Curso Ciencia Viva. CPR de Oviedo
06/Marzo/2017

Robots Biomédicos.

Ejemplos en medicina traslacional.

Juan Luis Fernández-Martínez

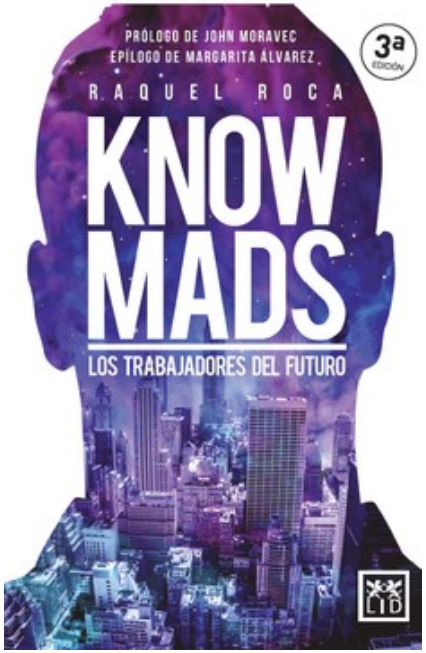
Departamento de Matemáticas.

Grupo de Problemas Inversos,

Optimización y Aprendizaje Automático

UNIVERSIDAD DE OVIEDO
Departamento de Matemáticas

KNOWMADS



KNOWMADS

El término ***knowmad*** es un neologismo que combina las palabras **know y nomad** y que da cuenta del perfil del sujeto capaz de ser un nómada del conocimiento. Fue creado por John Moravec para referirse a los nómadas del conocimiento y la innovación. **Se caracteriza por ser innovador, imaginativo, creativo, capaz de trabajar en colaboración con casi cualquier persona, en cualquier momento y lugar**.

Un *knowmad* es valorado por su conocimiento personal, lo que le proporciona una ventaja competitiva con respecto a otros trabajadores.

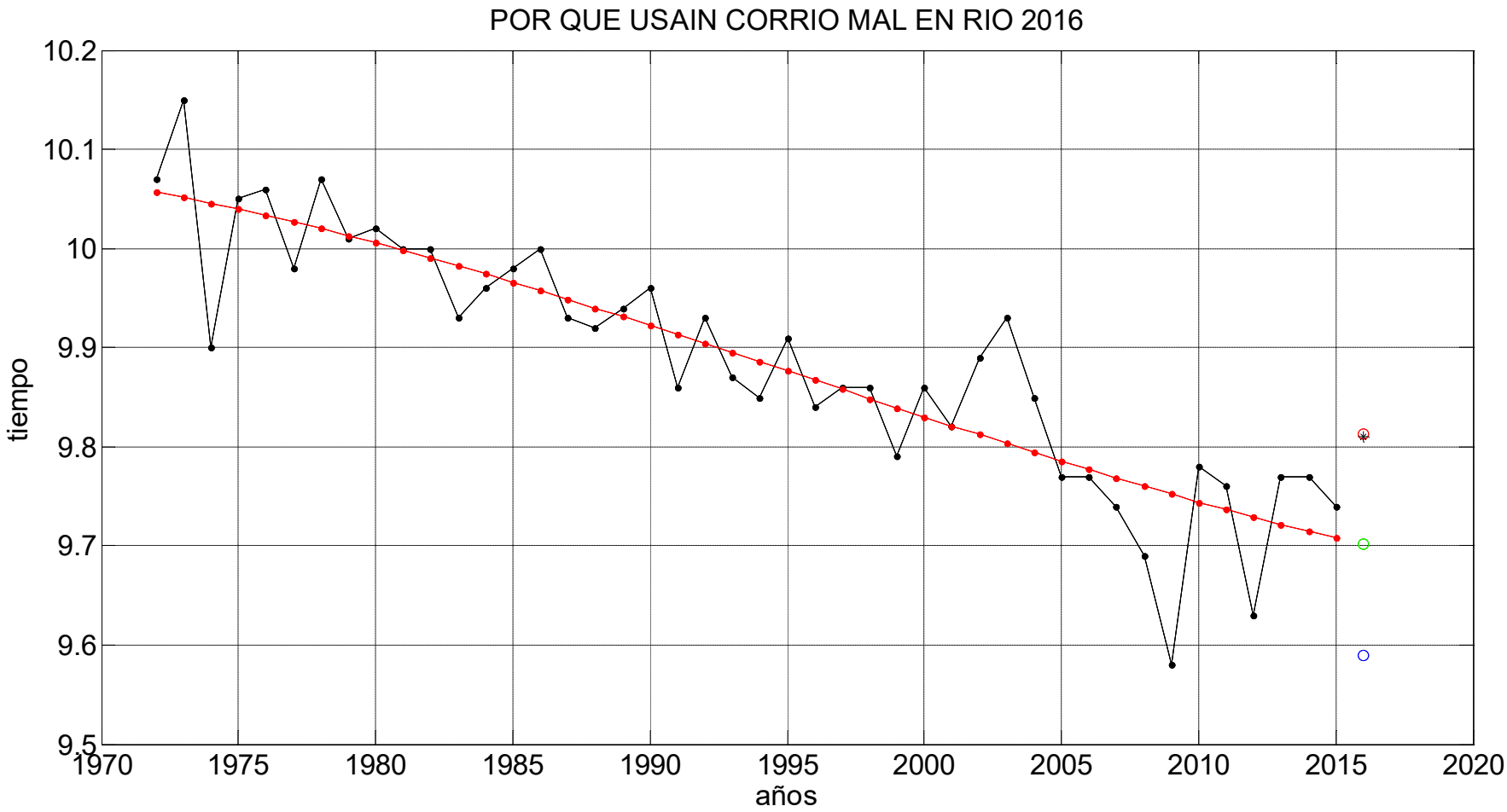
I’m a Knowmad

1. Soy Dr. Ingeniero de Minas (Escuela de Minas de Oviedo, 1994) e Ingeniero de Exploración petrolífera de formación (Ecole Nationale du Pétrole et des Moteurs (1988), Imperial College of Science and Technology de Londres (1989).
2. He vivido, estudiado y trabajado en varios países diferentes: España, Francia, Inglaterra y Estados Unidos.
3. No soy esclavo de ninguna idea a priori y todo me interesa en el mundo tecnológico y humanístico.
4. Creo en la autoría y sé gestionar el fracaso.

5. En el grupo de Problemas Inversos trabajamos problemas relativos a:
   - **Inversión Geofísica y sus aplicaciones medioambientales.**
   - **Ingeniería de Reservorios y Optimización de las técnicas de producción**.
   - **Problemas de diseño industrial bajo condiciones de incertidumbre.**
   - **Técnicas de tratamiento digital de señales y Biometría.**
   - **Problemas Biomédicos y su relación con la salud y el diseño de nuevos fármacos. PROYECTO FINISTERRAE.**
   - **Problemas de predicción y análisis de incertidumbre en finanzas.**

¿PARA QUE DIANTRES SIRVEN LAS MATEMÁTICAS?

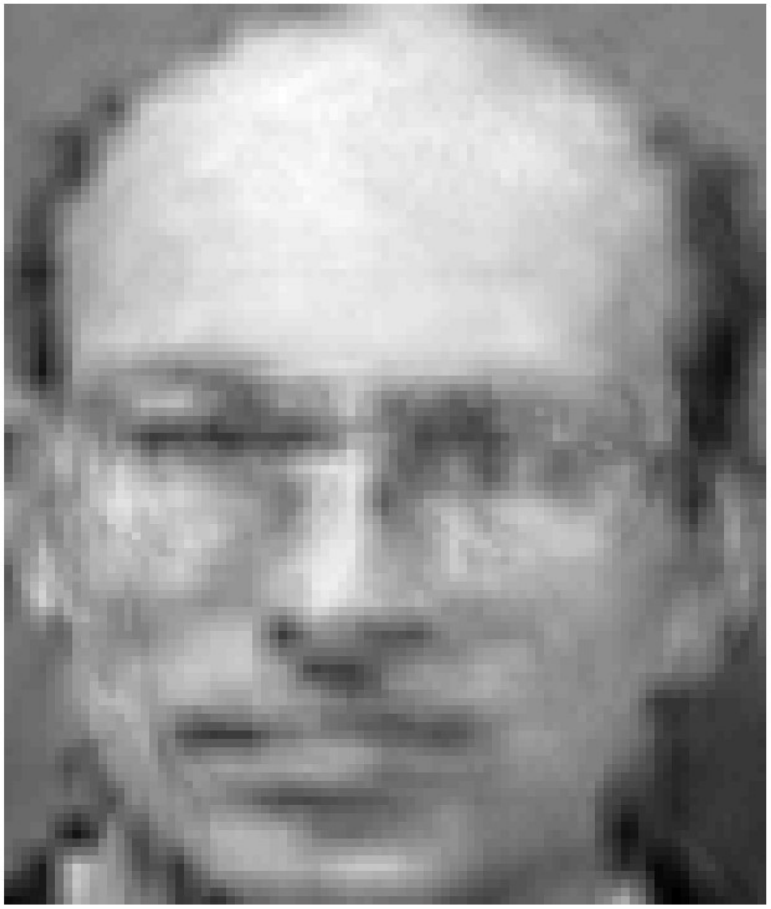
¿Para que sirve predecir?



