

Predictibilidad de heladas agrometeorológicas en Uruguay en base a técnicas de aprendizaje automático



Santiago de Mello
IV Jornadas de estadística aplicada

Octubre 2017

Contenido

1. Introducción
2. Objetivos
3. Datos y metodología
4. Resultados
5. Conclusiones

Contenido

1. **Introducción**
2. Objetivos
3. Datos y metodología
4. Resultados
5. Conclusiones

1. Introducción

Definiciones:

Definiciones:

- Helada Agrometeorológica (HA) ($T_{\min} < 2^{\circ}\text{C}$)

Definiciones:

- **Helada Agrometeorológica (HA)** ($T_{\min} < 2^{\circ}\text{C}$)
- **Helada Agrometeorológica Generalizada (HAG)** (HA en el 75% de las estaciones disponibles)

Definiciones:

- **Helada Agrometeorológica (HA)** ($T_{\min} < 2^{\circ}\text{C}$)
- **Helada Agrometeorológica Generalizada (HAG)** (HA en el 75% de las estaciones disponibles)

Relevancia:

- **Impacto en el agro, ecosistemas, sector eléctrico, etc.**
- **Pocos estudios a escala intraestacional.**

Contenido

1. Introducción
2. **Objetivos**
3. Datos y metodología
4. Resultados
5. Conclusiones

Objetivos:

1. Mejorar la comprensión entre las HAG y los patrones de circulación general de la atmósfera y los océanos

Objetivos:

1. Mejorar la comprensión entre las HAG y los patrones de circulación general de la atmósfera y los océanos
2. Explorar la predictibilidad de estos fenómenos a escala intraestacional.

Contenido

1. Introducción
2. Objetivos
- 3. Datos y metodología**
4. Resultados
5. Conclusiones

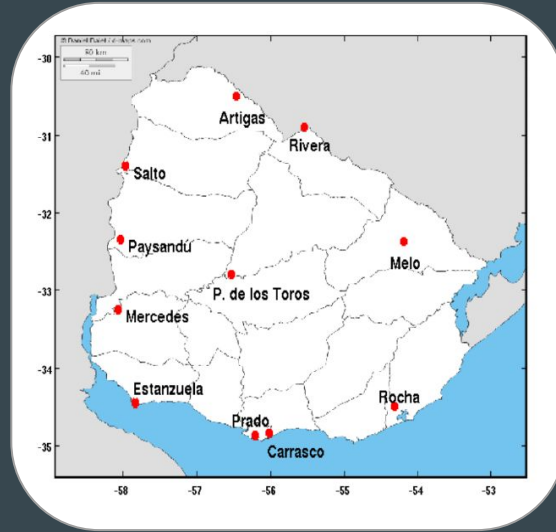
**Datos para patrones de
circulación y océanos**

