

Herramientas para la gestión de la información meteorológica y continuación del trabajo de la “base espejo”

Jose Luis Araya Lopez

Departamento de Redes y Procesamiento de Datos

Instituto Meteorologico Nacional

20 de Marzo del 2019

La necesidad de este trabajo

Hay un consenso de que se requiere de un método sistemático de homogeneización y rellenado. Estos métodos de homogenización consisten en eliminar tendencias superfluas en la serie, las cuales dificultan la interpretación de los resultados y conducen a malas conclusiones sobre la evolución de la media y la varianza de los procesos.

Utilidad de este trabajo

- Los datos se pueden vender como PRODUCTOS.
- Mejoramiento íntegro de la información que se ofrece.
- Se puede satisfacer las necesidades de ciertos usuarios internos y dar opciones para mejorar la calidad del servicio que se brinda.

Lo que se quiere hacer

- 1) Definir con precisión las necesidades de usuarios internos y externos en términos de productos derivados a partir de los datos.
- 2) Extraer los datos de la base de datos (con niveles de control de calidad iniciales ya sobrepasados).
- 3) Generar matriz de datos.
- 4) Aplicación de técnicas de agrupamiento sobre los datos. Selección de grupos de estaciones y criterios asociados.
- 5) Control de calidad espacial.
- 6) Homogeneización de grupos.
- 7) Relleno de datos.
- 8) Generación de productos a partir de esta información (gráficos, mapas, etc).

Objetivo

Dotar a usuarios internos y externos de productos relevantes. En particular:

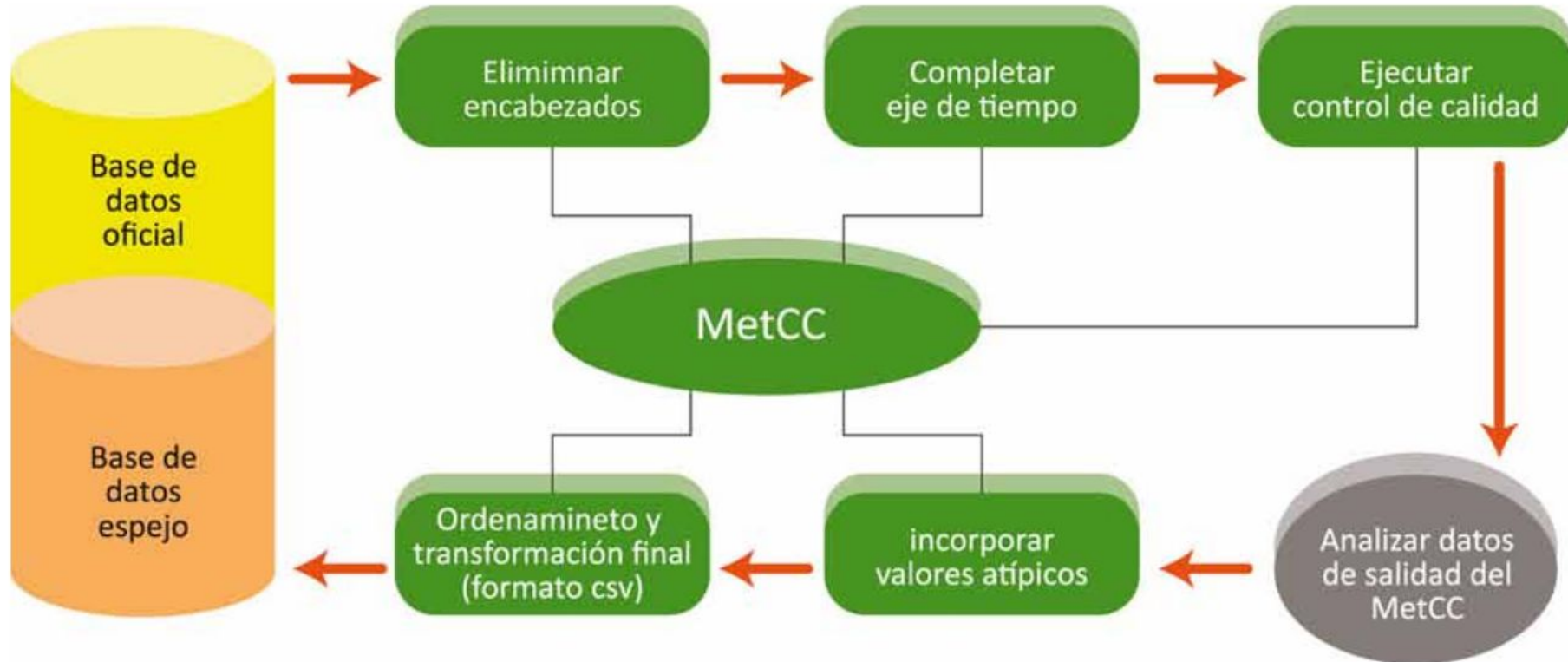
- Series de datos rellenas
- Series de datos homogeneizadas
- Otros productos según necesidades futuras