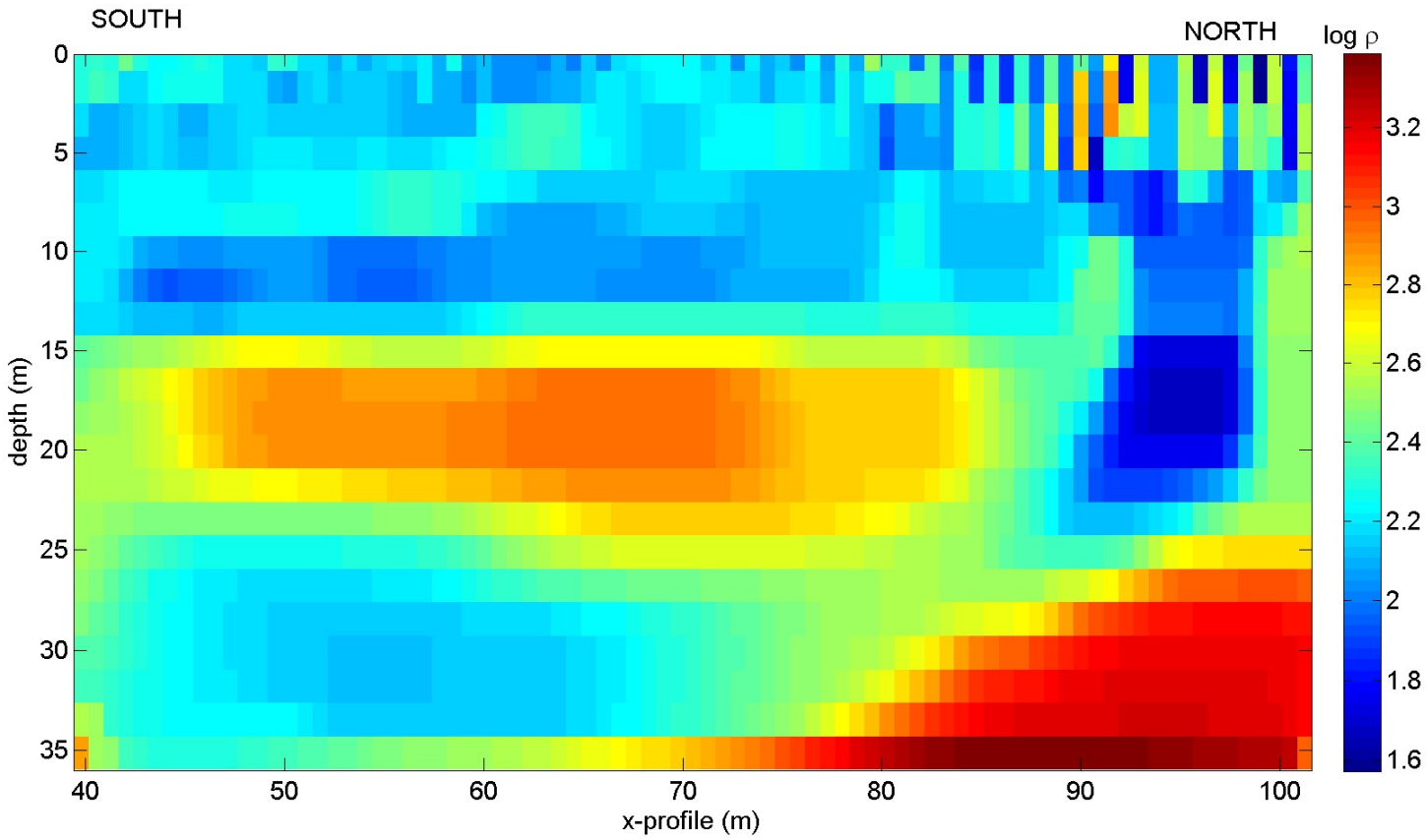
CSI

Imagen test original

Imagen test reconstruida

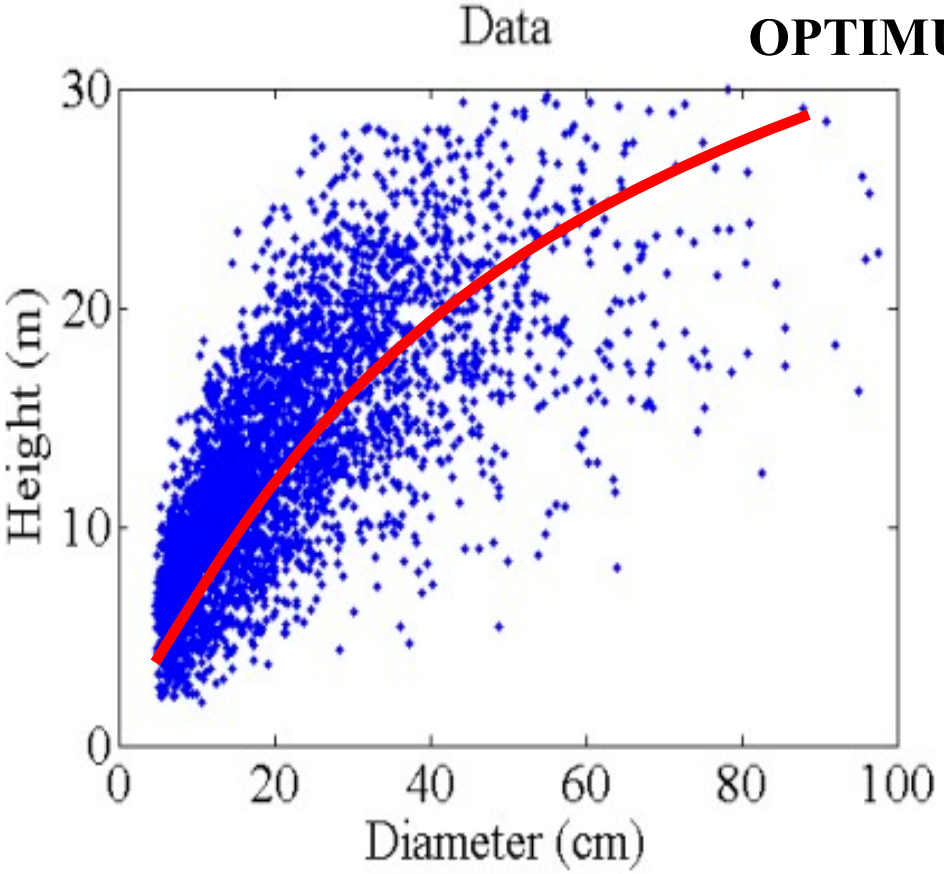


¿De dónde viene el CO_2 en la cueva de Lascaux?

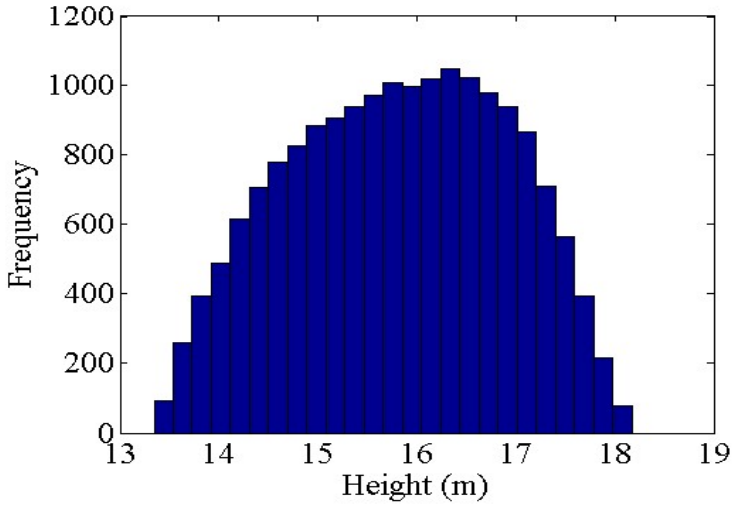


¿Cuánto mide el haya?