Datos para definir HAG

Datos para definir HAG

T_{\min} 11 estaciones meteorológicas diaria

Período: 1950 - 2014

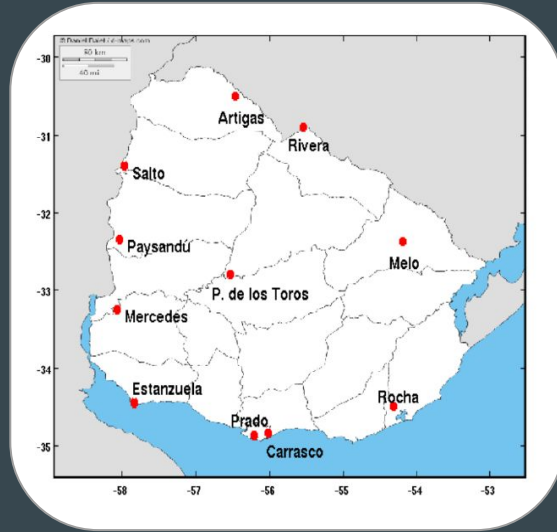


Datos para patrones de circulación y océanos

Datos para definir HAG

T_{\min} 11 estaciones meteorológicas diaria

Período: 1950 - 2014



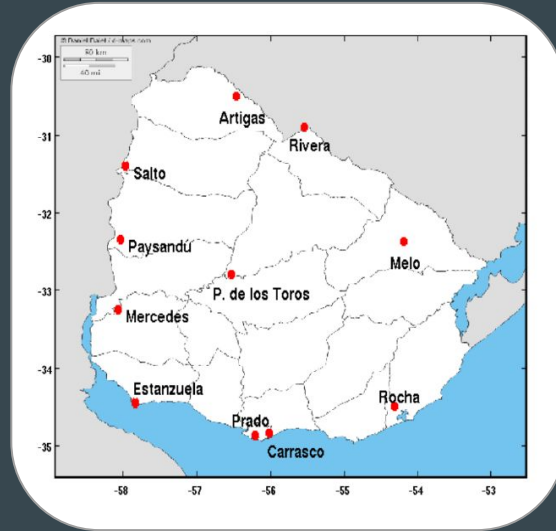
□ HAG

Datos para patrones de circulación y océanos

Datos para definir HAG

T_{\min} 11 estaciones meteorológicas diaria

Período: 1950 - 2014



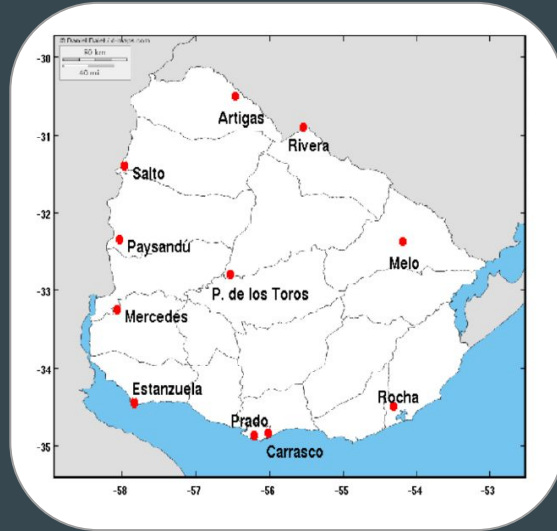
- HAG
- Índice: Cantidad de HAG por mes

Datos para patrones de circulación y océanos

Datos para definir HAG

T_{\min} 11 estaciones meteorológicas diaria

Período: 1950 - 2014



□ HAG □ Índice: Cantidad de HAG por mes

Datos para patrones de circulación y océanos

Datos mensuales de reanálisis ERA-interim 1979 - 2014

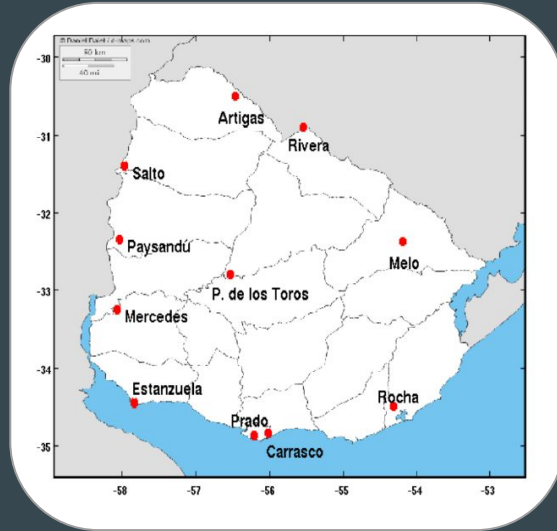
Anomalía de:

- SLP (Presión en superficie)
- OLR (Radiación de onda larga)
- SST (Temperatura del mar)

Datos para definir HAG

T_{\min} 11 estaciones meteorológicas diaria

Período: 1950 - 2014



□ HAG □ Índice: Cantidad de HAG por mes

Datos para patrones de circulación y océanos

Datos mensuales de reanálisis ERA-interim 1979 - 2014

Anomalía de:

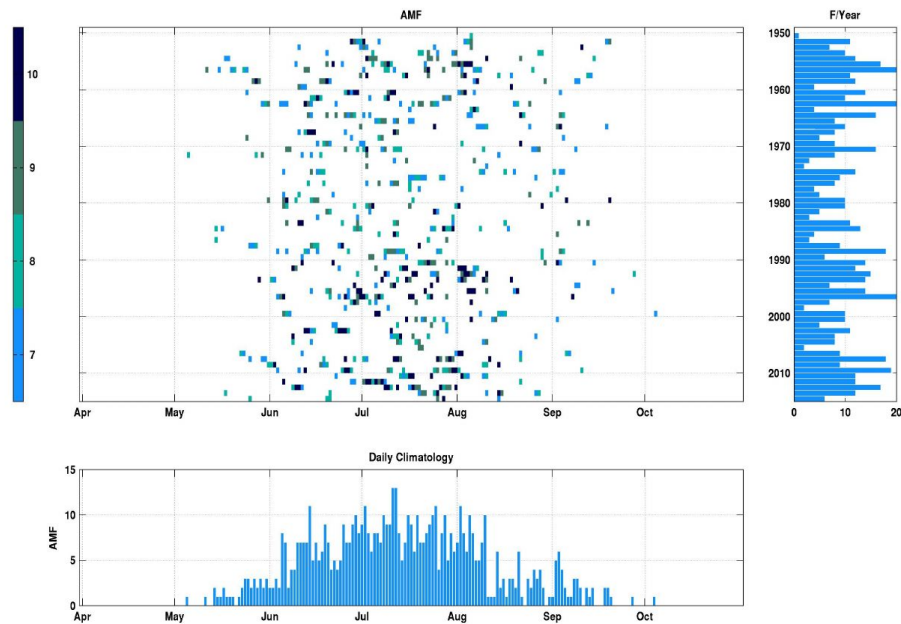
- SLP (Presión en superficie)
- OLR (Radiación de onda larga)
- SST (Temperatura del mar)