Propuesta

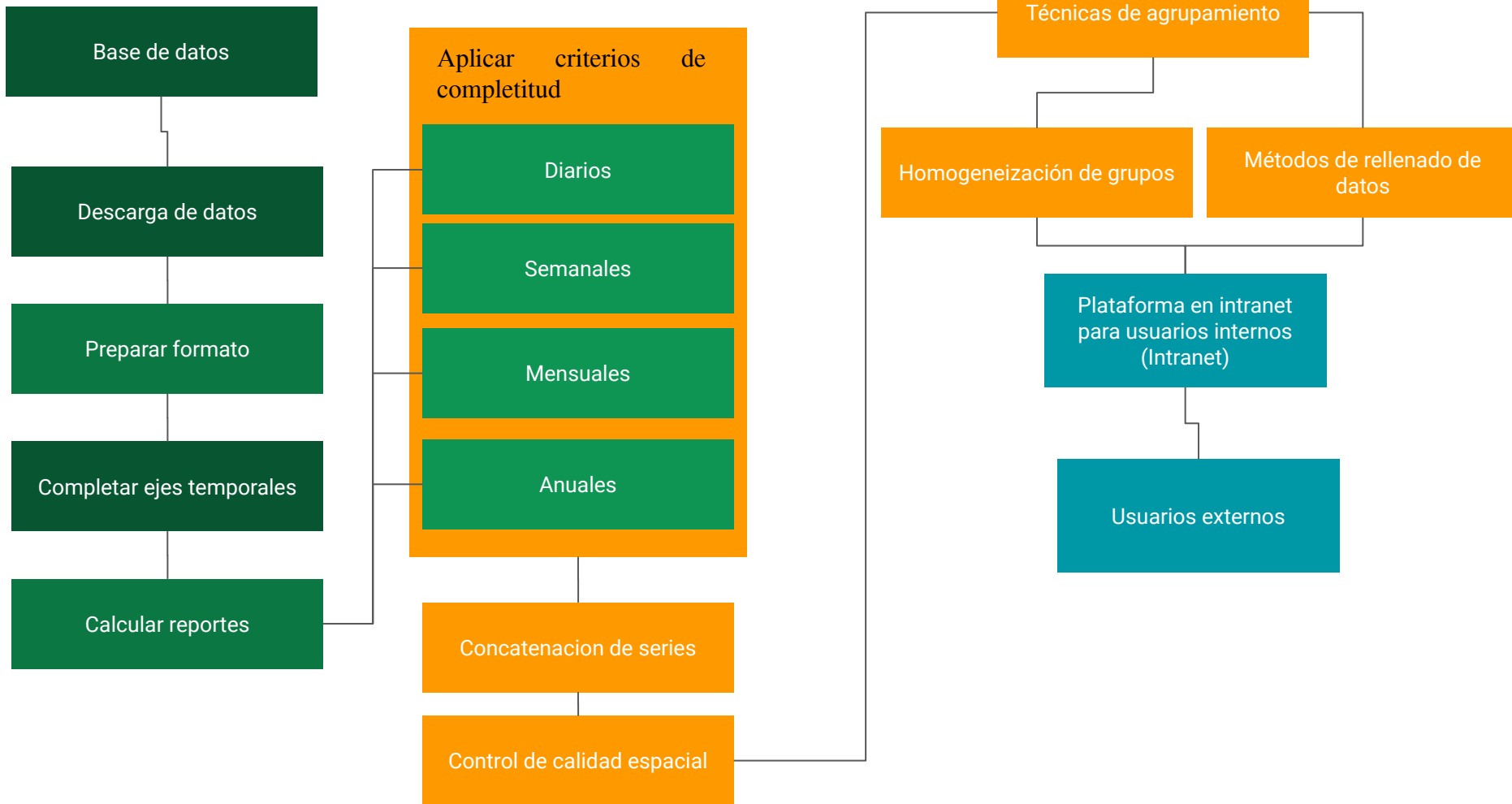
Desarrollo de una herramienta (software o programa) que permita realizar el relleno y homogenización de los datos almacenados en la base oficial del IMN, y sistematizar la generación regular de ciertos productos



