



Fecha de presentación del Informe: Día Mes Año

Datos generales del Proyecto

Código del proyecto: 2842			
Título del proyecto: Modelación estadística de la contaminación del aire por partículas de diámetro aerodinámico menor que 2.5 μm (PM _{2.5})			
Facultad o Instituto Académico: Facultad Ingeniería			
Departamento o Escuela: Escuela de Estadística			
Grupo (s) de investigación: Inferir			
Investigadores ¹	Nombre	Tiempo asignado (Horas Semanales)	Tiempo dedicado
Investigador Principal	Javier Olaya Ochoa	15	15
Coinvestigadores			
Otros participantes	Camilo Jiménez	10	10
Otros participantes	Andrés Caicedo	10	10
Otros participantes	Estefanía Santiago	10	10
Otros participantes	Fabian Chicangana	10	10
Otros participantes	Elizabeth Cabezas	10	10
Otros participantes	Johann Alexis Ospina	0	0
Otros participantes	Alejandro Delgado	15	15
Otros participantes	Cristhian Leonardo Urbano	15	15



FACULTAD DE INGENIERÍA
Vicedecanatura de
Investigación y Posgrados

04 MAY 2017

RECIBIDO

¹ Todas las personas relacionadas en el informe y que participen en el proyecto deben haber suscrito el acta de propiedad intelectual de acuerdo con los formatos establecidos.

1. Resumen ejecutivo:

La contaminación ambiental de manera general afecta gravemente la salud humana, específicamente el material particulado (PM), el cual puede ser clasificado de acuerdo a su diámetro aerodinámico en 2 tipos, definidos de manera oficial como: Partículas gruesas las cuales poseen un diámetro inferior a $10\ \mu\text{m}$ y mayor a $2.5\ \mu\text{m}$ (PM_{10}) y partículas finas como aquellas que presentan un diámetro aerodinámico menor a $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$). Estas últimas según la Organización Mundial de la Salud (OMS) se encuentran asociadas a la aparición de enfermedades respiratorias crónicas como: cáncer de pulmón, asma, pulmonía, etc. Debido a ello, se han implementado normas para la medición del contaminante, lo cual, requiere protocolos de monitoreo y vigilancia permanente.

En la ciudad de Cali se cuenta con un sistema de vigilancia de la calidad del aire (SVCASC) compuesto por 9 estaciones; no obstante, este sistema mide las partículas con diámetro menor a $10\ \mu\text{m}$ incluyendo las partículas gruesas y finas y se clasifica como PM_{10} ; mientras que las partículas finas si se encuentran medidas en concordancia con la definición oficial. Vale aclarar, que no todas las estaciones se encuentran habilitadas para la medición de partículas finas y las que se encuentran habilitadas no reportan las mediciones de manera continua y regular debido a fallas de tipo técnico y mantenimiento de los equipos. Por tal motivo, en este proyecto se hace uso de técnicas de imputación de datos faltantes para aumentar la cantidad de datos disponibles, y se caracteriza el comportamiento diario de la concentración de partículas finas mediante el uso de técnicas de datos funcionales, con lo cual se logra visualizar los momentos del día en los que se presenta la mayor contaminación y la presencia de niveles atípicos. Adicionalmente, se explora el uso de un Modelo Aditivo Generalizado (GAM) a fin de observar el efecto que tienen variables climatológicas como: velocidad y dirección del viento, temperatura, humedad, radiación solar y lluvia sobre la concentración del contaminante.

2. Síntesis del proyecto:

Tema:

Modelación estadística de la contaminación del aire por partículas de diámetro aerodinámico menor que $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$).

Objetivo General

Caracterizar el comportamiento a lo largo de un día de las concentraciones horarias de las partículas suspendidas en el aire cuyos diámetros aerodinámicos son menores que $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$).

Objetivos Específicos

- Evaluar y proponer mecanismos de imputación de datos faltantes de $\text{PM}_{2.5}$.
- Proponer una curva típica del comportamiento diario de los niveles de contaminación en el aire por $\text{PM}_{2.5}$ en las estaciones de monitoreo de la ciudad de Cali, que permita estimar los niveles a cualquier hora del día.
- Explorar el uso de un modelo aditivo generalizado para la estimación de $\text{PM}_{2.5}$ a partir de variables de clima.