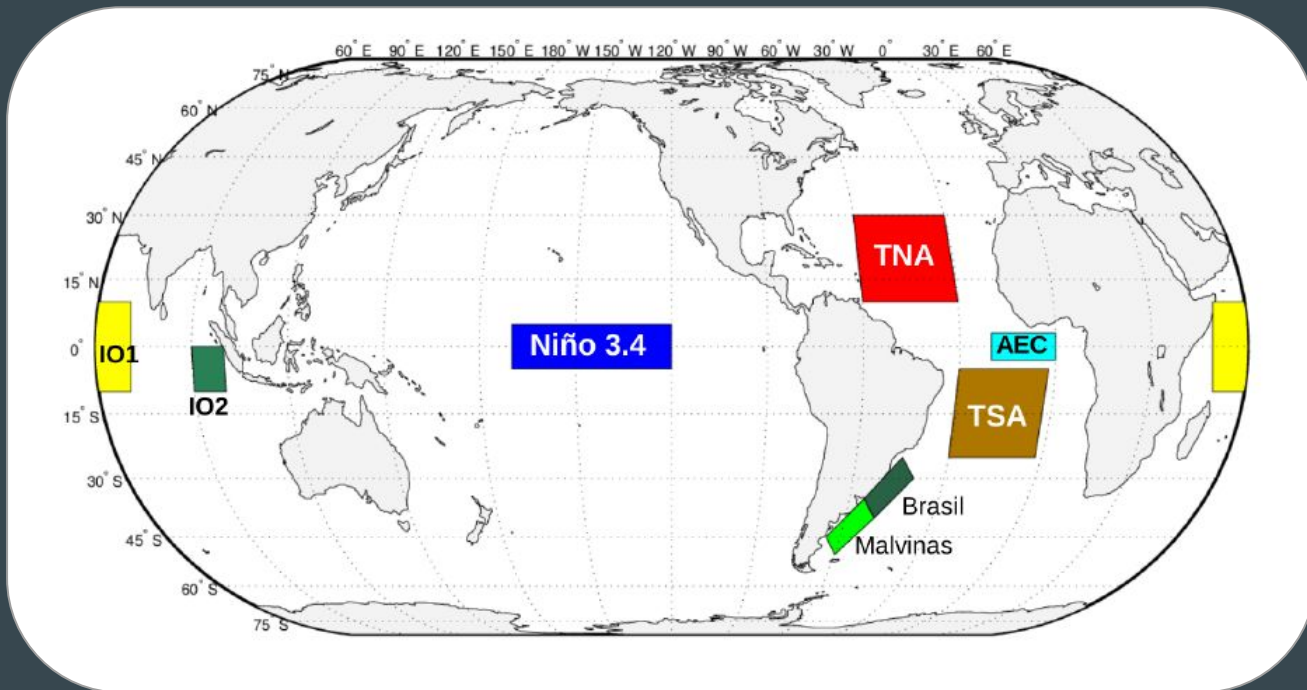
Índices oceánicos: Niño 3.4, ...

