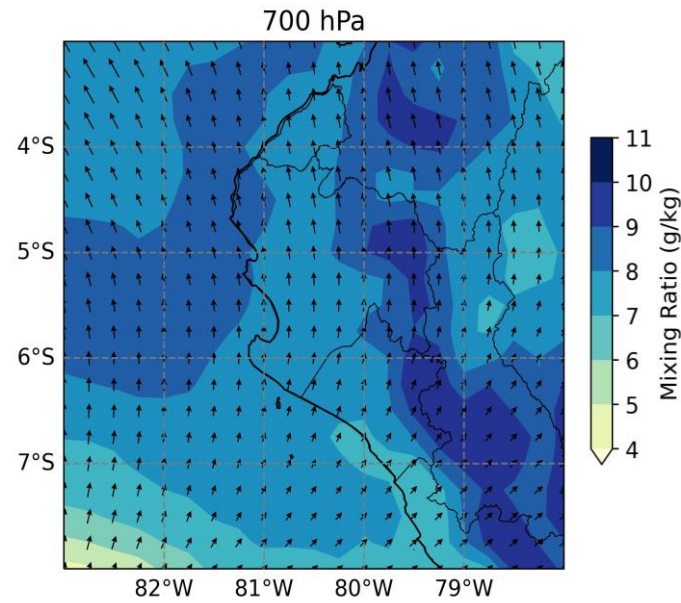
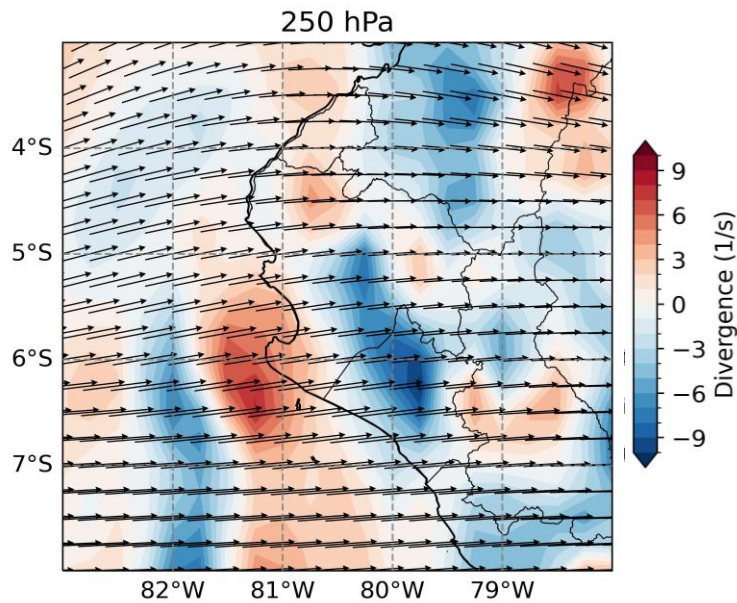
Casos de estudio (14 de Abril, 2023)