- Explorar el uso de un modelo de regresión funcional para la estimación de $PM_{2.5}$ a partir de variables de clima.
- Avanzar en el estudio del problema de la estimación de la varianza en los modelos propuestos.
- Promover la medición generalizada y permanente de PM_{10} y $PM_{2.5}$ en todas las estaciones del sistema de vigilancia de la calidad del aire de Cali.

Metodología

El estudio realizado se centró en la caracterización de la contaminación por partículas suspendidas en el aire $PM_{2.5}$, las cuales son consideradas por la Organización Mundial de la Salud (OMS) como una de las principales causas para la adquisición de enfermedades, especialmente las de tipo respiratorio.

Los datos se producen de manera discreta, uno cada 10 segundos, y se reportan acumulados como datos horarios. Para cada hora se dispone de la medición promedio, constituyendo una serie temporal. Vale decir, que el análisis no se condujo mediante la técnica de Análisis de Series de Tiempo, debido a que en estudios previos (Barrientos, 2007) se indicó que tal vía no es la más recomendable en este tipo de estudios, por tal motivo la metodología utilizada para alcanzar los objetivos de este proyecto fue la siguiente:

Revisión de Literatura:

En la fase de revisión de literatura se contó con diversas fuentes de información dentro de las cuales, algunas se descartaron y otras resultaron de vital importancia para la realización de los objetivos. Por ejemplo, para la imputación de datos faltantes Chiou et al. (2014), Meléndez et al. (2015), Junger & de León (2015), Junninen et al. (2004), Gómez-Carracedo et al. (2014) y Plaia & Bondi (2006) se encontraron relevantes. Entre tanto, para el análisis exploratorio funcional se consideraron los trabajos de Franklin et al. (2007), Laden et al. (2006), Franklin et al. (2008), Londoño & Vasco (2008), Gaviria et al. (2011), Enkhmaa et al. (2014), Wang et al. (2006), Bedoya & Martínez (2009), Giraldo (2007), Staniswalis et al. (2009), Torres et al. (2011), Bi & Kuesten (2013), Din & Rambely (2014), Suhaila (2014) y Shaadan et al. (2014^a). Por otro lado, para el ajuste del Modelo Aditivo Generalizado (GAM), los trabajos de Aldrin & Hobæk 2005, Liu et al. 2009, Belusic et al. 2015 y Song et al. 2015 fueron apropiados. Vale decir que la revisión de la bibliografía se efectuó de forma permanente a lo largo de todo el proyecto.

Recolección y Validación de Datos:

El material particulado es definido de manera oficial como partículas gruesas las cuales poseen un diámetro inferior a $10\ \mu m$ y mayor a $2.5\ \mu m$ (PM_{10}) y partículas finas como aquellas que presentan un diámetro aerodinámico menor a $2.5\ \mu m$ ($PM_{2.5}$).

Para este proyecto se utilizó la información del Sistema de Vigilancia de la Calidad del Aire de Santiago de Cali (SVCASC) suministrada por el DAGMA entre los años 2013 a 2015. Los niveles de $PM_{2.5}$ se miden rutinariamente en cuatro estaciones de la ciudad y el SVCASC tiene 8 estaciones fijas en operación, así como una unidad móvil, ubicadas en diferentes puntos de la ciudad. Sin embargo, este sistema mide las partículas con diámetro menor a $10\ \mu m$ incluyendo

las partículas gruesas y finas y se clasifica como PM_{10} ; mientras que las partículas finas si se encuentran en concordancia con la definición oficial.

Adicionalmente, si bien las muestras recolectadas son obtenidas mediante mediciones continuas y permanentes de las partículas, no se contó con las bases de datos completas debido a la ausencia de mediciones reportadas, lo cual motivó uno de los objetivos del proyecto.

Procesamiento y análisis de datos

La información se procesó mediante el uso del software de libre distribución R, diseñado y mantenido por el Proyecto R para la Computación Estadística (R Core Team 2014). Los métodos estadísticos necesarios para la ejecución del proyecto están disponibles en este software. Los análisis se basaron en los métodos estándar de suavización y regresión no-paramétrica.