Índices atmosféricos: SAM, SOI y MJO

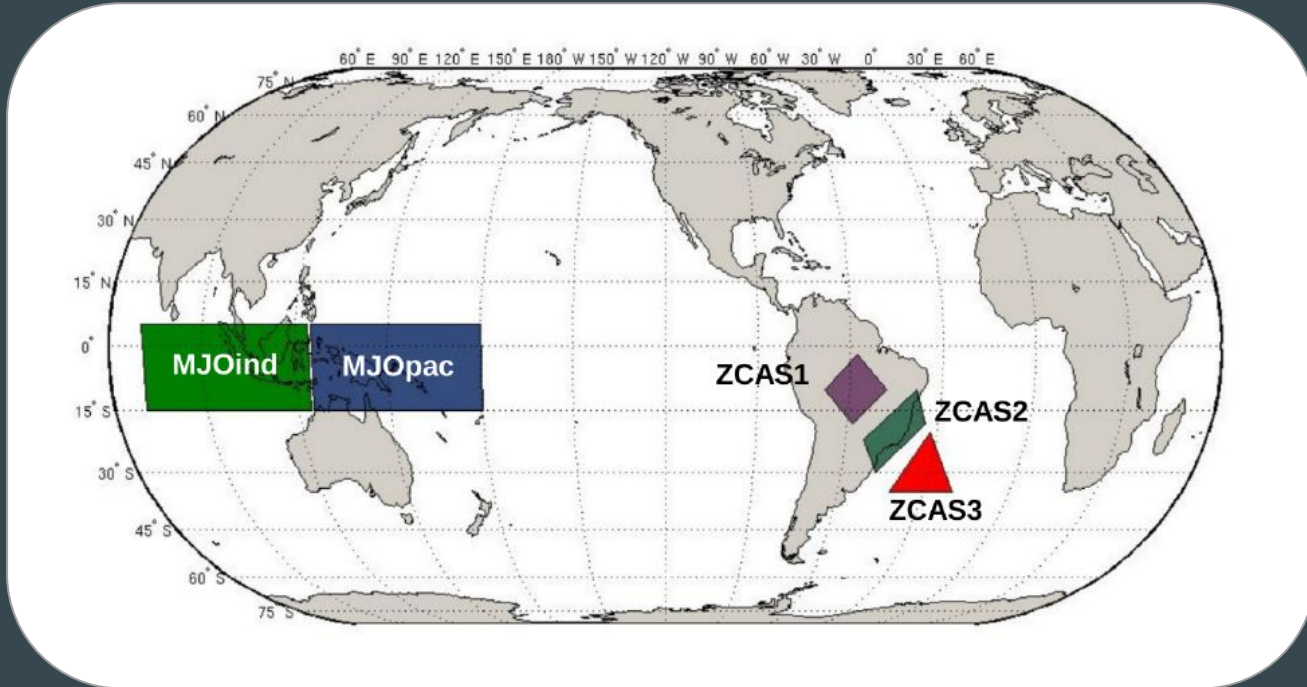
Distribución de HAG



Índices oceánicos

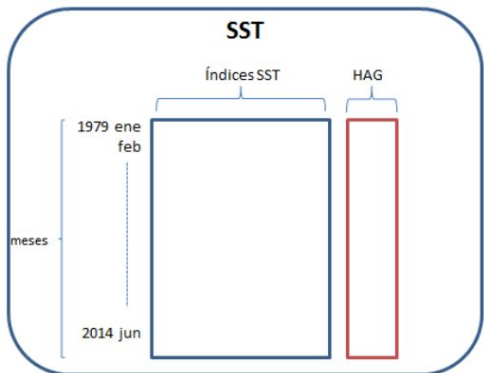


Índices atmosféricos



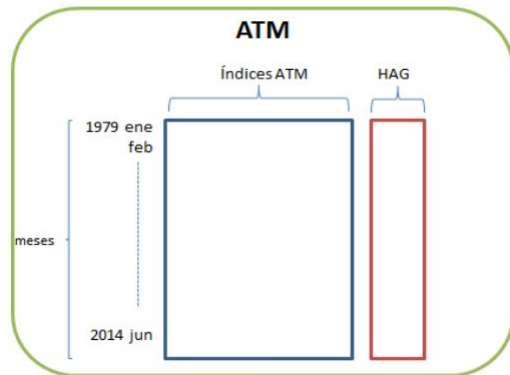
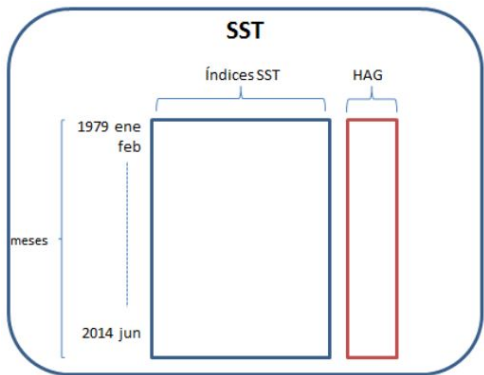
+ SAM + SOI (presión en superficie)

Bases de datos:



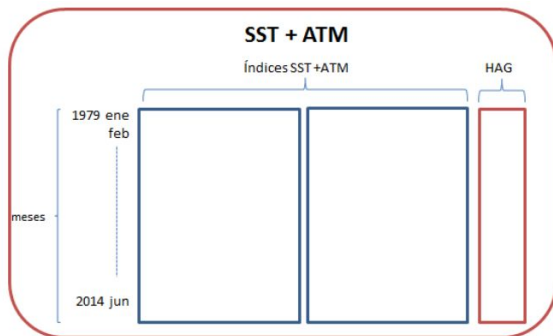
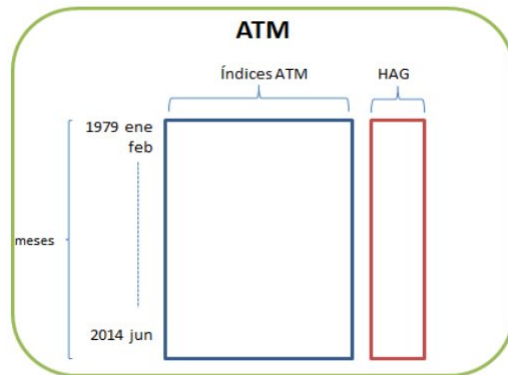
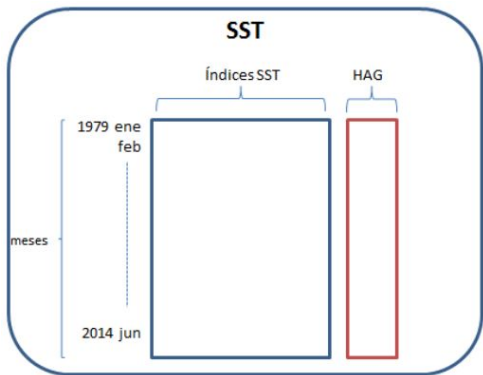
SST: $(7+1) \times 479$

Bases de datos:



SST: $(7+1) \times 479$

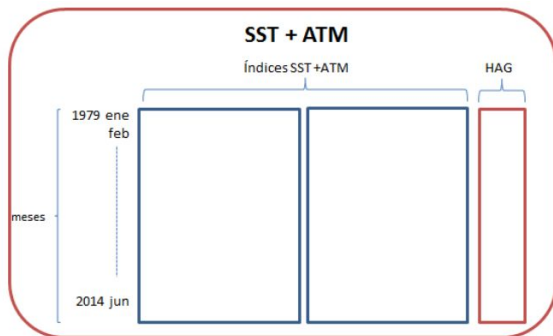
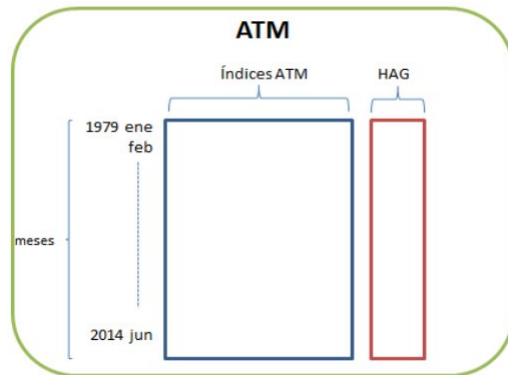
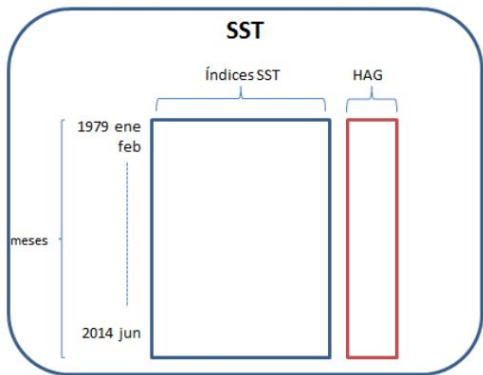
Bases de datos:



SST: (7+1) x 432

**SST + ATM:
(14+1) x 432**

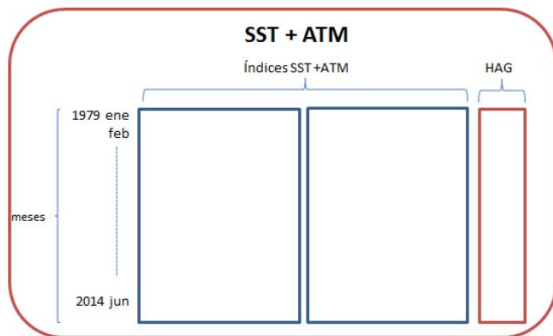
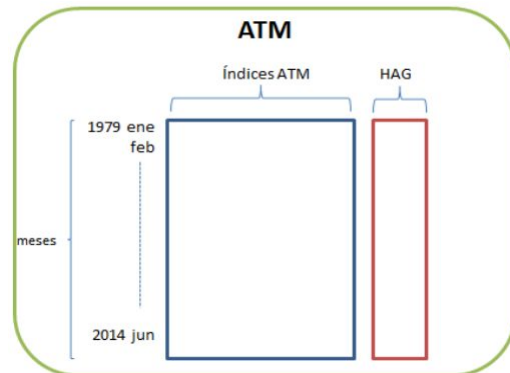
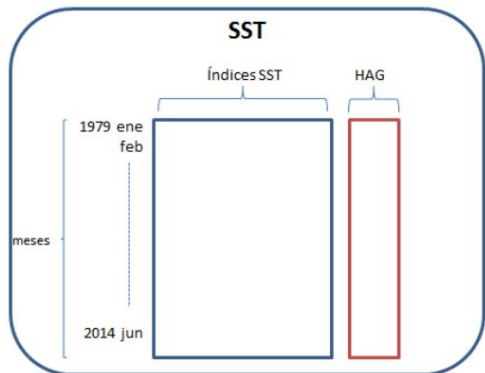
Bases de datos:



SST: (7+1) x 432

**SST + ATM:
(14+1) x 432**

Bases de datos:



SST: (7+1) x 432

**SST + ATM:
(14+1) x 432**

(aprox 1.100.000)

Metodología:

3. Datos y metodología

Metodología:

1. Reducción de las variables explicativas por ACP

ACP a SST y ATM tomando los CP que explican más del 50 % de la varianza.
(Se verifican los supuestos previos: test Bartlett e índice KMO)

Metodología:

1. Reducción de las variables explicativas por ACP

ACP a SST y ATM tomando los CP que explican más del 50 % de la varianza.
(Se verifican los supuestos previos: test Bartlett e índice KMO)

Metodología:

2. Inferir los índices más relevantes para las HAG

Modelos utilizados: LM, GLM , GAM y CART a las 4 bases de datos.

Se verifican los supuestos (gráfica y test de Shapiro) y analiza la bondad del ajuste mediante: R^2 , R^2 ajustado, P-valor de las variables, y F global, Xerror , ECM y SD.

Se seleccionan los mejores modelos por stepAIC y criterio CP óptimo 1-SE.

3. Predictibilidad de las HAG

Modelos: GLM, GAM, CART y RF.

Se compara por muestra de prueba (3/4 y 1/4) en 50 iteraciones y se comparan los modelos por ECM y SD.

Se pronostica (1) la cantidad de heladas a ocurrir en un mes; (2) por clases (p75 y p25)

Metodología:

1° se aplican las metodologías 1, 2 y 3 para todos los meses

Metodología:

1° se aplican las metodologías 1, 2 y 3 para todos los meses

2° se repite la metodología 1, 2 y 3 para el período Jun-Jul-Ago

Metodología:

1° se aplican las metodologías 1, 2 y 3 para todos los meses

2° se repite la metodología 1, 2 y 3 para el período Jun-Jul-Ago

Contenido

1. Introducción
2. Objetivos
3. Datos y metodología
- 4. Resultados**
5. Conclusiones

Resultados (Jun-Jul-Ago)

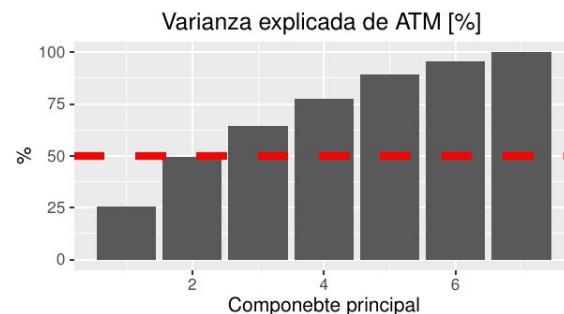
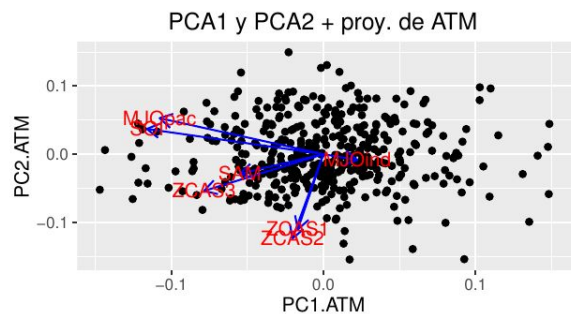
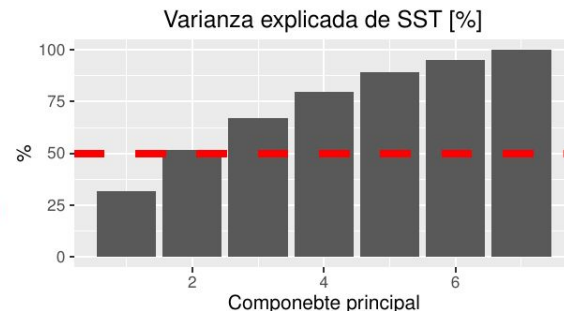
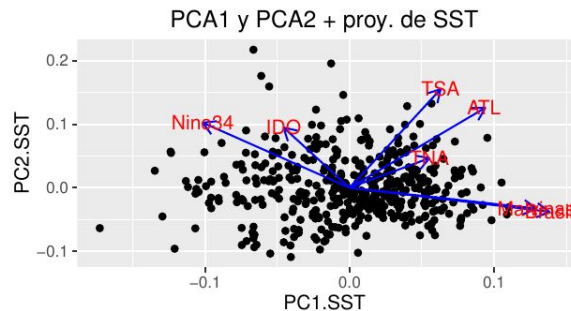
1. Reducción de variables ACP