Replantando la idea de la “base espejo”



Generalidades de la aplicación



Tecnologías a utilizar



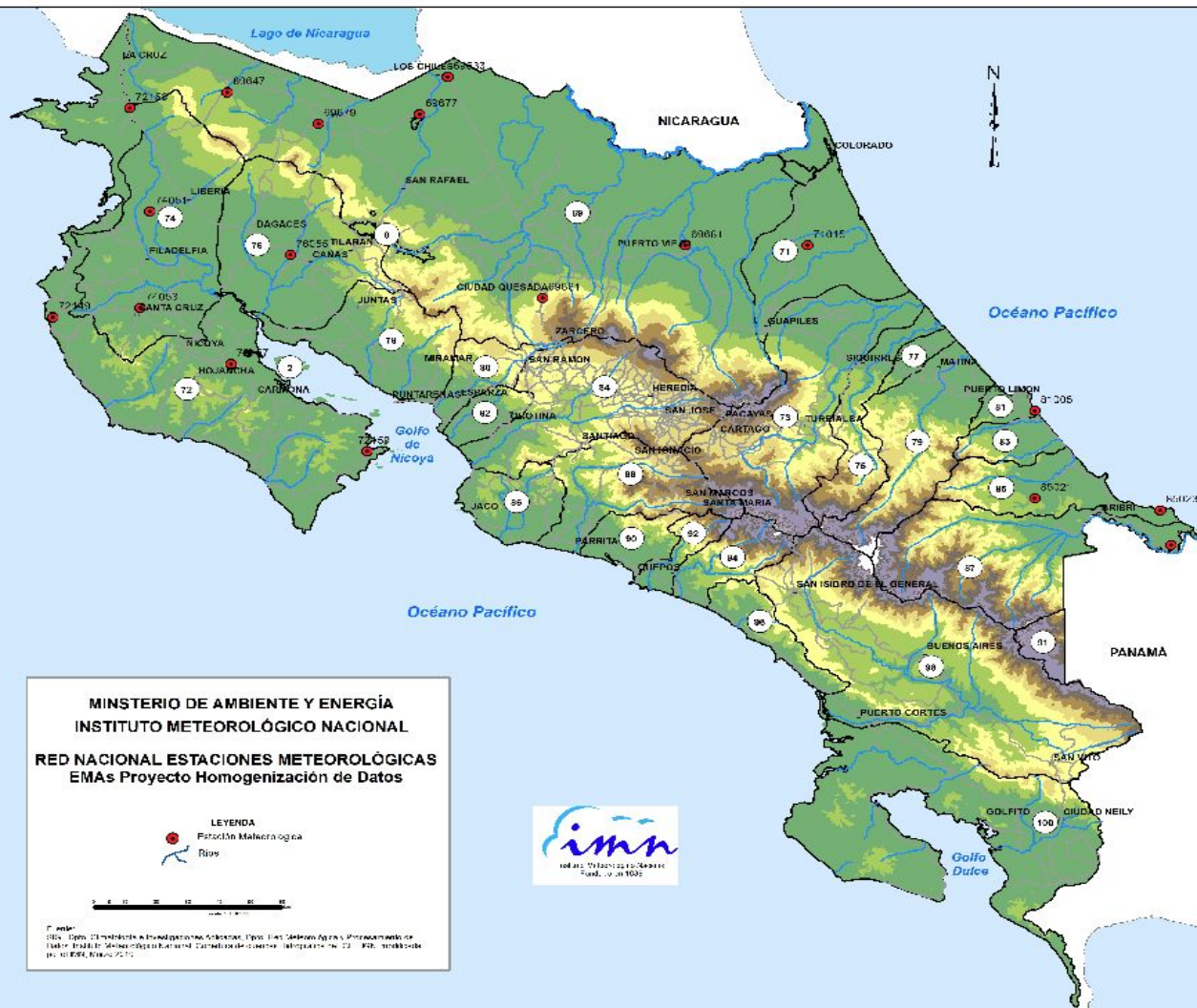
- Lenguajes de programación
- Modelos
- Reanalysis
- Técnicas de aprendizaje de máquina



