



## Plan de Gestión de Datos

| <b>INFORMACIÓN SOBRE EL PROYECTO</b>  |  |
|---|--|
| <b>1. – Datos del Proyecto</b>  |  |
| <b>- Título del Proyecto (en castellano)</b>  |  |
| Datos complejos de alta dimensión: predicción, regresión eficiente y reducción suficiente de dimensiones.   |  |
| <b>- Título del Proyecto (en inglés)</b>  |  |
| Complex high-dimension data: prediction, efficient regression and sufficient dimension reduction.   |  |
| <b>- Descripción del Proyecto (en castellano) Resumen</b>   |  |
| <p>El objetivo de este plan de trabajo es proveer soluciones originales desde la estadística matemática a problemas de regresión y predicción con predictores y/o respuestas pertenecientes a espacios de alta dimensión y/o espacialmente correlacionados. En particular, en problemas de regresión con predictores y/o respuestas pertenecientes a espacios de alta dimensión, cuando la distribución de los datos pueda describirse por medio de modelos gráficos probabilísticos, la propuesta busca explotar la capacidad de modelado junto con las ventajas computacionales que ellos presentan para desarrollar nuevos métodos de regresión eficiente y de reducción suficiente de dimensiones para problemas con datos complejos. La motivación es dar soluciones estadísticamente rigurosas a problemas concretos de las ciencias médicas, ciencias sociales, ciencias de la tierra, entre otras. Estas aplicaciones se caracterizan por presentar variables de distinta naturaleza con una fuerte estructura de dependencia entre ellas, sobredispersión, y/o muchos ceros en las observaciones, que no son abordables con la metodología actual. Además, se busca desarrollar métodos de predicción espacial paramétrica y semiparamétrica para predictores multivariados que puedan llegar a pertenecer a espacios de alta dimensión. Una de las propiedades buscadas a partir de los métodos propuestos es la flexibilidad en términos de las variables a utilizar, las cuales podrán ser continuas, de conteo o categóricas, y de los puntos geográficos necesarios para poder predecir la variable de interés en el sitio no muestreado. Gracias a esa flexibilidad, el nuevo método podrá aplicarse para predecir variables en áreas como ser la ecología, epidemiología, ciencias de la tierra, climatología, ciencias sociales, economía, entre otras. Por ejemplo, en el contexto de medición del bienestar socioeconómico de la población, este método puede ser utilizado como una herramienta actualizable de monitoreo y evaluación de pobreza mediante la construcción de mapas de privaciones socio-económicas locales. En el contexto geoestadístico, el método podrá aplicarse para crear mapas de concentración de ciertos metales en la tierra, como por ejemplo de zinc, a partir de la medición de otros metales, características del suelo y otras variables ambientales.</p> |  |
| <b>- Descripción del Proyecto (en inglés) Resumen</b>   |  |
| <p>The objective of this plan is to provide original solutions from the mathematical statistics to regression problems with predictors and/or responses belonging to high-dimensional and/or spatially correlated spaces. In particular, in regression problems with predictors and/or responses belonging to high-dimensional spaces, when the distribution of the data can be described by probabilistic graphic models, the proposal seeks to exploit the modeling capacity together with the computational advantages they present, to develop new efficient regression</p>   |  |



and sufficient dimension reduction methods for complex data problems. The motivation is to provide statistical rigorous solutions to specific problems of medical sciences, social sciences, earth sciences, among others. These applications are characterized by presenting variables of different nature with a strong structure of dependence among them, overdispersion, and/or many zeros in the observations, which are not approachable with the current methodology. In addition, this plan seeks to develop parametric and semi-parametric spatial prediction methods for multivariate predictors that may belong to high-dimensional spaces. One of the properties sought from the proposed methods is flexibility in terms of the variables to be used, which may be continuous, counting or categorical, and of the geographical points necessary to perform the prediction in non-sampled site. Thanks to this flexibility, the new method can be applied to predict variables in areas such as ecology, epidemiology, earth sciences, climatology, social sciences, economics, among others. For example, in the context of measuring the socioeconomic welfare of the population, this method can be used as an updatable tool for monitoring and assessing poverty by building maps of local socio-economic deprivations. In the geostatistical context, the method can be applied to create maps of concentration of certain metals in the earth, such as zinc, from the measurement of other metals, soil characteristics and other environmental variables.

Throughout the execution of the project, efforts will be made to provide comprehensive answers to the problems raised, ranging from the analysis of the theoretical properties of the proposed estimators, tools for their computation and applying the proposed methodologies to the specific problems mentioned above.