Requisitos previos: análisis de la colinealidad → Barlett y KMO “acceptables”

Utilizando matriz de correlación y covarianza → muy similares → bien representada la variabilidad y dispersión de las series.

Resultados (Jun-Jul-Ago)

1. Reducción de variables ACP



Resultados (Jun-Jul-Ago)

1. Reducción de variables ACP

Pasamos de 14 variables a 4, representando más del 50% de la variabilidad.

Resultados (Jun-Jul-Ago)

1. Reducción de variables ACP

Pasamos de 14 variables a 4, representando más del 50% de la variabilidad.

La contribución de los índices en los CP no es clara, destacando que el índice Brasil - Malvinas contribuye de gran manera al PC1 y el Niño 3.4 con la segunda componente (consistente con Muller, 2009).

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
LM			
GLM			
GAM			
CART			

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
LM	No cumple con los supuestos ...		
GLM			
GAM			
CART			

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
LM	No cumple con los supuestos ...		
GLM	TSA, ATL, Malvinas, Brasil	ZCAS1, MJOind, MJOpac	TSA, Malvinas, Brasil, ZCAS1, MJOind, MJOpac
GAM			
CART			

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
LM	No cumple con los supuestos ...		
GLM	TSA, ATL, Malvinas, Brasil	ZCAS1, MJOind, MJOpac	TSA, Malvinas, Brasil, ZCAS1, MJOind, MJOpac
GAM	Niño 3.4, Malvinas, Brasil	ZCAS1, ZCAS3, MJOpac, MJOind	IDO, Brasil, Malvinas y ZCAS2
CART			

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
LM	No cumple con los supuestos ...		
GLM	TSA, ATL, Malvinas, Brasil	ZCAS1, MJOind, MJOpac	TSA, Malvinas, Brasil, ZCAS1, MJOind, MJOpac
GAM	Niño 3.4, Malvinas, Brasil	ZCAS1, ZCAS3, MJOpac, MJOind	IDO, Brasil, Malvinas y ZCAS2
CART	criterio 1-SE → flexibiliza umbral → Inestable...		

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
GLM	TSA, ATL, <u>Malvinas, Brasil</u>	ZCAS1, MJOind, MJOpac	TSA, <u>Malvinas, Brasil</u> , ZCAS1...
GAM	Niño 3.4, <u>Malvinas, Brasil</u>	ZCAS1, ZCAS3, MJOind	IDO, <u>Brasil, Malvinas</u> , ZCAS2...

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
GLM	TSA, ATL, <u>Malvinas, Brasil</u>	ZCAS1, MJOind, MJOpac	TSA, <u>Malvinas, Brasil</u> , ZCAS1...
GAM	Niño 3.4, <u>Malvinas, Brasil</u>	ZCAS1, ZCAS3, MJOind	IDO, <u>Brasil, Malvinas</u> , ZCAS2...

Se comparan por ANOVA

Resultados (Jun-Jul-Ago)

2. Inferencia (Más relevante → Menos Relevante)

Modelo	SST	ATM	ATM + SST
GLM	TSA, ATL, <u>Malvinas, Brasil</u>	ZCAS1, MJOind, MJOpac	TSA, <u>Malvinas, Brasil</u> , ZCAS1...
GAM	Niño 3.4, <u>Malvinas, Brasil</u>	ZCAS1, ZCAS3, MJOind	IDO, <u>Brasil, Malvinas</u> , ZCAS2...