14 de Abril del 2023: 16h hora local

High

Middle

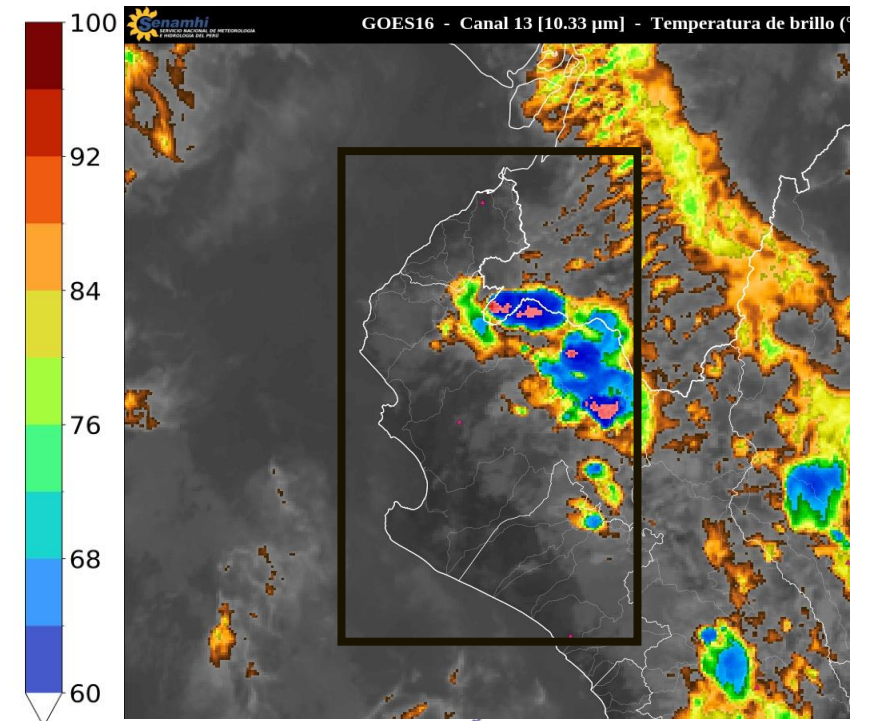
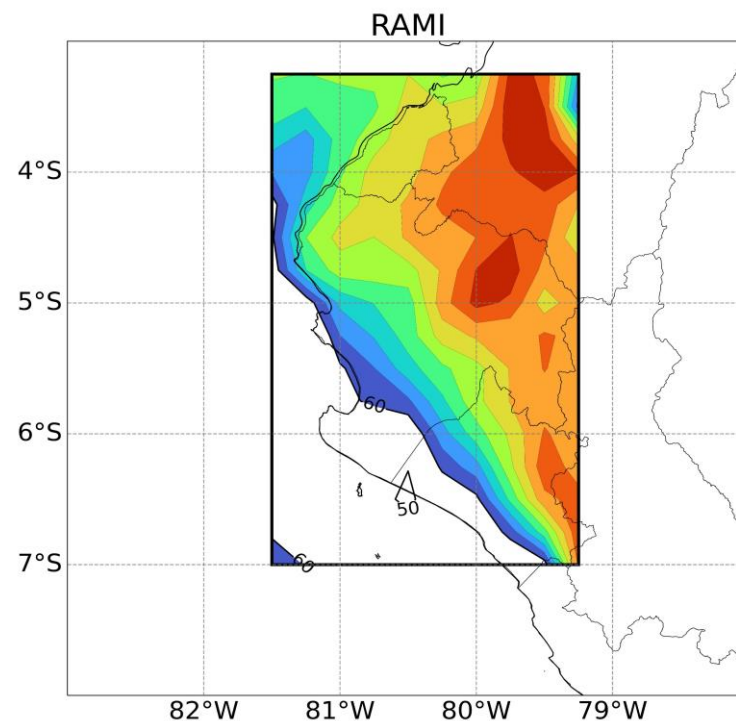
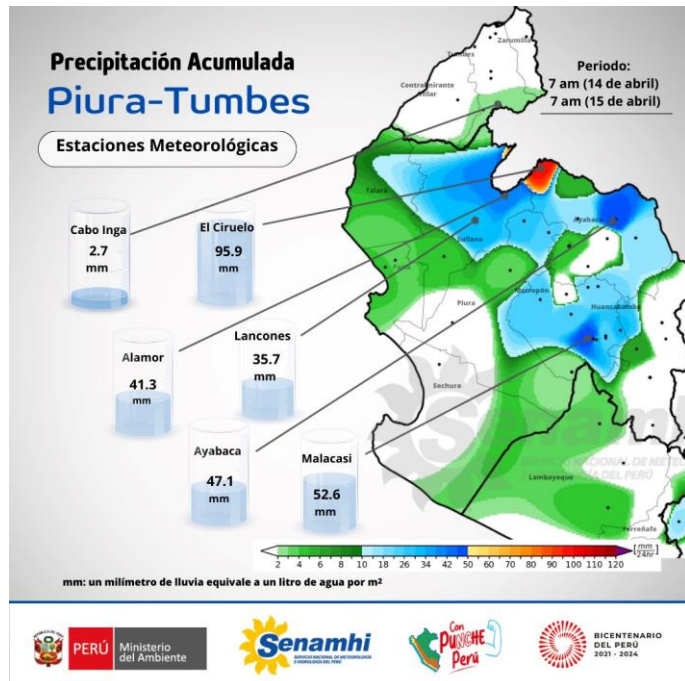
Low



Casos de estudio (14 de Abril, 2023)

14 a las 7H hasta las 15 a las 7H del 2023

14 de Abril 2023: 16h hora local



Conclusiones

- LR fue una herramienta o técnica útil para comprender la relación de las variables meteorológicas y su precisión para la predicción de la lluvia. Ayudó a seleccionar la mejor combinación de región, umbrales y variables meteorológicas.
- La mejor combinación de cuadro de color, PT, VT y variables meteorológicas se determinó mediante validación con un subconjunto de datos. Eran: Área total, PT=50, VT=60 y las variables: Relación de mezcla en 700 hPa, divergencia en 950 hPa, divergencia en 250 hPa y el GDI.

Conclusiones

- La validación de pixel por pixel mostró un desempeño decente del RAMI con valores altos de POD, aunque sobreestima la predicción de lluvia, lo que fue señalado por los altos valores de FAR.
- RAMI presentó un buen comportamiento al compararlo con las tormentas de 2023, identificando las zonas donde se produjo la convección con valores de RAMI superiores a 80.

Referencias

- Vannia Aliaga-Nestares, Diego Rodriguez-Zimmermann, and Nelson Quispe Gutierrez. Behavior of the ITCZ second band near the Peruvian coast during the 2017 coastal el niño. *Atmosfera*, 2022.
- Bureau of Meteorology. (n.d.). The three phases of the El Niño–Southern Oscillation (ENSO). The three phases of ENSO. Retrieved from <http://www.bom.gov.au/climate/enso/history/In-2010-12/three-phases-of-ENSO.shtml>
- Draper, D. W., Newell, D. A., Wentz, F. J., Krimchansky, S., & Skofronick-Jackson, G. M. (2015). The Global Precipitation Measurement (GPM) Microwave Imager (GMI): Instrument Overview and Early On-Orbit Performance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(7), 3452-3462. doi:10.1109/JSTARS.2015.2403303

Referencias

- Jose Manuel Galvez and Michel Davison. The galvez-davison index for tropical convection, 2016.
- Kelita del Rocio Quispe Vega. El niño costero 2017: Precipitaciones extraordinarias en el norte de peru. Master's thesis, 2018.
- NOAA's Climate Prediction Center. Noaa's climate prediction center, Jan 2001.
- L'Heureux, M. (2014). What is the El Niño-Southern Oscillation (ENSO) in a nutshell? NOAA Climate.gov. Retrieved from <https://www.climate.gov/news-features/blogs/enso/what-el-ni%C3%B1o%E2%80%93southern-oscillation-enso-nutshell>