Estado actual de este trabajo

- Se ha implementado una versión del algoritmo ePCA para datos multivariados en las 18 series temporales seleccionadas para el entrenamiento.
- Se ha implementado una versión de un algoritmo de agrupamiento para selección semi-automática de una cantidad dada de un número de series temporales. Se debe profundizar más en este aspecto.
- Se está evaluando cómo funciona el algoritmo para diferentes resoluciones temporales de estos datos.

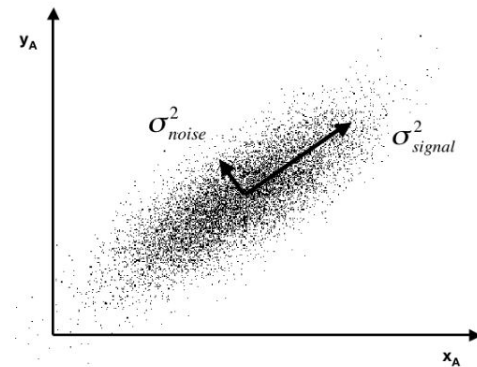
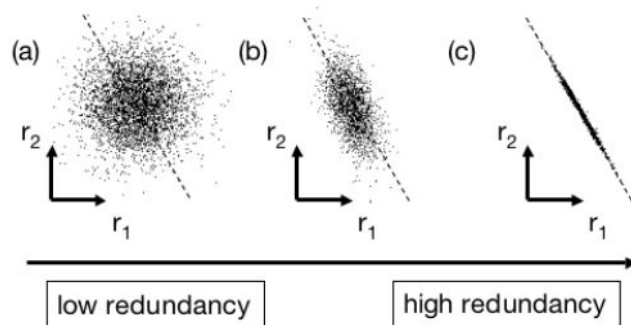
Veamos algunos ejemplos preliminares....



