



Universidad de la República - CSIC

Formulario de Informe final del Programa de Apoyo a la  
Investigación Estudiantil

Edición 2014



### DATOS DEL PROYECTO

- Título del Proyecto: Análisis de aglomeraciones, análisis espectral e interpretación de datos meteorológicos tri-variados de alta frecuencia
- Número ID del proyecto: 115
- Área de conocimiento: Científica
- Facultad o Servicio: Facultad de Ingeniería, UDELAR
- Nombre completo de los-as Integrantes del equipo: Vanessa Bentancur, Gianina Giacosa
- Correo electrónico del/de la estudiante referente: vane.bentancur@gmail.com
- Nombre completo del/de la docente orientador-a: Rafael Terra
- Correo electrónico del/de la docente orientador-a: rterra@fing.edu.uy

## INFORME FINAL

(desde ítem 1 a 7 la extensión máxima POR ÍTEM es de una carilla)

- Transcriba los objetivos del proyecto tal cual figuraban en la solicitud financiada.

Como objetivo general buscamos desarrollar técnicas estadísticas a partir de una base de datos, las cuales nos permitan interpretar meteorológicamente los datos en estudio. Dentro de esta interpretación, incluiríamos una comparación con estudios existentes para evaluar la consistencia de los resultados obtenidos, además de identificación de comportamientos, de alguna o varias variables, que se desvíen de lo esperado.

En particular, los datos a analizar provienen de una estación meteorológica en Buena Unión, localizada en el kilómetro 481 de la ruta 5, en el departamento de Rivera. Se trata de una estación perteneciente a UTE, la cual forma parte de un grupo de estaciones pertenecientes a una campaña, en conjunto con la UdelaR, cuyo fin era el determinar las zonas en el Uruguay con mayor potencial eólico. Los datos provenientes de la estación situada en Buena Unión corresponden a un período de 3 años y medio y tienen una resolución de 10 minutos, para las variables velocidad del viento, temperatura y radiación solar.

Si bien este set de datos fue generado con un objetivo específico, son muy valiosos desde el punto de vista meteorológico debido a la frecuencia de los mismos. Sin embargo, deberían estudiarse primero las condiciones de la obtención de dichos datos, ya que existen normas mundiales que implican un procedimiento riguroso en el momento de realizar mediciones meteorológicas. Por lo tanto, como primer objetivo específico, se encuentra el evaluar la calidad de los datos, en un primer contacto con los mismos.

En cuanto a los objetivos más específicos, se incluye el aprendizaje de los métodos estadísticos a utilizar. En particular, las técnicas utilizadas serían "Análisis de Aglomeraciones" o "Cluster Analysis", por un lado, y "Análisis espectrales", por otro. En ambos casos, el estudio se realizará primero con un enfoque mono-variado, donde se analizará el comportamiento de cada una de las variables por separado, y luego multivariado, donde se combinarán las tres variables, velocidad del viento, temperatura y radiación solar.

Respecto a la estudio de la primera técnica, cuya finalidad es dividir un conjuntos de objetos en un grupos (clusters) de forma que los perfiles de los objetos en un mismo grupo sean muy similares entre sí y los de los objetos de clusters diferentes sean bien distintos; se incluye el conocer sus variantes (métodos jerárquicos, métodos no jerárquicos).

En segunda instancia, se realizará el análisis espectral, técnica que permite descomponer una señal o serie de tiempo en componentes senoidales de diferentes frecuencias y amplitudes. Buscaremos identificar frecuencias e interpretar estos valores.

Además de conocer nuevos métodos estadísticos, ampliando así nuestros conocimientos en lo que respecta a las bases de datos, apuntamos a reforzar nuestra formación académica a través del análisis y de las lecturas a realizar. Como objetivo personal, esto incluiría fundamentalmente aplicar los conocimientos teóricos frente a situaciones reales y aprender a interpretar los resultados obtenidos.

- Enumere y describa las principales actividades desarrolladas en el marco de su proyecto.

Se realizó un análisis de aglomeraciones monovariado para la temperatura y la intensidad del viento en dos estaciones de Uruguay: Colonia Rubio y José Ignacio.