Resultados obtenidos

A partir del trabajo de imputación de datos realizado por los estudiantes Camilo Jiménez y Andrés Caicedo, se encontró que las concentraciones de $PM_{2.5}$ son más bajas los días lunes y de manera similar el domingo. Entre tanto, las concentraciones son mayores en los demás días de la semana con presencia de máximos que sobrepasan los valores permisibles por las autoridades ambientales. Se evidencia también, un comportamiento marcado de niveles más bajos en las primeras horas del día con aumentos significativos entre las 7:00 y las 9:00 y aproximadamente, a partir de las 15:00 la concentración de $PM_{2.5}$ se estabiliza alrededor de un valor, seguido por valores bajos en horas de la noche. También se refleja un efecto de rezago en los niveles de $PM_{2.5}$, ya que estos tienden a ser más altos una hora después de los momentos con mayor afluencia vehicular.

Otro resultado importante es el papel que juega la proporción de datos faltantes en un día, ya que cuando esta es mayor el método resulta ser ineficiente y por lo tanto, no recoge muy bien la variabilidad inherente al comportamiento. Sin embargo, en los días con mayor homogeneidad como los lunes y domingos el proceso de imputación resultó óptimo incluso con una mayor proporción de datos faltantes. Se procede entonces a usar el método de imputación para la concentración de $PM_{2.5}$ reportada por la estación Compartir en el período 1 de Diciembre 2014 al 30 de Noviembre 2015 imputando 97 registros horarios y como máximo 6 por día, con ello se consiguió pasar de un 46.2% a un 61.9% de días utilizables para este período. Trabajo que fue realizado por los estudiantes Fabian Chicangana y Estefanía Santiago con la finalidad de caracterizar el comportamiento del contaminante de interés para su posterior análisis, del cual podemos resaltar el hallazgo de 3 horas del día con picos alrededor de las 8:00, las 14:00 y las 21:00 horas, lo cual parece responder a los 3 momentos del día con mayor afluencia vehicular con un rezago aproximadamente de una hora; sin embargo este rezago, puede proceder de la forma en la que son registradas las mediciones, ya que si bien, se toma cada 10 segundos solo se reporta el valor promedio de la contaminación en la hora inmediatamente anterior.

En términos de la variabilidad de las concentraciones promedio horarias de $PM_{2.5}$, se encontró una evidente diferencia entre las primeras 11 horas del día. Adicionalmente, se resalta la

similitud en el comportamiento de las curvas de los días lunes con las obtenidas para los días domingo. Cabe resaltar la presencia de mediciones atípicas en la concentración de $PM_{2.5}$ el día primero de enero, fenómeno que parece obedecer a la quema de pólvora por motivo de la celebración de fin de año.

Vale decir que los resultados obtenidos hasta aquí son de carácter descriptivo del comportamiento del contaminante de interés ignorando el posible efecto de otras variables climatológicas sobre éste. Debido a esto, surgió la necesidad de explorar la utilización de un modelo GAM en presencia de variables como: velocidad y dirección del viento, temperatura, humedad, radiación solar y lluvia, el cual estuvo a cargo de los estudiantes Camilo Cuchumbé y Leydy Ortega. Como primera medida se encontró la correlación lineal usando el coeficiente de correlación de Pearson, del cual se pudo notar que las variables de clima no presentan correlación lineal estadísticamente significativa, por lo tanto el modelo ajustado no resulta adecuado para mostrar el efecto de las variables climáticas mencionadas sobre la concentración de $PM_{2.5}$, así, incluyendo todas las variables de clima se obtuvo un 16% de la variación explicada por el modelo, lo cual es ineficiente.

Principales conclusiones y recomendaciones

- A pesar de que el comportamiento de la concentración del contaminante de interés presenta cierto patrón a determinadas horas del día, este parece ser no periódico ni regular y tener relación con el día de la semana.
- La caracterización de las curvas reveló la presencia de 3 momentos del día con concentraciones elevadas de $PM_{2.5}$ en instantes de tiempo posterior a las horas de mayor afluencia vehicular.
- Cabe resaltar que la concentración de material particulado fino, a causa de la celebración de fin de año alcanza niveles atípicos, los cuales se ven reflejados el 1 de enero en la hora 1:00.
- Las técnicas de modelación usuales en el enfoque no paramétrico mostraron resultados débiles, por lo cual, se recomienda el uso de técnicas propias del análisis de datos funcionales para modelar la concentración de material particulado y comparar su eficiencia con los resultados obtenidos en este proyecto.
- Las técnicas de imputación propuestas en este trabajo no resultan ser de mucha ayuda si la proporción de datos faltantes por día es grande, por tal motivo se recomienda explorar los resultados obtenidos tomando únicamente los datos disponibles y comparar las conclusiones con las alcanzadas usando la técnica de imputación.
- Se recomienda una mayor conceptualización sobre la construcción e interpretación de los boxplots funcionales.