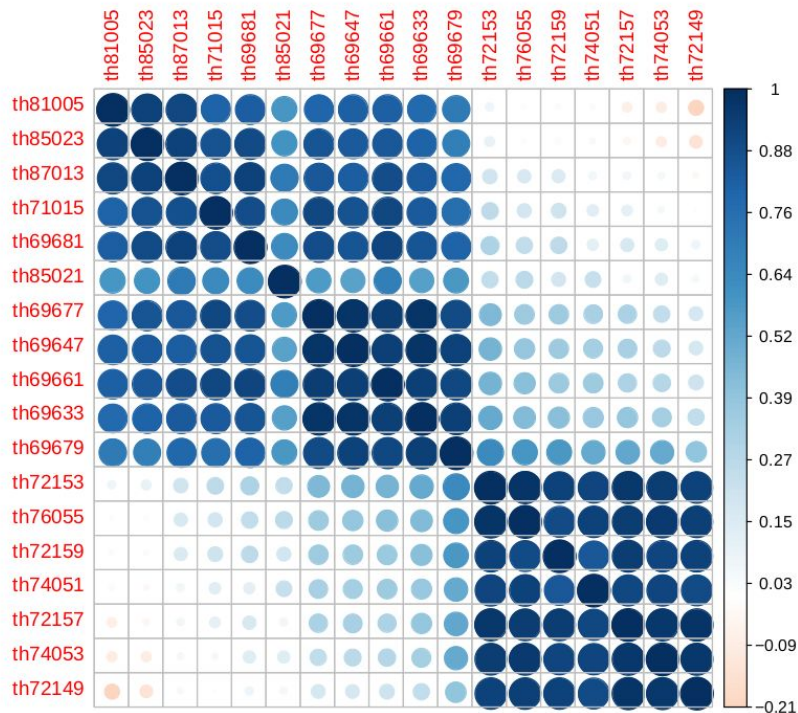
NÚMERO	ESTACIÓN
69633	COMANDO LOS CHILES
69647	FINCA BRASILIA DEL ORO
69661	CIUDAD CUESADA
69677	LAGUNA, CAÑO NEGRO
69679	UPALA
69681	LA REUSCA
71015	CANTA GALLO
72149	SAN JOSE, PINILLA
72153	SANTA ELENA, LA CRUZ
72157	FINCA LA CEIBA
72159	PAQUERA
74051	AEROP. LIBERIA OESTE 07
74053	SANTA CRUZ
76055	HDA. MOJICA
81005	AEROP. LIMON
85021	HITCYCFRFRF
85023	MANZANILLO
87013	SIXAOLA

Componentes principales temporales espacio-temporales (ePCA)

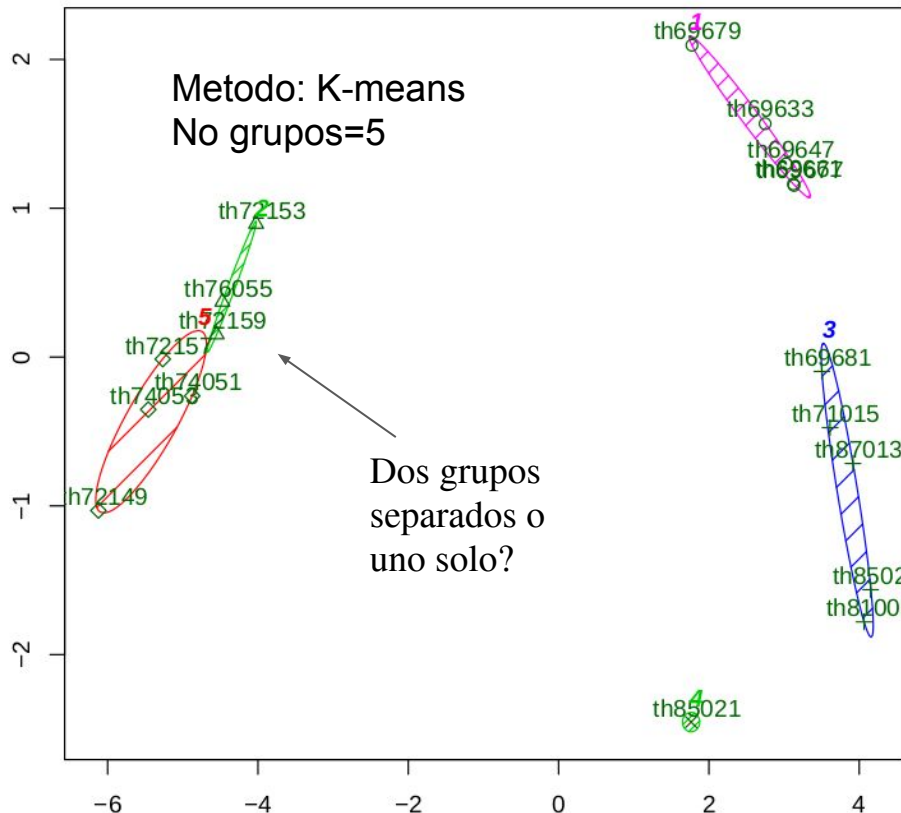
Este método es una extensión de PCA que permite codificar la información temporal y espacial simultáneamente.