Las principales tareas realizadas en el marco del proyecto se pueden dividir en dos etapas: la primera es la preparación de los datos a ser estudiados mediante el análisis de aglomeraciones y la segunda es la realización del mismo junto con análisis e interpretación de resultados obtenidos.

Dentro de la primera etapa se encuentran las siguientes actividades:

- 1) Se verificó que el comportamiento de los datos sea de la forma esperada mediante cálculos de ciclos diarios y finalmente se los pasó a datos horarios.
- 2) En el caso de la temperatura se obtuvo el ciclo anual para obtener las anomalías con respecto al mismo, ya que se probó realizar el análisis cluster con los datos originales y se observó una gran influencia de la estacionalidad en los resultados. La intensidad de viento no demostró mucha variación y se trabajó con los datos originales.
- 3) Para realizar el análisis se eligió un método no jerárquico. Inicialmente se probó realizar un método llamado “k-means”, al cual se le debe asignar un número de semillas iniciales de valores aleatorios, pero se comprobó que los resultados no siempre eran los mismos y por ello no se los podía interpretar.

Dentro de la segunda etapa se encuentran las siguientes actividades:

- 4) Se decidió realizar un método llamado “Nucleated Agglomerative Clustering”, una mezcla entre el “k-means” y un método aglomerativo. La razón de la utilización del mismo es que los resultados fueron más robustos, es decir, realizando varias pruebas se llegaba siempre a lo mismo.
- 5) Una vez obtenidos los resultados y las semillas finales de cada cluster se las graficó para ver el comportamiento de cada ciclo y se realizaron cálculos de media, amplitud y tendencia.
- 6) Para poder visualizar mejor la distribución de los días en cada grupo o cluster, se asignaron los días en una cuadrícula con colores indicando cada día el color del grupo correspondiente. Luego se realizó un histograma de rachas, es decir donde se observa para cada cluster, cuántas veces hubo una cantidad  $x$  de días seguidos en el mismo.
- 7) Se calcularon probabilidades de transición, es decir probabilidad de que al pasar de un día a otro se pase de un cluster a otro (o a sí mismo), para observar cómo se relacionan los cluster entre ellos. Para tener una mayor seguridad de los resultados, se utilizó la función binomial acumulada, para ver la significancia estadística y luego se realizaron grafos en base a estos resultados. Estos cálculos se realizaron por semestre y anualmente.
- 8) Finalmente se hizo un cruce de variables, donde se intenta buscar coincidencias entre grupos de intensidad de viento y de temperatura, comparando día a día cuántas veces coinciden cada cluster de una variable con la otra. Se utilizó nuevamente la función binomial acumulada para corroborar la significancia estadística y luego se realizaron grafos a partir de los resultados. Estos cálculos también fueron realizados por semestre y anualmente.
  - Indique si se han efectuado todas las etapas planteadas en el cronograma de ejecución del proyecto. En caso de que su cronograma haya sufrido alteraciones o no se haya podido cumplir con todas las etapas definidas en el cronograma, aclare los motivos de tal situación.

En cuanto a las etapas del cronograma planteadas, en general se pudieron efectuar. Sin embargo, solo se realizaron para una de las técnicas que fue el análisis de aglomeraciones con enfoque mono-variado.

Como se indicó en la parte anterior se realizó la observación y calidad de los datos y se realizaron lecturas sobre el análisis a realizar ya que no teníamos total conocimiento

del mismo. En cuanto a esta parte nos llevó un poco más de tiempo del esperado debido a la falta de experiencia que teníamos con dichas técnicas y siendo prioritario el adecuado entendimiento de las mismas. Como al inicio del proyecto contábamos con datos de varias estaciones redefinimos nuestro set de datos a estudiar, para simplificar el estudio y poder realizar comparaciones del comportamiento de variables medidas en distintos puntos del país. Elegimos dos estaciones, Colonia Rubio y José Ignacio, y dos variables, temperatura e intensidad de viento.