Referencias:

- Estimación de la demanda y del precio por kilovatio de energía eléctrica (Barrientos, González y Olaya, 2007; y Herrera, Lozano y Olaya, 2009).
- Chiou, J.-M., Zhang, Y.-C., Chen, W.-H. & Chang, C.-W. (2014), "A functional data approach to missing value imputation and outlier detection for traffic flow data", *Transportmetrica B: Transport Dynamics* 2(2), 106-129.
- Meléndez, R. R., Bolívar Atuesta, S., Morales Ospina, V. & Rojano, R. (2015), "Imputación de valores faltantes y detección de valores atípicos para datos funcionales", XXV Simposio Internacional de Estadística. 2015-08-05/2015-08-08. Armenia, Quindío.
- Junger, W. & de Leon, A. P. (2015), "Imputation of missing data in time series for air pollutants", *Atmospheric Environment*, 102, 96-104. URL: <http://www.sciencedirect.com/science/article/pii/S1352231014009145>.
- Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. & Kolehmainen, M. (2004), "Methods for imputation of missing values in air quality data sets", *Atmospheric Environment* 38(18), 2895-2907.

URL: <http://www.sciencedirect.com/science/article/pii/S1352231004001815>.

- Gómez-Carracedo, M., Andrade, J., López-Mahía, P., Muniategui, S. & Prada, D. (2014), "A practical comparison of single and multiple imputation methods to handle complex missing data in air quality datasets", *Chemometrics and Intelligent Laboratory Systems* 134(0), 23-33.

URL: <http://www.sciencedirect.com/science/article/pii/S016974391400032X>.

- Plaia, A. & Bondí, A. (2006), "Single imputation method of missing values in environmental pollution data sets", *Atmospheric Environment* 40(38), 7316-7330.

URL: <http://www.sciencedirect.com/science/article/pii/S1352231006006959>.

- Franklin, M., Zeka, A. & Schwartz, J. (2007), 'Association between pm2.5 and all-cause and specific-cause mortality in 27 us communities', *Journal of Exposure Science and Environmental Epidemiology* 17(3), 279-287.
- Laden, F., Schwartz, J., Speizer, F. E. & Dockery, D. W. (2006), 'Reduction in fine particulate air pollution and mortality: extended follow-up of the harvard six cities study', *American journal of respiratory and critical care medicine* 173(6), 667-672.
- Londoño, C. A. E. & Vasco, G. J. M. (2008), 'Relación entre las partículas finas (pm2. 5) y respirables (pm10) en la ciudad de Medellín', *Revista Ingenierías Universidad de Medellín* 7(12), 23-42.
- Gaviria, G., Carlos, F., Benavides, C., Paula, C. & Tangarife, C. A. (2011), 'Contaminación por material particulado (pm2, 5 y pm10) y consultas por enfermedades respiratorias en Medellín (2008-2009)', *Revista Facultad Nacional de Salud Pública* 29(3), 241-250
- Enkhmaa, D., Warburton, N., Javzandulam, B., Uyanga, J., Khishigsuren, Y., Lodoysamba, S., Enkhtur, S. & Warburton, D. (2014), 'Seasonal ambient air pollution correlates strongly with spontaneous abortion in mongolia', *BMC pregnancy and childbirth* 14(1), 1.