**- Palabras Claves descriptivas del Proyecto (en castellano)**

|                                   |                     |                          |
|-----------------------------------|---------------------|--------------------------|
| DATOS COMPLEJOS DE ALTA DIMENSION | REGRESION EFICIENTE | REDUCCION DE DIMENSIONES |
|-----------------------------------|---------------------|--------------------------|

**- Palabras Claves descriptivas del Proyecto (en inglés)**

|                  |                      |                     |
|------------------|----------------------|---------------------|
| COMPLEX BIG DATA | EFFICIENT REGRESSION | DIMENSION REDUCTION |
|------------------|----------------------|---------------------|

**2 – Datos del Director/ar del Proyecto**

**- Nombre y Apellido**

Pamela Llop

**- Unidad Académica**

Faculta de Ingeniería Química

**- Teléfono oficial de contacto**

342 4571177 int 34

**-Teléfono móvil de contacto**

342 5170573

**-E-mail del Director/a del Proyecto**

lloppamela@gmail.com

**DATOS RESULTANTES DE LA EJECUCIÓN DEL PROYECTO**

**-Describe la toma de muestras / datos a realizar**

|  |
|--|
|  |
|--|



**– Datos: ¿Existe alguna razón por la cual los datos declarados no deban ser puestos a disposición de la comunidad/ser de acceso público? (marque X)**

|  |  |
|--|--|
|  | <b>NO</b>  |
|  | <b>SI. Elija una de las opciones:</b>  |
|  | a) Se encuentra en evaluación de protección por medio de patentes                      |
|  | b) No se inició el proceso de evaluación de patentabilidad, pero podría ser protegible |
|  | c) Existe un contrato con un tercero que impide la divulgación                         |
|  | d) Otro. Justifique.   |

**– Período de Confidencialidad: Es el período durante el cual los datos no deberían ser publicados, contado a partir del momento de la toma de los mismos. El período máximo para la no publicación es de 5 (CINCO) años posteriores a su obtención. Luego de este periodo, los datos estarán disponibles para la comunidad/serán de acceso público.**

**Si Ud. considera que este tiempo es insuficiente, y necesita prorrogar el período de confidencialidad, indique sus motivos y la cantidad de años adicionales que considera necesarios. Marque su opción con “X”.**

|  |                       |
|--|-----------------------|
|  | <b>1 (UN) año</b>     |
|  | <b>2 (DOS) años</b>   |
|  | <b>3 (TRES) años</b>  |
|  | <b>4 (CUATRO) año</b> |
|  | <b>5 (CINCO) años</b> |
|  | <b>Otro.</b>          |
|  | <b>Motivos:</b>       |

Pamela Llop



## **INSTRUCTIVO PARA COMPLETAR EL PLAN DE GESTIÓN (PGD)**

El PGD no es un documento definitivo, sino que se desarrollará a lo largo del ciclo de vida del proyecto.

### **INFORMACIÓN SOBRE EL PROYECTO**

#### **1 – Datos del Proyecto**

**Título del Proyecto (en castellano):** Deberá ingresar el título completo del proyecto (en castellano), indicando además el código asignado por la SCAYT.

**Título del Proyecto (en inglés):** Deberá ingresar el título completo del proyecto en inglés.

**Descripción del Proyecto (en castellano):** Deberá ingresar la descripción del Proyecto en castellano.

**Descripción del Proyecto (en inglés):** Deberá ingresar la descripción del Proyecto en inglés.

**Palabras Claves descriptivas del Proyecto (en castellano):** Deberá ingresar tres palabras claves descriptivas del Proyecto, en castellano.

**Palabras Claves descriptivas del Proyecto (en inglés):** Deberá ingresar tres palabras claves descriptivas del Proyecto, en inglés.

#### **2- Datos del Director/a del Proyecto**

**Nombre y Apellido del Titular del Proyecto:** Nombre completo y apellido del Titular del Proyecto.

**Unidad Académica:** Nombre de la Unidad Académica a la que pertenece el/la directora/a del Proyecto.

**Teléfono oficial de contacto:** Número de teléfono de la oficina/laboratorio/Institución del Director/a del Proyecto, donde pueda ser contactado, incluyendo número de área/país (ej: Para Santa Fe: + 54 9 342 4999-9999).

**Teléfono móvil de contacto:** Número de teléfono móvil del director/ar del Proyecto, donde pueda ser contactado, incluyendo número de área/país.

**E-mail del Director/a del Proyecto:** Correo electrónico de contacto del Director/a del Proyecto.

### **DATOS RESULTANTES DE LA EJECUCIÓN DEL PROYECTO**



**Describa la toma de muestras/datos a realizar:** Información descriptiva sobre la toma de muestras que resultarán en datos/conjuntos de datos. La descripción deberá incluir información de contexto (lugar de toma de los datos; instrumentos, etc.)

**Datos:** ¿Existe alguna razón por la cual los datos declarados no deban ser puestos a disposición de la comunidad/ser de acceso público? Deberá marcar con una “X” la opción correcta. En caso de responder afirmativamente, deberá justificar debidamente, comprendiendo que sólo en casos de extrema excepcionalidad esta restricción de acceso a los datos resulta practicable/aceptable.

**Período de Confidencialidad:** Es el periodo durante el cual los datos no deberían ser publicados, contado a partir del momento de la toma de los mismos. El periodo máximo para la no publicación es de 5 (CINCO) años posteriores a su obtención. Luego de este periodo, los datos estarán disponibles para la comunidad/serán de acceso público.

Si Ud. considera que este tiempo es insuficiente, y necesita prorrogar el período de confidencialidad, indique sus motivos y la cantidad de años adicionales que considera necesarios.

Deberá indicar los años que considera necesario prorrogar el período de confidencialidad y explicar los motivos.