Luego se comenzó con el análisis, se continuó y finalizó aproximadamente como dice el cronograma. Pero al finalizar con el mismo, debido a las diferentes interpretaciones que se podía obtener, se decidió profundizar más en éste, y realizar más cálculos a partir de los resultados, ya que no se podían interpretar de forma segura. Es así que se priorizó la interpretación física de los resultados para lograr una lectura final unificada de los mismos, con análisis originalmente no previstos, respecto a iniciar análisis con otras técnicas (p. e. análisis espectral) y otras variables (radiación) con objetivos diferentes. A futuro se realizará la etapa de difusión del proyecto.

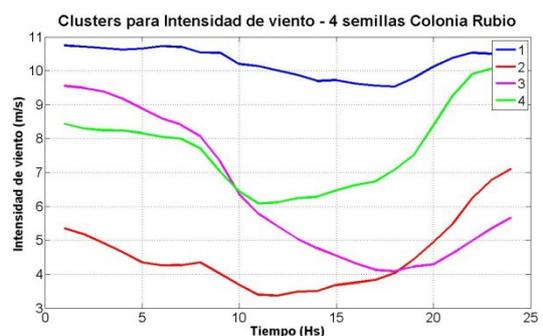
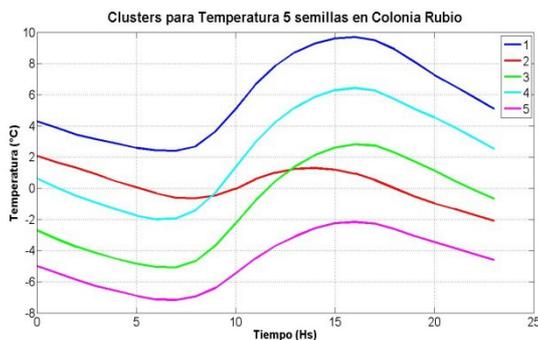
- Indique los principales resultados obtenidos. Aclare hasta qué punto coinciden - o no - con los resultados esperados por parte del equipo.

Para cada estación (Colonia Rubio y José Ignacio) y para cada variable (temperatura e intensidad de viento) se obtuvieron los siguientes resultados:

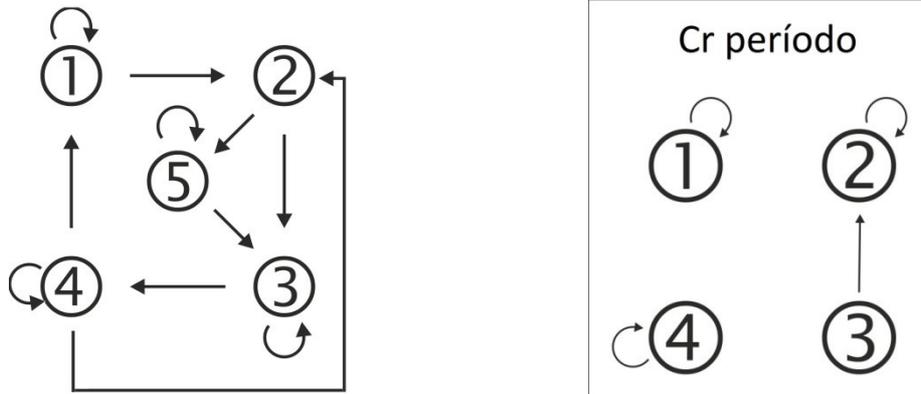
- Caracterización de los regímenes diurnos típicos de cada variable. Además del análisis de aglomeración en sí, se presenta frecuencia, amplitud, tendencia, histograma de rachas, distribución de los clusters en el período estudiado. Todos estos cálculos fueron realizados tanto para todo el período de datos, como separando los datos en semestres frío y cálido.
- Análisis de transiciones entre regímenes de cada variable. A partir de las probabilidades de transición empíricas y, analizando la significancia estadística en comparación con la distribución binomial, se realizaron transiciones preferidas, las cuales se resumieron en grafos.
- Caracterización unificada de ambas variables e interpretación física. Se realizó un cruce entre los regímenes diurnos de cada variable donde se halla la cantidad de días coincidentes entre los clusters de cada variable. A partir de esto se volvió a utilizar la función binomial acumulada con el criterio ya mencionado para determinar regímenes preferentemente asociados. En base a dicho análisis se intenta caracterizar la situación y evolución sinóptica.