- Wang, X., Bi, X., Sheng, G. & Fu, J. (2006), 'Chemical composition and sources of pm10 and pm2.5 aerosols in guangzhou, china', *Environmental Monitoring and Assessment* 119(1-3), 425-439.
- Bedoya, J. & Martínez, E. (2009), 'Calidad del aire en el valle de aburrá antioquia-colombia', *Dyna* 76(158), 7-15.
- Giraldo, R. (2007), 'Análisis exploratorio de variables regionalizadas con métodos funcionales', *Revista Colombiana de Estadística* 30(1), 115-127.
- Staniswalis, J. G., Yang, H., Li, W.-W. & Kelly, K. E. (2009), 'Using a continuous time lag to determine the associations between ambient pm2.5 hourly levels and daily mortality', *Journal of the Air & Waste Management Association* 59(10), 1173-1185.
- Torres, J. M., Nieto, P. G., Alejano, L. & Reyes, A. (2011), 'Detection of outliers in gas emissions from urban areas using functional data analysis', *Journal of hazardous materials* 186(1), 144-149.
- Bi, J. & Kuesten, C. (2013), 'Using functional data analysis (fda) methodology and the R package 'fda' for sensory time-intensity evaluation', *Journal of Sensory Studies* 28(6), 474-482.
- Din, W. R. W. & Rambely, A. S. (2014), Functional data analysis on ground reaction force of military load carriage increment, in 'Proceedings of the 3rd International Conference on Mathematical Sciences', Vol. 1602, AIP Publishing, pp. 83-88.
- Suhaila, J. (2014), Statistical analysis on mean rainfall and mean temperature via functional data analysis technique, in 'Statistics and Operational Research International Conference (SORDIC 2013)', Vol. 1613, AIP Publishing, pp. 368-379.
- Shaadan, N., Jemain, A. A. & Deni, S. M. (2014a), The construction of control chart for pm10 functional data, in 'Proceedings of the 3rd International Conference on Mathematical Sciences', Vol. 1602, AIP Publishing, pp. 1161-1167.
- Aldrin, M. & Hobæk, I. (2005), 'Generalised additive modelling of air pollution, traffic volumen and meteorology', *Atmospheric Environment* (39), 2145,2155.
- Liu, Y., Paciorek, C. & Koutrakis, P. (2009), 'Estimating regional spatial and temporal variability of pm2.5 concentrations using satellite data, meteorology, and land use information', *Environmental Health Perspectives* 117(6), 886,892.
- Belusic, A., HercegBulic, I. & Bencetic, Z. (2015), 'Using a generalized additive model to quantify the influence of local meteorology on air quality in zagreb', *Geofizika* 32, 47,77.
- Song, Y., Yang, H., Peng, J., Song, Y., Sun, Q. & Li, Y. (2015), 'Estimating pm2.5 concentrations in xi'an city using a generalized additive model with multi-source monitoring data'.



3. Productos:

Tabla No. 1. Cantidad y tipo de productos pactados en el *Acta de Trabajo y Compromiso* y productos finalmente presentados

TIPO DE PRODUCTOS	No. de PRODUCTOS PACTADOS		No. de PRODUCTOS PRESENTADOS	
Productos de nuevos conocimientos				
Artículo completo sometido para publicación en revistas A1 o A2	1		1	
Artículo completo publicados en revistas B	0		0	
Artículo completo publicados en revistas C	0		0	
Libros de autor que publiquen resultados de investigación	0		0	
Capítulos en libros que publican resultados de investigación	0		0	
Productos o procesos tecnológicos patentados o registrados	0		0	
☐ Prototipos y patentes	0		0	
☐ Software	0		0	
Productos o procesos tecnológicos usualmente no patentables o protegidos por secreto industrial	0		0	
Normas basadas en resultados de investigación	0		0	
Formación de recursos humanos	No. de estudiantes vinculados	No. de tesis	No. De estudiantes Vinculados	No. De tesis
Estudiantes de pregrado	2	1	4	2
Semillero de Investigación	Al menos 2	0	8	4
Estudiantes de maestría	1	0	2	0
Estudiantes de doctorado	0	0	0	0
Productos de divulgación				
Publicaciones en revistas no indexadas				

TIPO DE PRODUCTOS	No. de PRODUCTOS PACTADOS		No. de PRODUCTOS PRESENTADOS	
Ponencias presentadas en eventos (congresos, seminarios, coloquios, foros)	No. de ponencias nacionales	No. de ponencias internacionales	No. de ponencias nacionales	No. de ponencias internacionales
	1	1	2	2
Propuesta de investigación				
Propuestas presentadas en convocatorias externas para búsqueda de financiación.	1		1	

Tabla No. 2. Detalle de productos.