$$H = 1.3 + b_0\left(1 - \exp\left(-b_1 D\right)\right)^{b_2}$$

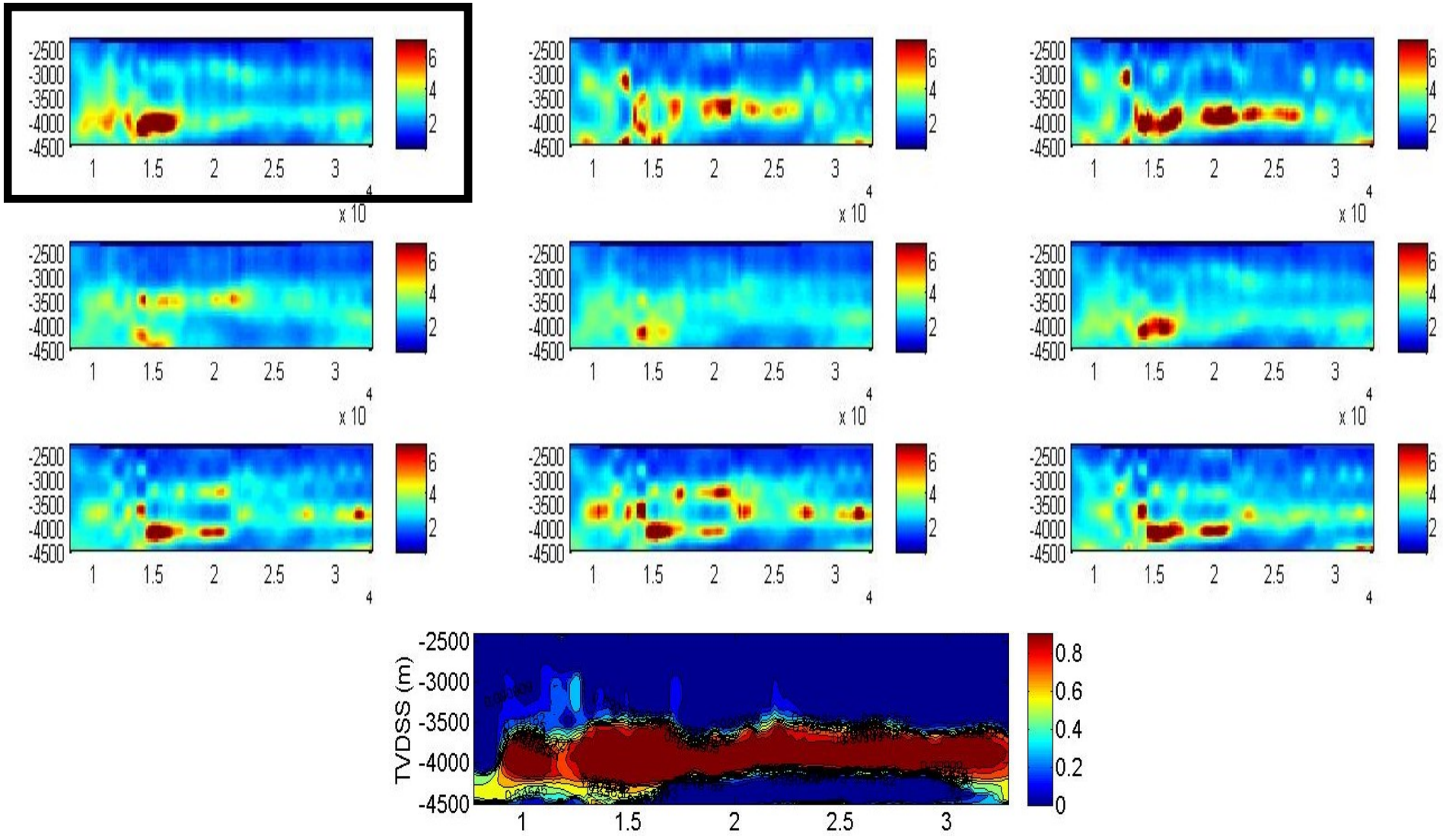


OPTIMUM FIT

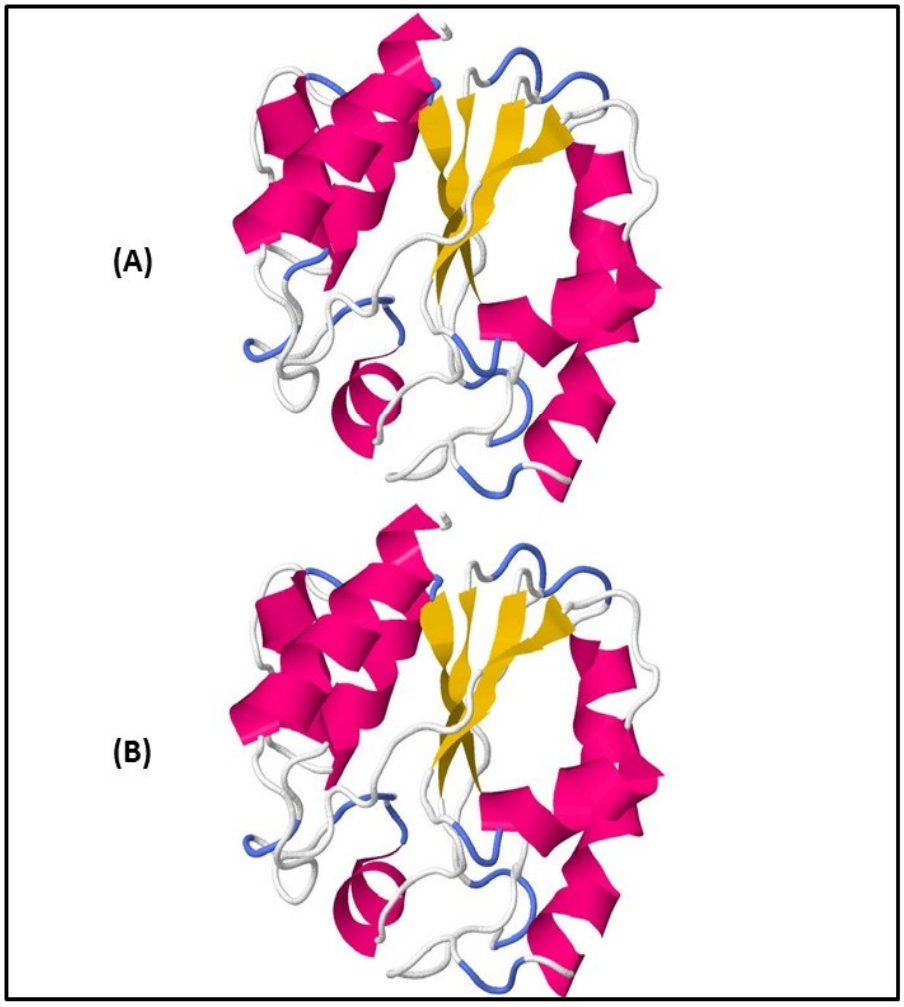


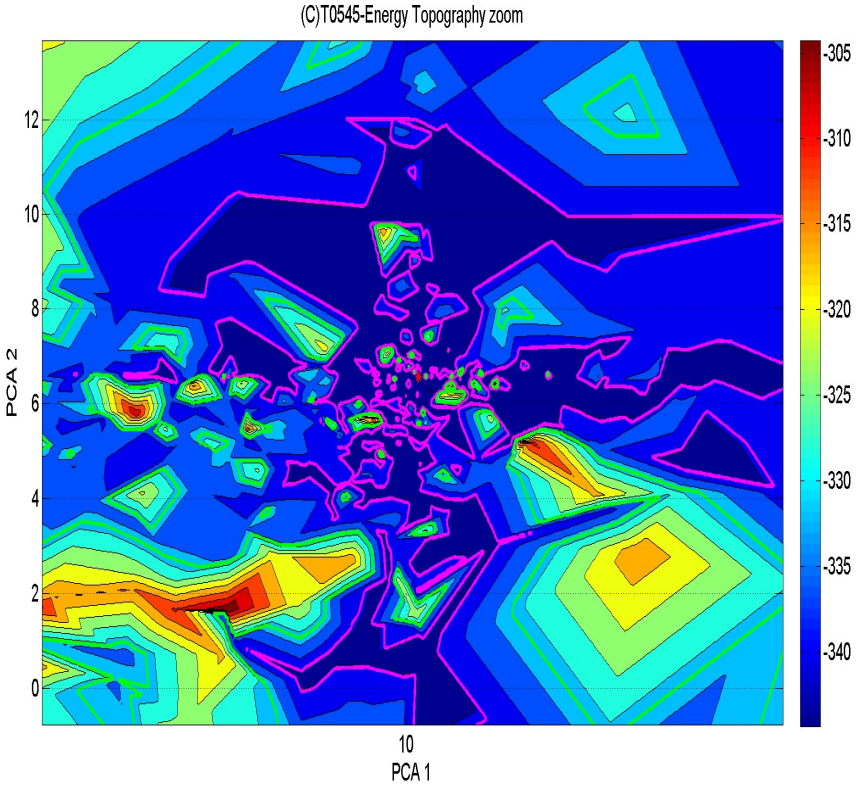
Altura para D=60

¿Qué posibilidad tengo de encontrar OIL?

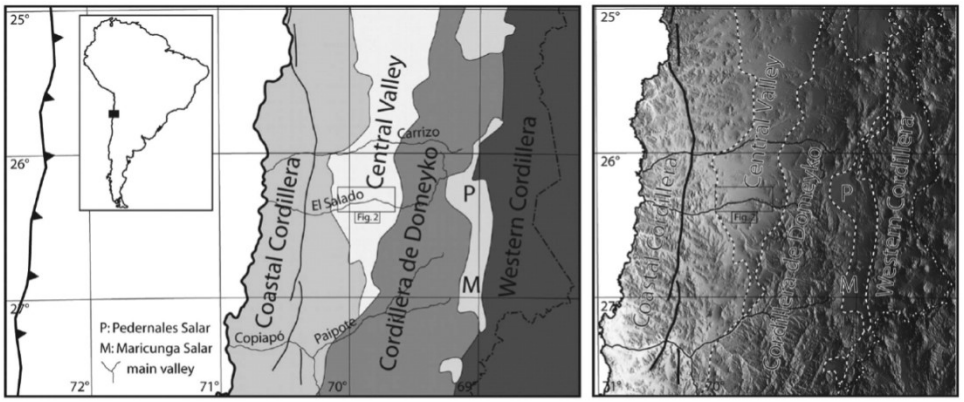


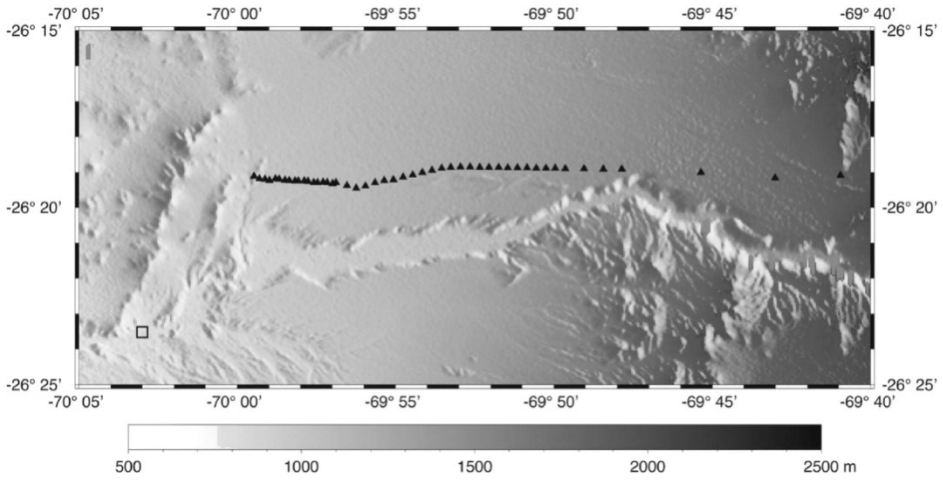
¿Cómo se predice la estructura terciaria de una proteína?