Se muestran a continuación algunos de estos resultados (por motivos de espacio solo para Colonia Rubio):

#### Gráficos de clusters de viento y temperatura:



Grafos de transiciones preferidas entre clusters:



Probabilidades de transición para las variables temperatura e intensidad de viento:

T	1	2	3	4	5	Total
1	<b>0.53</b>	<b>0.28</b>	0.03	0.12	0.05	1
2	0.01	0.15	<b>0.40</b>	0.11	<b>0.32</b>	1
3	0.02	0.12	<b>0.41</b>	<b>0.33</b>	0.10	1
4	<b>0.23</b>	<b>0.22</b>	0.07	<b>0.45</b>	0.03	1
5	0	0.02	<b>0.37</b>	0.02	<b>0.58</b>	1
V	1	2	3	4	Total	
1	<b>0.41</b>	0.08	0.24	0.27	1	
2	0.15	<b>0.28</b>	0.18	0.38	1	
3	0.08	<b>0.41</b>	0.15	0.36	1	
4	0.26	0.10	0.23	<b>0.40</b>	1	

Para el caso de la temperatura en Colonia Rubio se ve la asimetría entre el calentamiento gradual (5→3 →4→1) y el enfriamiento abrupto (1 ó 4→2→3 ó 5). Todos los clusters tienen una probabilidad de recurrencia significativa de un día al siguiente, salvo el 2 que parece estar asociado al pasaje de un frente.

Para el caso de viento, no se llegó a una interpretación tan clara, ya que no se observa a simple vista un comportamiento patrón en el ciclo de viento. Es parte de la investigación pendiente de este trabajo el poder analizar más detalladamente esta variable.

- Indique si los resultados parciales o finales del proyecto fueron difundidos a través de alguna actividad (charlas, seminarios, talleres, prensa, edición de materiales impresos, etc.).

En el primer semestre del año 2015 cursamos la asignatura Análisis Estadísticos y Datos Climáticos, el cual tenía como parte de la aprobación un trabajo final con presentación oral a los demás estudiantes. Adaptamos parte del trabajo realizado para el proyecto, y presentamos los resultados que habíamos obtenido hasta el momento (julio).

- En caso de haber enfrentado dificultades en el desarrollo del proyecto de investigación, realice una breve descripción de las mismas.

En un principio nos habíamos planteado explorar dos técnicas estadísticas, el análisis de aglomeración y el análisis espectral. Para poder implementarlas, primero fue necesario realizar el control de calidad de los datos e implementar cálculos básicos como promedios, anomalías, etc de nuestra serie de datos. Dado que no teníamos en ese momento experiencia previa en tratar con bases de datos reales ni en programación para ese tipo de datos, esta etapa inicial nos costó más tiempo de lo previsto.

Luego comenzamos con el análisis de aglomeración, además de ser una técnica nueva para nosotras, es un método que no da una solución única, esto es, que deja la posibilidad de considerar diferentes interpretaciones según sean las condiciones iniciales. Por estos motivos, y porque empezaron a emerger resultados novedosos, decidimos enfocarnos únicamente en esta técnica, para poder explorar en ella lo suficiente, y no realizamos el estudio de análisis espectral.

Nos dedicamos en cambio, a profundizar en los resultados obtenidos con el análisis cluster monovariado con otras herramientas no previstas dentro del cronograma inicial, ya que esto nos resultó interesante para interpretación.

- En base a su experiencia de trabajo en equipo en el marco de este Programa, le solicitamos que realice sugerencias o comentarios para ser tomados en cuenta en futuras ediciones del mismo.

Por ser un proyecto incluido en un programa de investigación para varias disciplinas, quizás hubiera sido interesante contar alguna instancia donde se muestran los trabajos en curso, durante los meses en que se desarrolla el proyecto. Por un lado, complementa a lo largo del proceso de investigación por ser una oportunidad de intercambiar diferentes modos de trabajar con otros equipos de estudiantes, y de ayudar a ver diferentes modos de resolver las dificultades planteadas en los diferentes equipos. Y por otro lado, permite conocer las distintas ideas e investigaciones que se realizan en la facultad.