Para cada uno de los productos obtenidos y relacionados en la tabla anterior, indique la información solicitada para cada uno, anexando copia de las respectivas constancias. Como anexo a esta guía encontrará el instructivo para la revisión de informes finales y productos

Tipo de producto:	Trabajo de grado
Nombre General:	Modelación mediante series de tiempo de la contaminación por PM2.5
Nombre Particular:	Modelación mediante series de tiempo de la contaminación por PM2.5
Ciudad y fechas:	Cali, 11 de marzo de 2016
Participantes:	Anderson Castañeda Trujillo y Javier Olaya Ochoa
Sitio de información:	Biblioteca Mario Carvajal, Universidad del Valle
Formas organizativas:	Grupo de Investigación en Estadística Aplicada – INFERIR, Escuela de Estadística, Universidad del Valle

Tipo de producto:	Trabajo de grado
Nombre General:	Métodos de imputación de datos faltantes en las mediciones de las concentraciones de $PM_{2.5}$ en la estación de monitoreo Univalle en Santiago de Cali
Nombre Particular:	Métodos de imputación de datos faltantes en las mediciones de las concentraciones de $PM_{2.5}$ en la estación de monitoreo Univalle en Santiago de Cali
Ciudad y fechas:	Cali, 11 de abril de 2016
Participantes:	Camilo Jiménez, Andrés Caicedo y Javier Olaya
Sitio de información:	Biblioteca Mario Carvajal, Universidad del Valle
Formas organizativas:	Grupo de Investigación en Estadística Aplicada – INFERIR, Escuela de Estadística, Universidad del Valle

Tipo de producto:	Trabajo de grado
Nombre General:	Análisis exploratorio funcional de las curvas diarias de concentraciones promedio horarias de $PM_{2.5}$ en el aire de la ciudad de Santiago de Cali para el periodo diciembre 2014 - noviembre 2015
Nombre Particular:	Análisis exploratorio funcional de las curvas diarias de concentraciones promedio horarias de $PM_{2.5}$ en el aire de la ciudad de Santiago de Cali para el periodo diciembre 2014 - noviembre 2015
Ciudad y fechas:	Cali, 6 de marzo de 2017
Participantes:	Fabian Chicangana, Estefanía Santiago y Javier Olaya
Sitio de información:	Biblioteca Mario Carvajal, Universidad del Valle
Formas organizativas:	Grupo de Investigación en Estadística Aplicada – INFERIR, Escuela de Estadística, Universidad del Valle

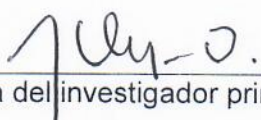
Tipo de producto:	Artículo
Nombre General:	Sometido el 03/05/2017 para publicación en la revista DYNA de la Universidad Nacional de Colombia.
Nombre Particular:	Acerca de la estimación de la fracción PM2.5/PM10
Ciudad y fechas:	Bogotá, Sometido el 03/05/2017.
Participantes:	Javier Olaya Ochoa, Diana Paola Ovalle Muñoz, Cristhian Leonardo Urbano Leon
Sitio de información:	Revista DYNA de la Universidad Nacional de Colombia y Biblioteca Mario Carvajal Universidad del Valle.
Formas organizativas:	Grupo de Investigación en Estadística Aplicada – INFERIR, Escuela de Estadística, Universidad del Valle.

4. Impactos actual o potencial:

A partir de los resultados de éste proyecto, podrían derivarse estudios principalmente en el ámbito de la salud, entre los cuales caben mencionar, el tratar de identificar las enfermedades ocasionadas por la contaminación del aire o también tratar de determinar si la zona donde se presenta mayor concentración de partículas suspendidas es aquella donde se presentan más casos de enfermedades, especialmente las de tipo respiratorio.

Un impacto que presentaría a largo plazo, sería el que se generen políticas de gestión ambiental para el control, seguimiento y prevención de la calidad del aire, mediante este proyecto o los derivados por este, en nuestra ciudad.

La presente versión del informe contiene las observaciones de los evaluadores:


Firma del investigador principal

VoBo. Vicedecano de Investigaciones

Por favor presente su informe impreso y en formato digital en hoja tamaño carta, letra arial 11, con espacios de 1 ½