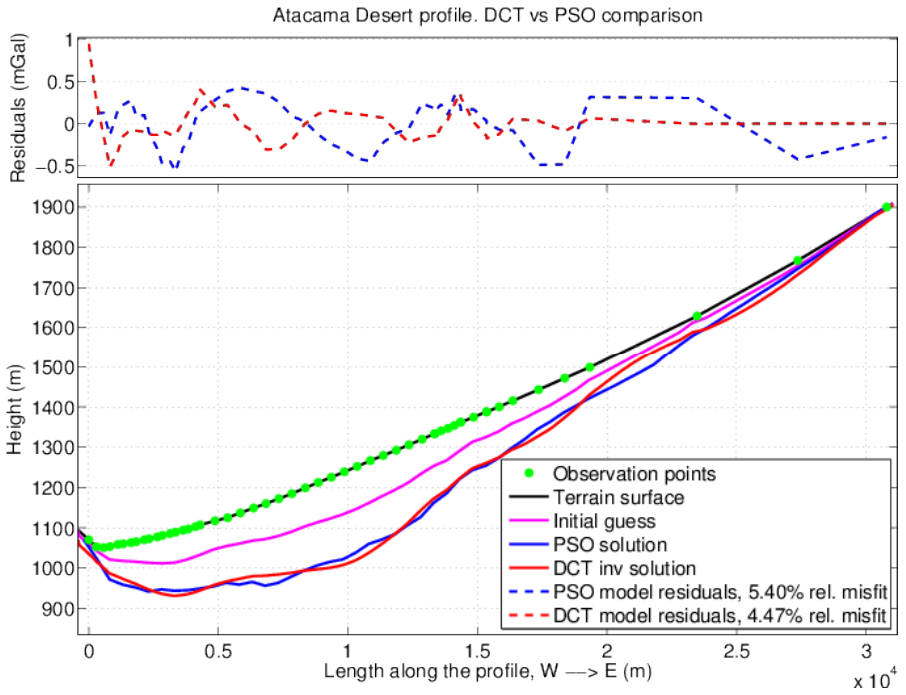




¿Qué hay debajo del desierto de Atacama?







PSO: ¿Como vuelan las manadas?

$$\mathbf{v}_i^{k+1} = \omega \mathbf{v}_i^k + \phi_{1i}^k (\mathbf{g}^k - \mathbf{x}_i^k) + \phi_{2i}^k (\mathbf{l}_i^k - \mathbf{x}_i^k),$$

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1}$$

i **particle index**
k **iteration index**

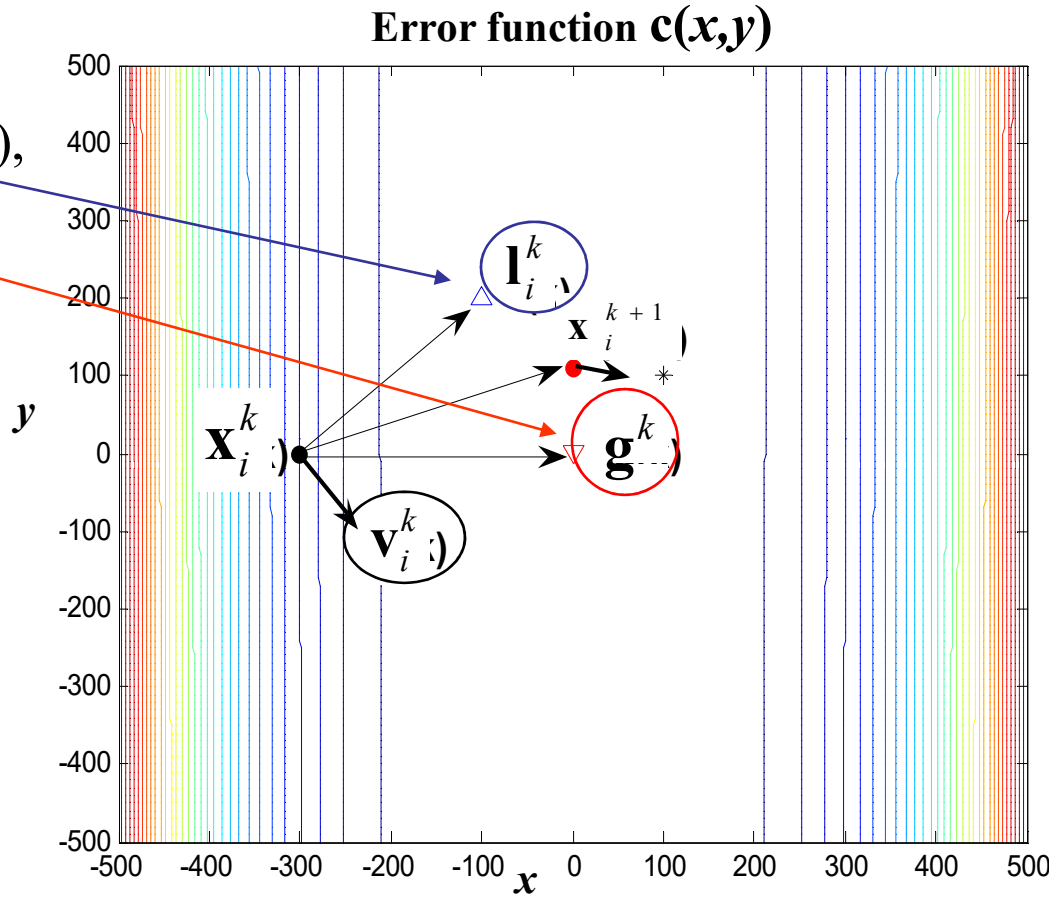
PSO PARAMETERS

ω	**Inertia weight**
a_g, a_l	**Global and Local acceleration**

Stochastic global and local acceleration $\phi_{1i}^k = r_{1i}^k a_g, \phi_{2i}^k = r_{2i}^k a_l$

Iteration dependent uniform random variables U(0,1) r_{1i}^k, r_{2i}^k

Size of the swarm N



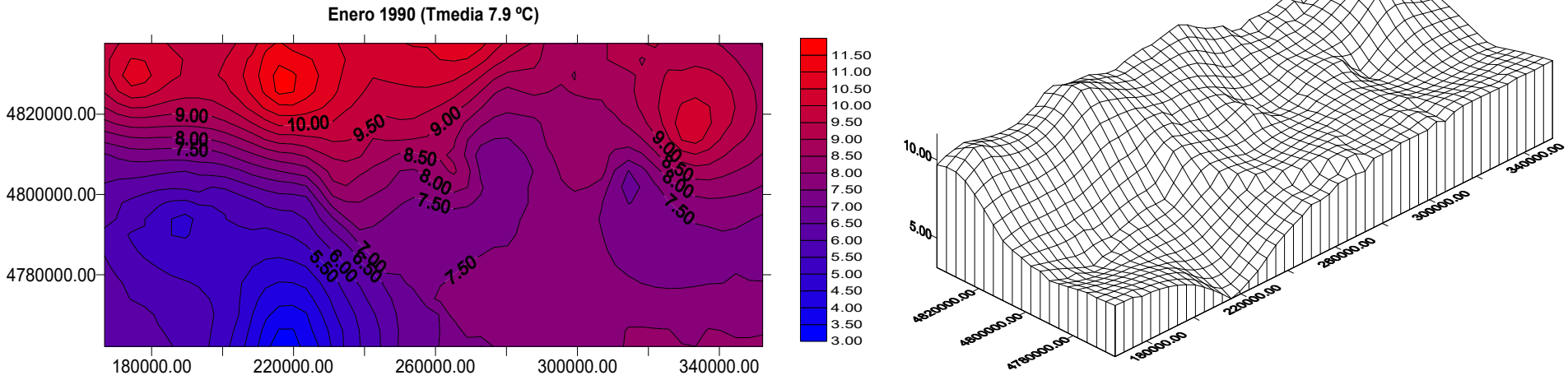
The influence of the error function enters via the local and global best

$$\mathbf{l}_i^k = \mathbf{x}_i^{k_b} : c\left(\mathbf{x}_i^{k_b}\right) = \min_{1 \le j \le k} c\left(\mathbf{x}_i^j\right) \quad i = 1, \ldots N$$

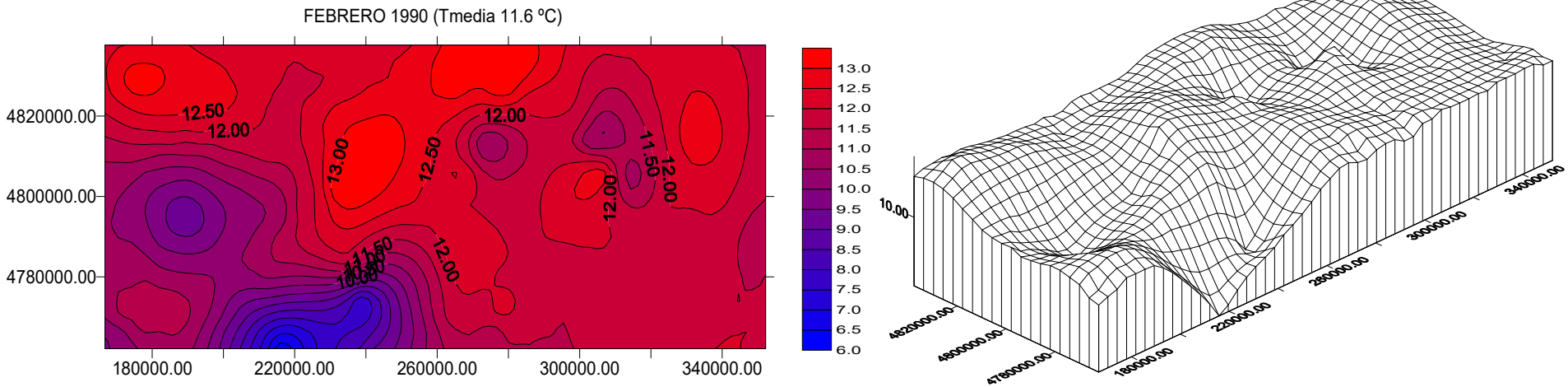
$$\mathbf{g}^k = \mathbf{l}_{i_g}^k : c\left(\mathbf{l}_{i_g}^k\right) = \min_{1 \le i \le N} c\left(\mathbf{l}_i^k\right)$$