- **Resumen publicable de no más de 250 palabras** que sea accesible para un público amplio, y en un lenguaje dirigido a no especialistas en la temática de la investigación. En este resumen se debe dar cuenta de los objetivos del proyecto, los pasos seguidos para cumplirlos y los principales resultados alcanzados.

**El resumen debe contener la siguiente información:**

Título del proyecto: Análisis e interpretación de datos meteorológicos

Servicio: Facultad de Ingeniería

Nombre de los integrantes del equipo: Vanessa Bentancur, Gianina Giacosa

Nombre del docente orientador: Rafael Terra

### **Resumen publicable:**

El principal objetivo fue familiarizarnos con información meteorológica de alta frecuencia y técnicas estadísticas, interpretando los resultados obtenidos para caracterizar el ciclo diurno. Partimos de una serie diezminutal de temperatura e intensidad de viento provenientes de dos estaciones situadas en puntos distantes del país. Comenzamos con un control de calidad de los mismos y verificamos que el comportamiento de las variables en estudio fuera el esperado.

Luego exploramos la técnica estadística “Análisis de Cluster”, la cual consiste en agrupar aquellos datos que posean características muy similares entre sí. Se tomó el ciclo diurno de cada variable para caracterizar cada día. Con esto se identifica para cada estación ciclos diarios de intensidad de viento y de temperatura, y se estudia después la frecuencia de ocurrencia de cada ciclo, y a partir de las transiciones preferidas entre los distintos ciclos se realizaron grafos. Finalmente comparamos el comportamiento de la misma variable entre las estaciones, y estudiamos la vinculación entre ambas variables en una misma estación.

Como ejemplo, en el caso de temperatura en Colonia Rubio, analizando el grafo correspondiente se ve la diferencia entre el calentamiento, que es gradual y el enfriamiento, que es abrupto. Todos los clusters tienen una probabilidad de recurrencia significativa de un día al siguiente, salvo un caso que parece estar asociado al pasaje de un frente.

Se observa que cada estación tiene su comportamiento específico, en principio porque una está cerca de la costa y la otra no posee la influencia directa del océano.



**Desde el 1/12/2015 y hasta el 15/12/2015 se deberá entregar a los Ayudantes I+D de los Servicios lo siguiente:**

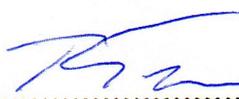
- Un CD con el **informe final** en formato .odt o .pdf. Y con el **póster** en su versión digital en formato .jpg o .pdf
- **Equipos y bibliografía adquiridos con fondos del PAIE** (declarados en la lista conformada en el ítem 8 de este documento)

  
.....  
FIRMA DEL ESTUDIANTE RESPONSABLE

Se solicita al **docente orientador** que brinde una **opinión general acerca del desempeño de su equipo de estudiantes** durante el transcurso de la investigación y que evalúe en forma breve los **resultados** expuestos a través de este informe y el contenido de su **resumen publicable**. (máx 200 palabras)

**Comentarios del docente orientador:**

El proyecto como experiencia de iniciación a la investigación claramente cumplió sus objetivos pues se notó claramente el proceso de maduración de las estudiantes. En ese sentido se pueden distinguir dos etapas. La primera en que se desarrolló los análisis tal cual fueron planificados originalmente en la propuesta, que implicó la familiarización con datos, el estudio de técnicas y su implementación. Las dificultades en esta etapa resultaron las asociadas con no entender algún aspecto teórico o con falta de agilidad en la implementación. Dados los interesantes resultados parciales obtenidos, el proyecto se desvió un poco del plan original intentando interpretar mejor los mismos con técnicas que se fue viendo como apropiadas a la situación. En este caso las dificultades fueron mayores, necesitaron más orientación, como es natural, pero también entiendo que el aprendizaje fue mayor. En todo momento se notó una alta motivación y buena dedicación al trabajo, con los altibajos naturales debido a otros compromisos estudiantiles. En cuanto a los resultados en sí, realmente son de mucho interés y es de interés continuar dicha línea de trabajo.

  
.....  
FIRMA DEL DOCENTE ORIENTADOR