Se comparan por ANOVA → Mejor ATM + SST

Resultados 3.1 Predictibilidad de la cantidad de HAG

Error medio y SD

	SST+ATM		ATM		SST	
	Media	STD	Media	STD	Media	STD
GLM	3.3	0.4	3.3	0.6	3.4	0.6
GAM	7.7	8.0	3.1	0.5	7.2	7.2
CART	4.1	0.49	4.0	0.6	4.1	0.6
RF	2.9	0.3	3.0	0.4	2.9	0.4

Resultados 3.1 Predictibilidad de la cantidad de HAG

Error medio y SD

	SST+ATM		ATM		SST	
	Media	STD	Media	STD	Media	STD
GLM	3.3	0.4	3.3	0.6	3.4	0.6
GAM	7.7	8.0	3.1	0.5	7.2	7.2
CART	4.1	0.49	4.0	0.6	4.1	0.6
RF	2.9	0.3	3.0	0.4	2.9	0.4

Valor medio 3.1 HAG/mes

Resultados 3.1 Predictibilidad de la cantidad de HAG

Error medio y SD

	SST+ATM		ATM		SST	
	Media	STD	Media	STD	Media	STD
GLM	3.3	0.4	3.3	0.6	3.4	0.6
GAM	7.7	8.0	3.1	0.5	7.2	7.2
CART	4.1	0.49	4.0	0.6	4.1	0.6
RF	2.9	0.3	3.0	0.4	2.9	0.4

Valor medio 3.1 HAG/mes → Mejor la climatología

Resultados 3.1 Predictibilidad por encima del **P75** de HAG

Tabla de contingencia → razón de casos mal pronosticados

Resultados 3.1 Predictibilidad por encima del P75 de HAG

Tabla de contingencia → razón de casos mal pronosticados

	SST+ATM		ATM		SST	
	Media	STD	Media	STD	Media	STD
GLM	0.21	0.05	0.18	0.3	0.24	0.03
GAM	0.13	0.03	0.15	0.04	0.20	0.04
CART	0.32	0.09	0.27	0.07	0.31	0.09
RF	0.23	0.03	0.24	0.04	0.24	0.03

Resultados 3.1 Predictibilidad por encima del P75 de HAG

Tabla de contingencia → razón de casos mal pronosticados

	SST+ATM		ATM		SST	
	Media	STD	Media	STD	Media	STD
GLM	0.21	0.05	0.18	0.3	0.24	0.03
GAM	0.13	0.03	0.15	0.04	0.20	0.04
CART	0.32	0.09	0.27	0.07	0.31	0.09
RF	0.23	0.03	0.24	0.04	0.24	0.03

Resultados 3.1 Predictibilidad por debajo del P25 de HAG

Tabla de contingencia → razón de casos mal pronosticados

	SST+ATM		ATM		SST	
	Media	STD	Media	STD	Media	STD
<u>GLM</u>	0.37	0.06	0.40	0.01	0.20	0.02
<u>GAM</u>	10.8	0.04	0.30	0.05	0.17	0.03
<u>CART</u>	0.43	0.09	0.48	0.08	0.29	0.08
<u>RF</u>	0.40	0.04	0.40	0.06	0.23	0.03

Resultados 3.1 Predictibilidad por debajo del P25 de HAG

Tabla de contingencia → razón de casos mal pronosticados

	SST+ATM		ATM		SST	
	Media	STD	Media	STD	Media	STD
GLM	0.37	0.06	0.40	0.01	0.20	0.02
GAM	10.8	0.04	0.30	0.05	0.17	0.03
CART	0.43	0.09	0.48	0.08	0.29	0.08
RF	0.40	0.04	0.40	0.06	0.23	0.03

Contenido

1. Introducción
2. Objetivos
3. Datos y metodología
4. Resultados
5. **Conclusiones**

Principales conclusiones:

- Relevancia de las corrientes Brasil y Malvinas en la ocurrencia de meses con HAG
- Mejoran levemente los modelos al considerar variables oceánicas y atmosféricas, dando la pauta de que la interacción océano atmósfera es relevante en la ocurrencia de HAG
- Los modelos son malos para pronosticar la cantidad de HAG
- Los modelos son muy buenos para pronosticar por clase las HAG

Trabajos a futuro:

- **Agregar nuevos modelos: redes neuronales, etc..**
- **Explorar otros índices**
- **Pronosticar por más de dos clases**

¡Gracias!

demello.santiago@gmail.com