Las 5 estaciones

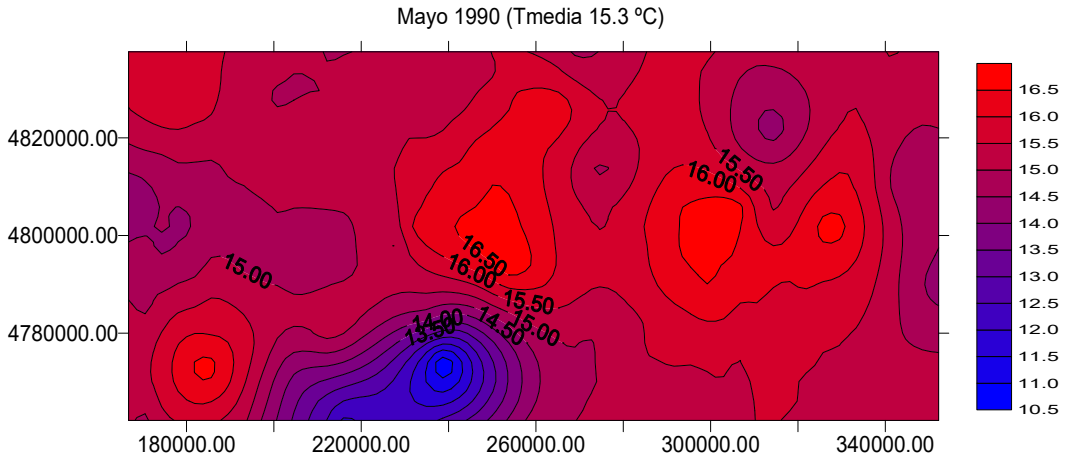
ENERO

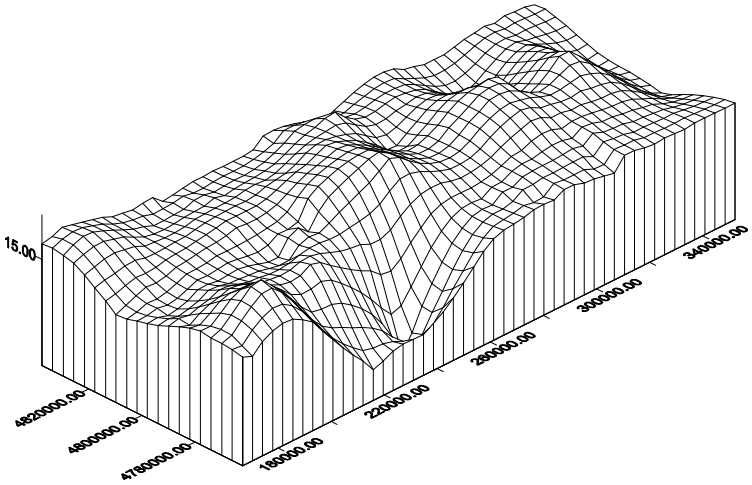


FEBRERO

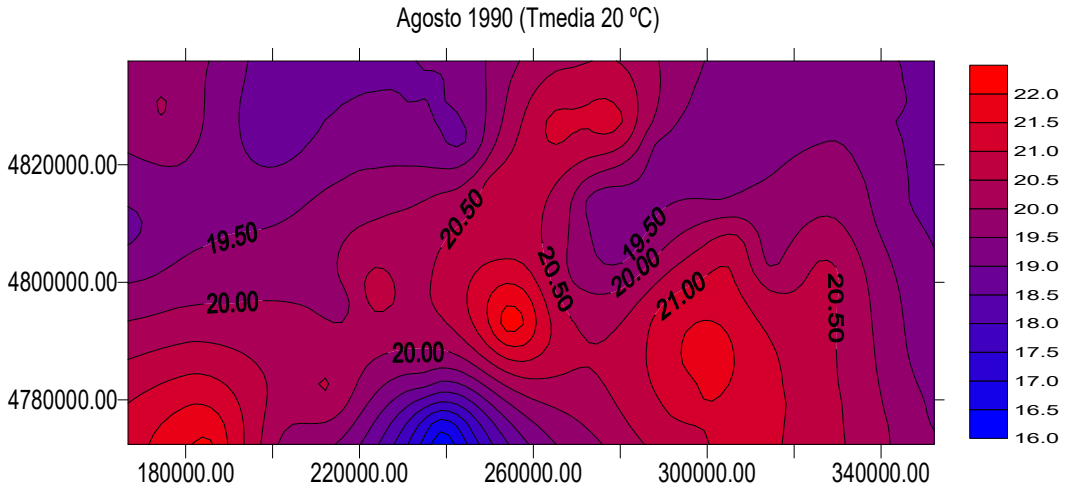


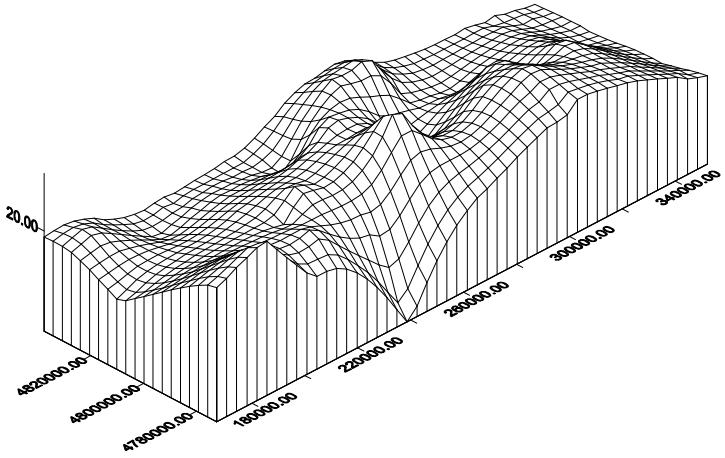
MAYO


Mayo 1990 (Tmedia 15.3 ºC)
4820000.00
4800000.00
4780000.00
180000.00
220000.00
260000.00
300000.00
340000.00
15.00
15.50
16.00
16.50
14.50
14.00
13.50
16.5
16.0
15.5
15.0
14.5
14.0
13.5
13.0
12.5
12.0
11.5
11.0
10.5

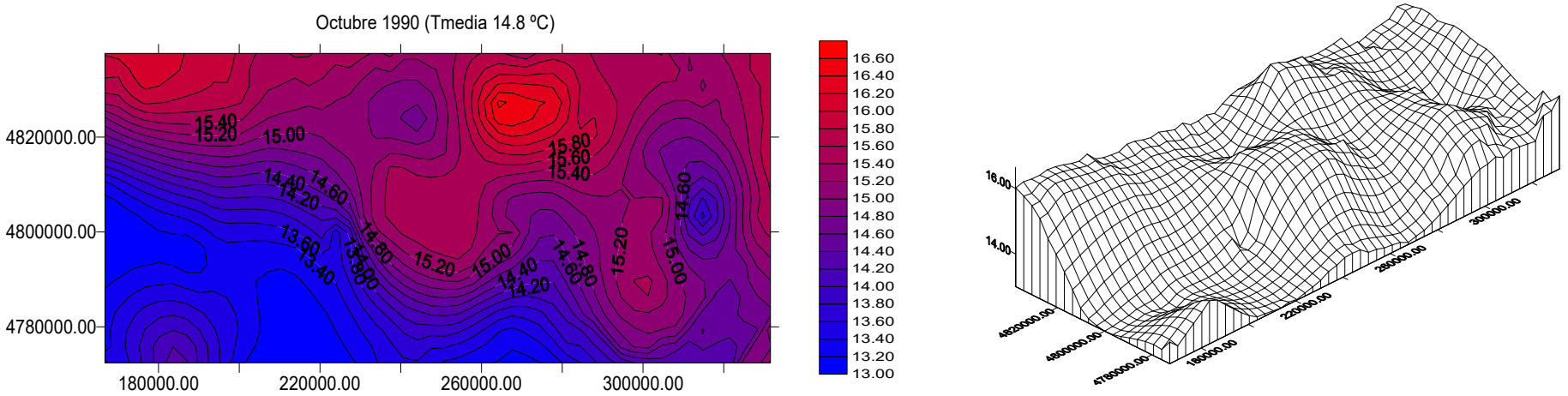

15.00
4820000.00
4800000.00
4780000.00
180000.00
220000.00
260000.00
300000.00
340000.00