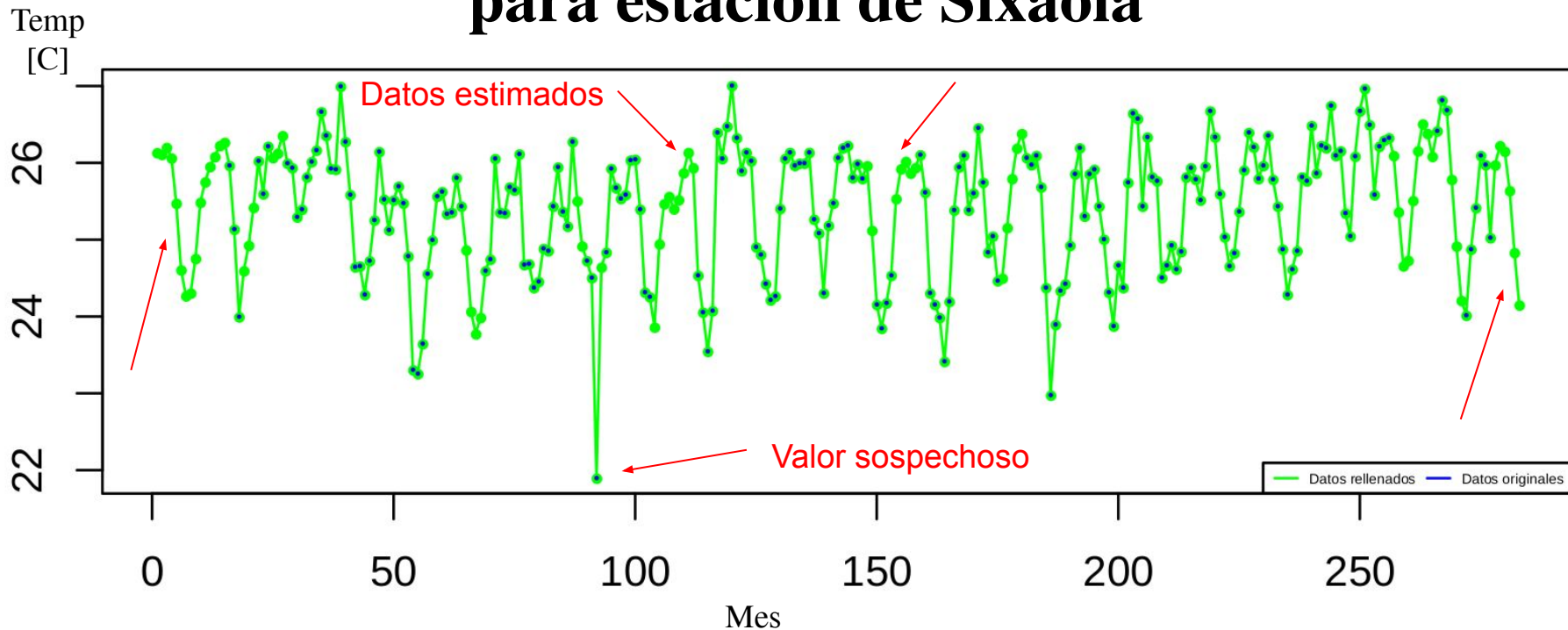
Probando un método de agrupamiento



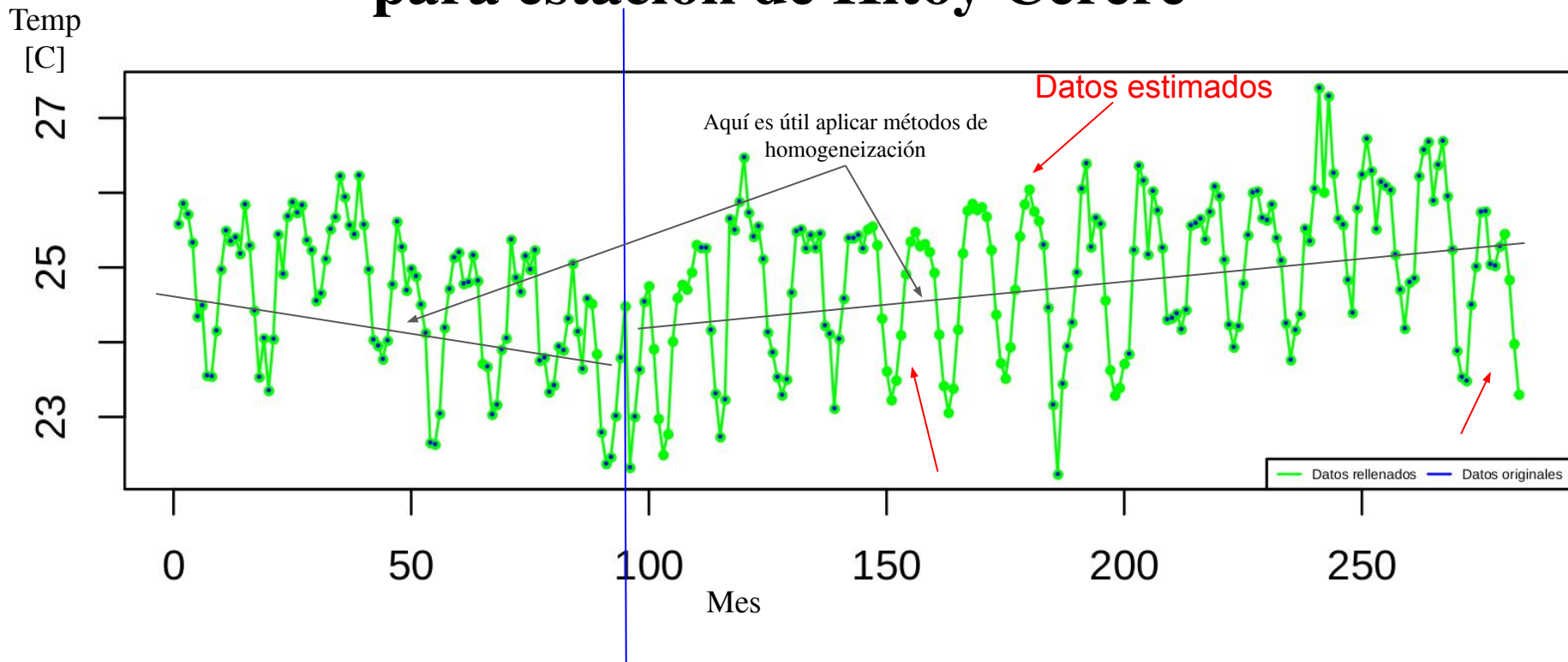
Matriz de correlación cruzada



Ejemplo de relleno de datos mensuales usando ePCA para estacion de Sixaola



Ejemplo de relleno de datos mensuales usando ePCA para estación de Hitoy Cerere



Que falta por hacer?

- Afinar esta metodología y aplicarla a otros parámetros meteorológicos, tomando en cuenta los resultados de la encuesta sobre necesidades de usuarios internos y externos.
- Tomar esta experiencia y organizar el código ya probado, darle mantenimiento y mejorarlo continuamente.
- Poner estos resultados a disposición de los usuarios a través de alguna plataforma tecnológica apropiada.
- Evaluar la utilidad de otras fuentes de información para el mejoramiento de estos métodos (aquí entra la idea de “datos híbridos”)

Gracias