AGOSTO

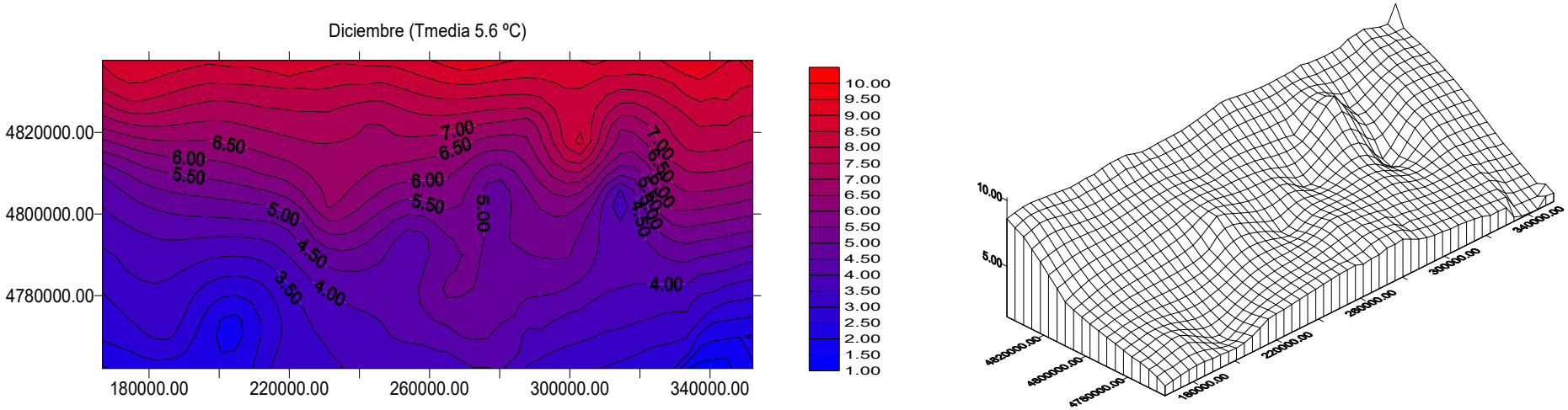

Agosto 1990 (Tmedia 20 ºC)
4820000.00
4800000.00
4780000.00
180000.00
220000.00
260000.00
300000.00
340000.00
19.50
20.00
20.50
21.00
22.0
21.5
21.0
20.5
20.0
19.5
19.0
18.5
18.0
17.5
17.0
16.5
16.0


20.00
4820000.00
4800000.00
4780000.00
180000.00
220000.00
260000.00
300000.00
340000.00

OCTUBRE



DICIEMBRE



Problemas Inversos

$$\mathbf{F}(\mathbf{m}) = \mathbf{d}$$

Forward Model

$$\mathbf{F}(\mathbf{m}) = (F_1(\mathbf{m}), F_2(\mathbf{m}), \ldots, F_s(\mathbf{m}))$$

Observed data

$\mathbf{d} \in \mathbb{R}^s$

Model Parameters

$\mathbf{m} \in \mathbb{R}^n$

$$\mathbf{m}_{\text{sol}} = \min_{\mathbf{m} \in M} \left\| \mathbf{d} - \mathbf{F}(\mathbf{m}) \right\|_{N_D}$$

El problema directo es lo que se denomina problema de predicción, es decir, dada la causa predecir el efecto.

Un problema inverso consiste en adivinar la(s) causa(s) dado el efecto y para ello es necesario resolver un sistema de ecuaciones.

El problema inverso más sencillo es la resolución de un sistema de ecuaciones lineales.

Ejemplo: el problema del muelle elástico

$$\begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix} = k \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \Rightarrow \text{Hallar } k \text{ tal que } \left\| \mathbf{F} - k\mathbf{x} \right\|_2 \text{ sea mínima.}$$

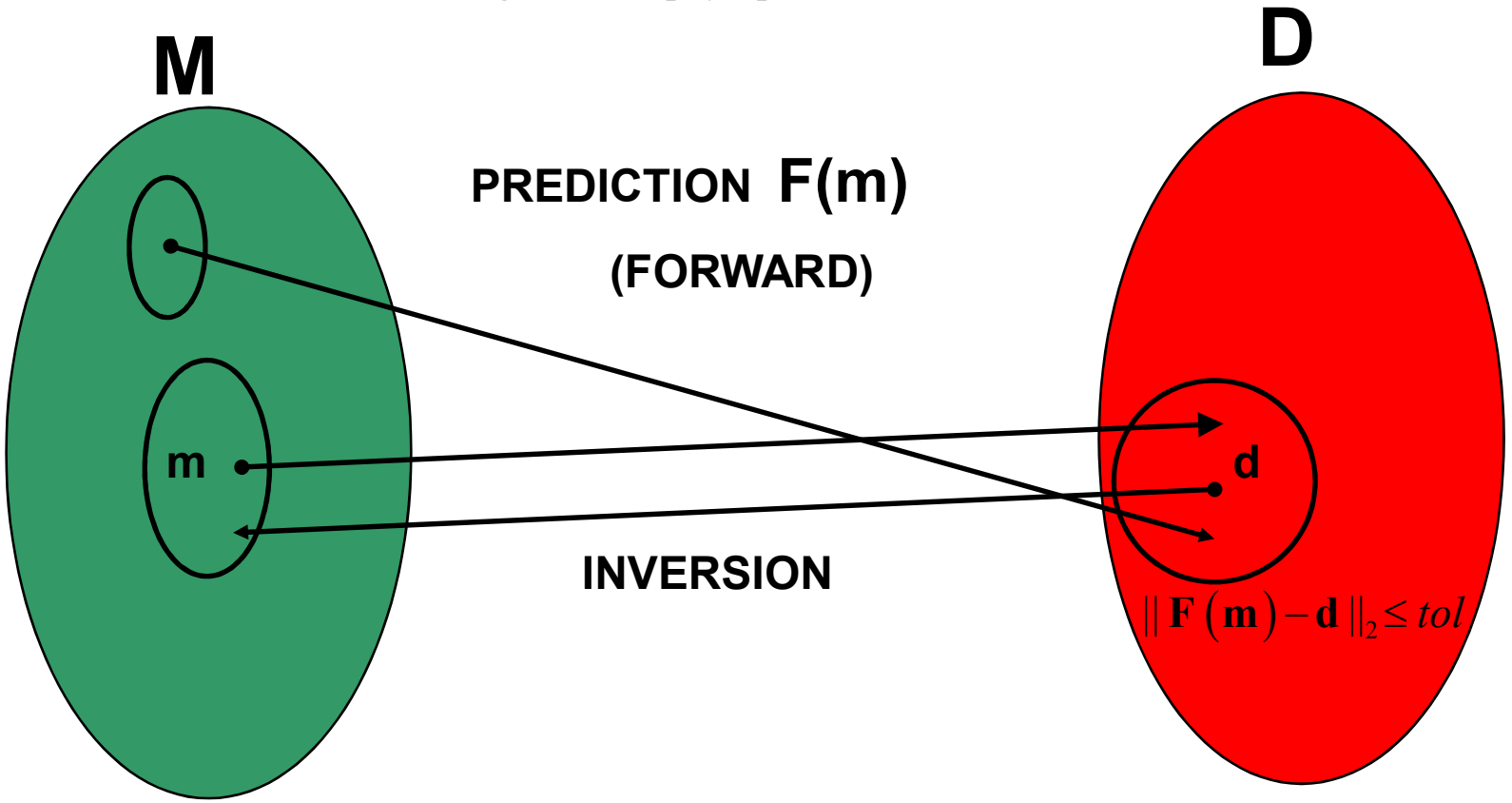
Para obtener k solo hay que proyectar ortogonalmente el vector $\mathbf{F}$ sobre el $\langle \mathbf{x} \rangle$.

El **ALGEBRA LINEAL** y la **GEOMETRÍA** son unas de las disciplinas más importantes en Modelización.

El concepto de proyección ortogonal es la base de toda la teoría de la aproximación, es decir, de la Ingeniería.

Inverse/classification problems are **ill-posed**

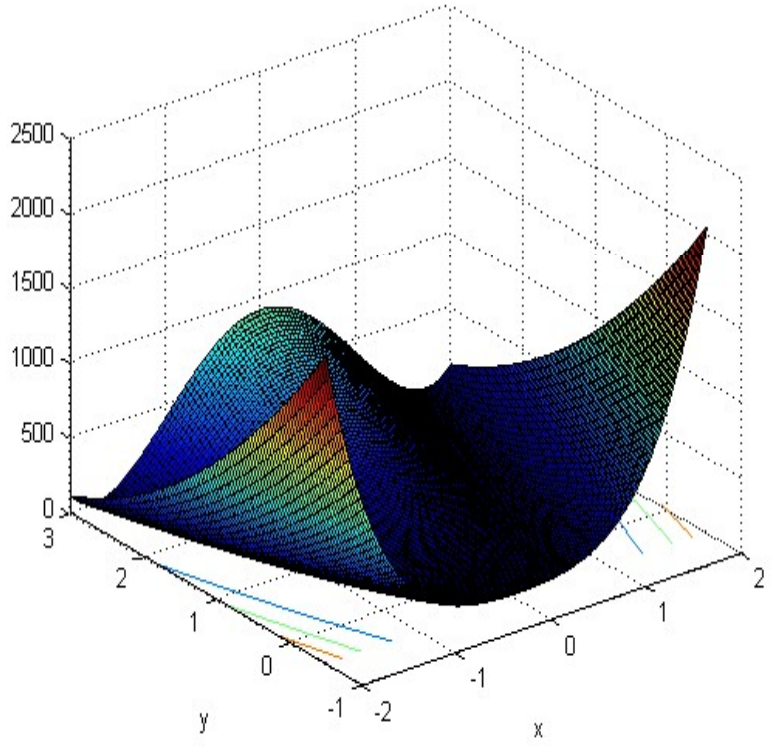
J. Hadamard, 1902. Sur les problèmes aux dérivées partielles et leur signification physique. Princeton Univ. Bull.



$$M_{tol} = \mathbf{F}^{-1}\left(|| \mathbf{F}(\mathbf{m}) - \mathbf{d} ||_2 \leq tol \right)$$

La topografía de la función objetivo $C(\mathbf{m}) = \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2$

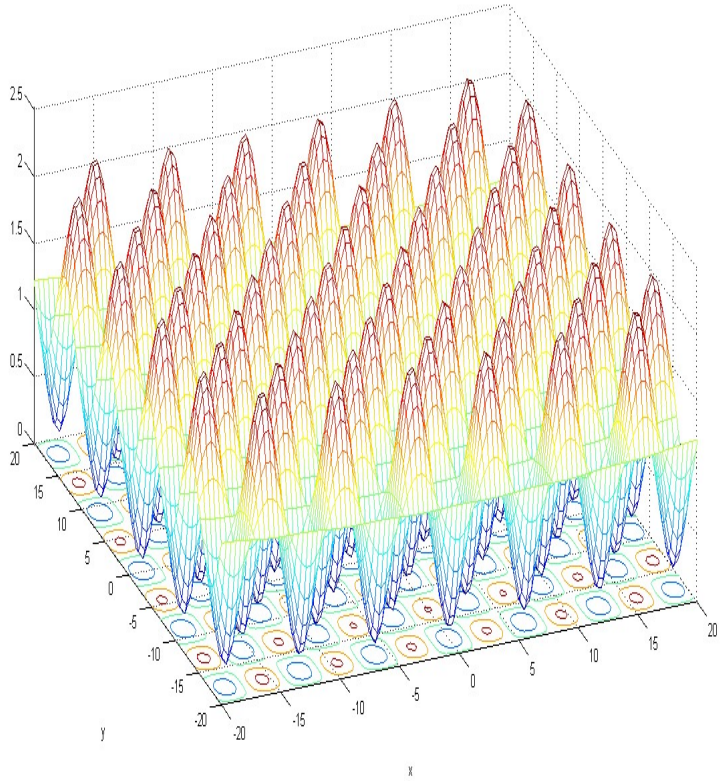
Problemas Subdeterminados



Función de tipo valle

(Rosenbrock)

Problemas Sobredeterminados
(problemas de ajuste)



Función multimodal

(Griewank)

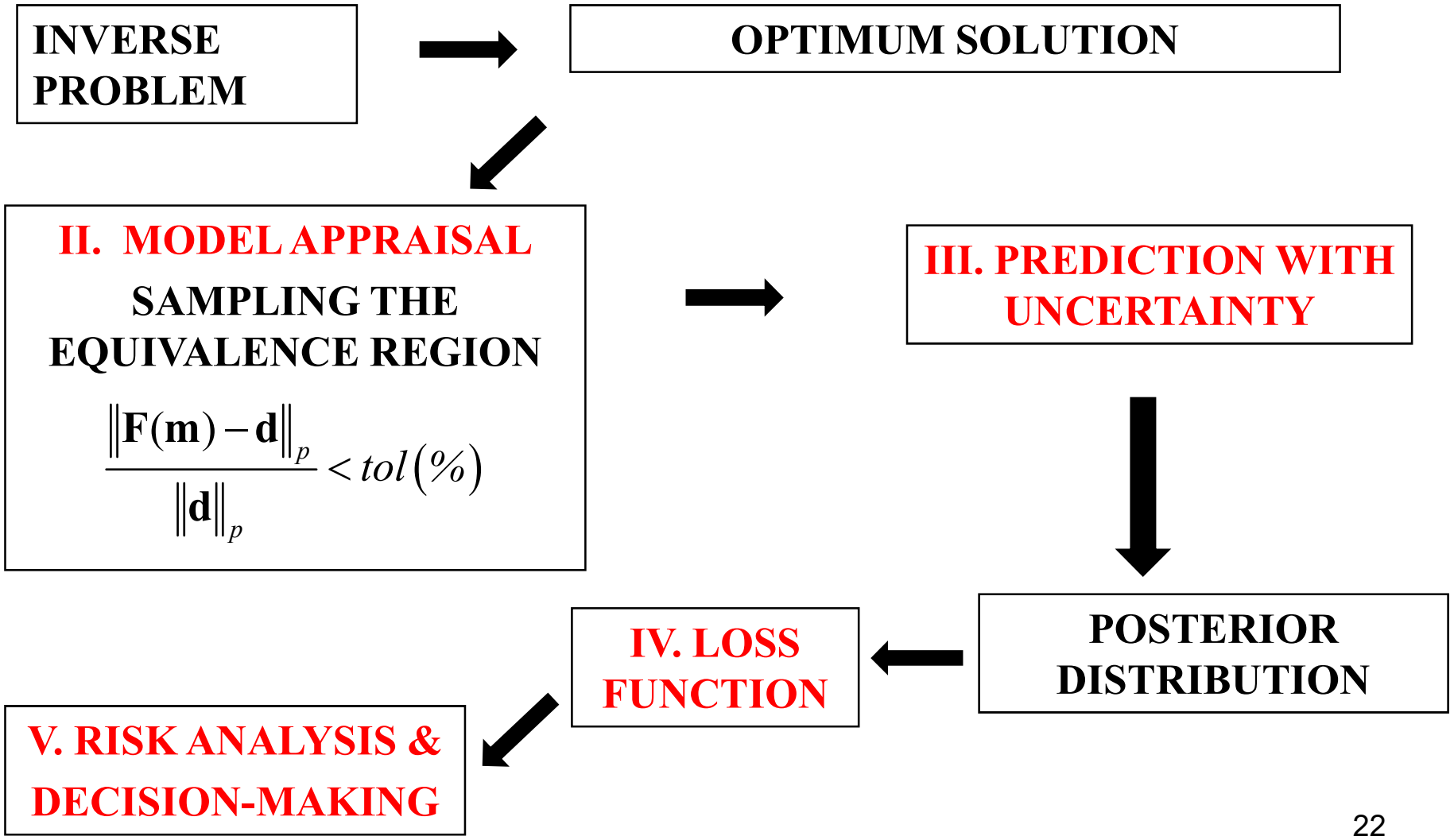
Uncertainty is inherent to any parameter identification problem due to:

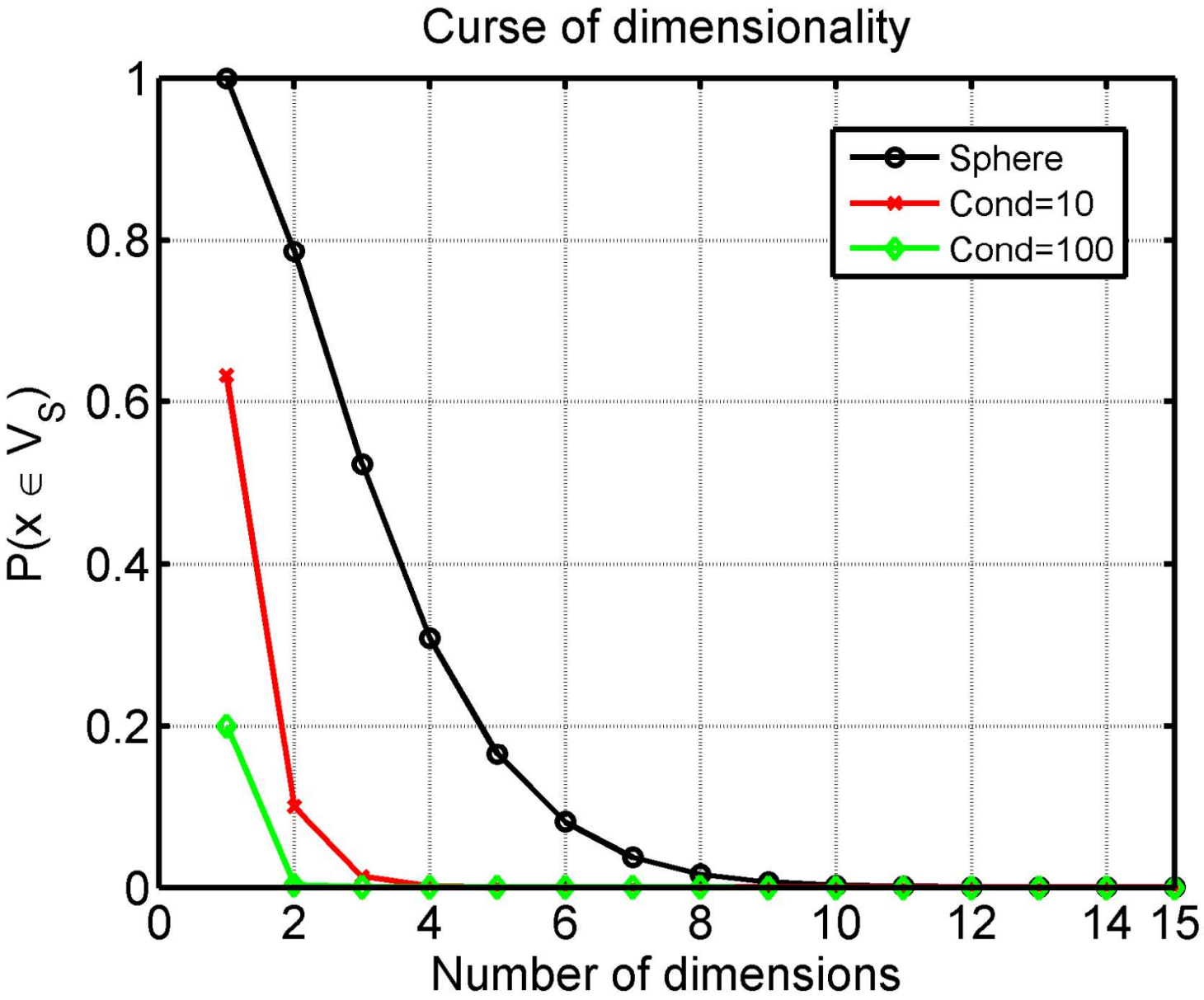
- **Noise** in observed data.
- Discrete and **incomplete data coverage**.
- **Modeling hypothesis**, **approximate physics** & **numerical approximations made in F**.

Fernández Martínez J.L., Fernández-Muñiz Z., Tompkins M., 2012. On The topography of the cost function in linear and nonlinear inverse problems. GEOPHYSICS. VOL. 77, NO. 1 (JANUARY-FEBRUARY 2012); P. 1–15.

Fernández-Martínez, J.L., Fernández-Muñiz, Z., Pallero, J.L.G., Pedruelo-González, L.M, 2013. From Thomas Bayes to Albert Tarantola. New insights to understand uncertainty in inverse problems from a deterministic point of view. Journal of Applied Geophysics, 28:62-72.

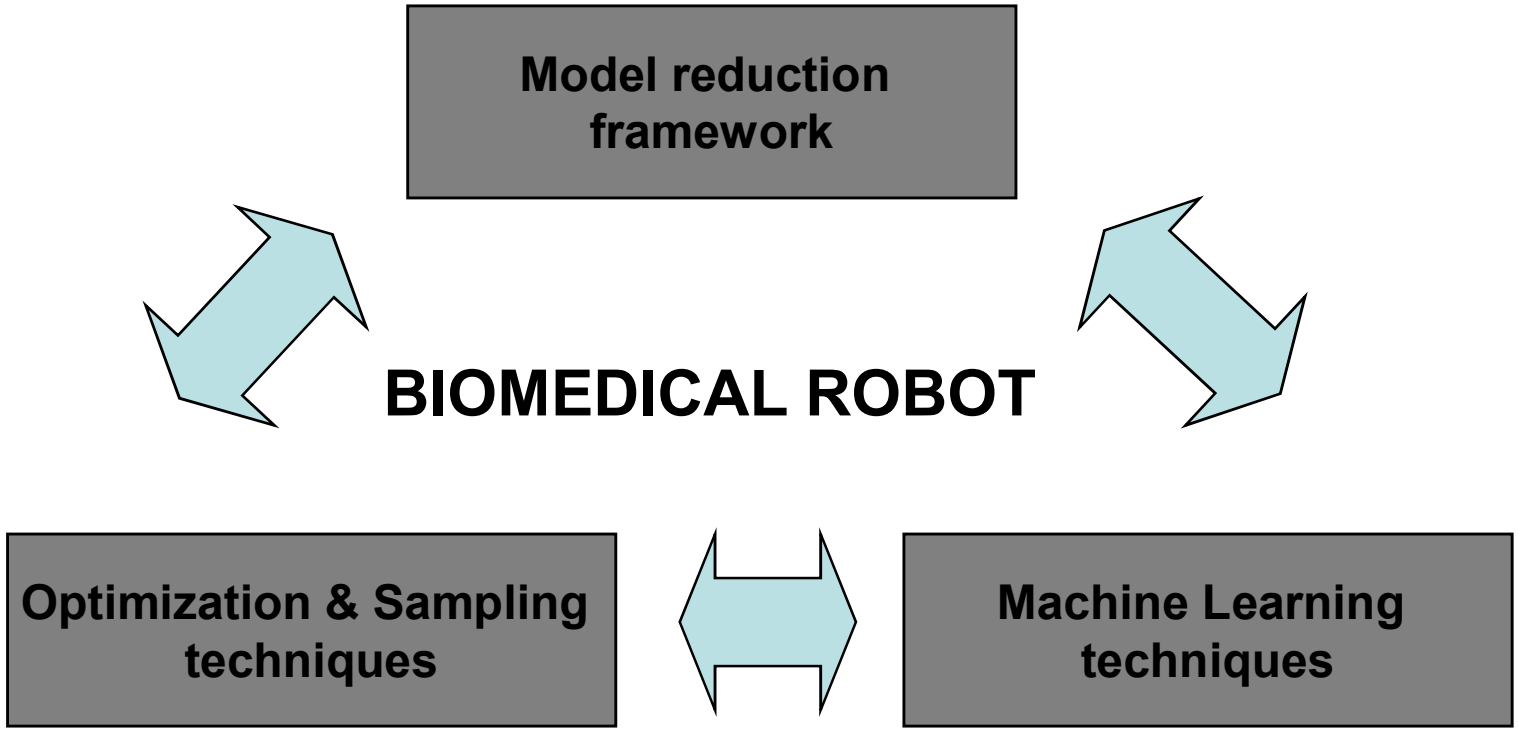
Epistemological RISK ANALYSIS (from engineering to managing)




Curse of dimensionality
Sphere
Cond=10
Cond=100
P(x ∈ V_S)
Number of dimensions
0
0.2
0.4
0.6
0.8
1
0
2
4
6
8
10
12
14
15

ROBOTS BIOMÉDICOS

How to design biomedical robots?

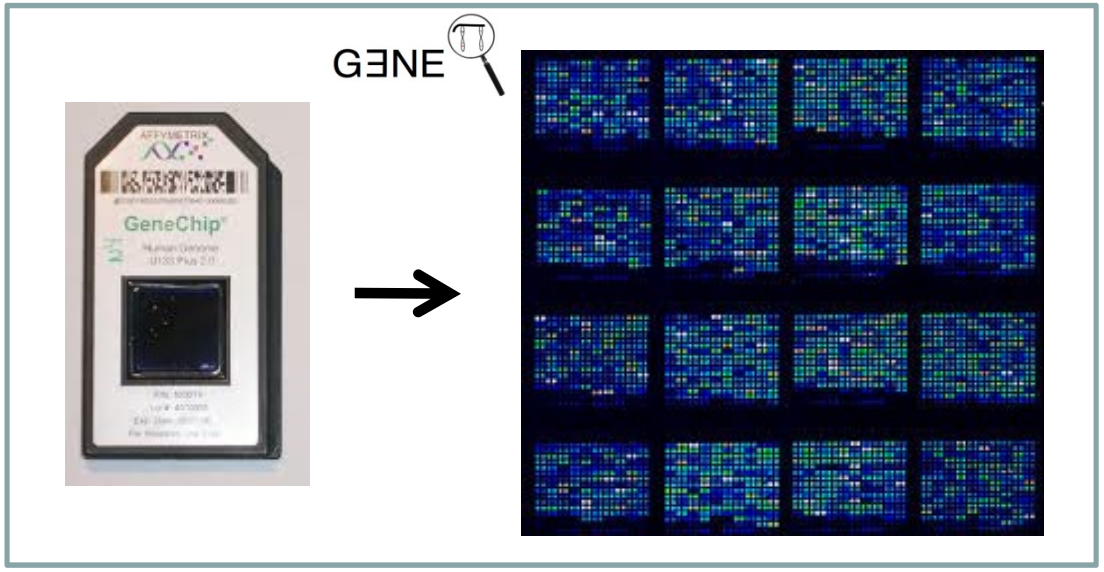


- **Dynamic learning (no fixed rules)**
- **Prediction with uncertainty**
- **Consensus decision making and risk analysis**

The challenge

Big data gathering
Genetic Data/
Hospital Data

THE ROBOT



- **Diagnosis**
- **Treatment Optmization**
- **Prognosis**

PREDICTION
+
UNCERTAINTY

MIT Technology Review

Illumina's Bid to Beat Cancer with DNA Tests

The DNA sequencing giant will launch a new company, Grail, to develop blood tests to detect cancer.

By Antonio Regalado on January 10, 2016

The world's largest DNA sequencing company says it will form a new company to develop blood tests that cost $1,000 or less and can detect many types of cancer before symptoms arise.

Some Typical Problems

- **DESIGN OF SMALL-SCALE SIGNATURES**

AIM: Finding new biomarkers in the case of cancer, rare and neurodegenerative diseases for early diagnosis, optimization of treatment and prognosis. Understanding of the biological pathways.

- **DRUG DESIGN & TREATMENT OPTMIZATION**

For every **5,000 to 10,000 potential drug candidates** that enter the discovery research stage only about **2.5 to 5%** will make it through to the **preclinical phase**. Of that percentage, only **0.05 to 0.1%** will enter the **clinical trial testing phase.**

AIM: Optimization of the pharmaco-genomics and pharmaco-cinetic aspects of drugs.

- **DESIGN OF ROBUST CLINICAL PREDICTIVE MODELS**

AIM: Development of support systems for clinical decision making and personalized treatment.

Informe
BIG DATA
Y SALUD

Sumario

INTRODUCCIÓN

1 La revolución del Big Data en la Sanidad ____ 2

SALUD

2 El poder del Real World Data
SALVADOR PEIRÓ ____ 3

3 "Si queremos hacer sostenible el sistema estamos abocados a la medicina de precisión"
BERNARDO VALDIVIESO ____ 8

4 La hora del Big Data en Salud
MIGUEL ÁNGEL MAÑEZ ____ 14

5 "Analizar millones de datos en las redes sociales es muy valioso"
PALOMA MARTÍNEZ ____ 18

6 "Tenemos que ser capaces de hacer las preguntas correctas para avanzar en Big Data"
JULIO MAYOL ____ 22

7 Retos futuros para los Sistemas de Información en el Entorno Sanitario
JOSÉ A. GUERRERO ____ 28

PRIVACIDAD

8 Los datos son del paciente ____ 34

9 "Hay que hacer cambios tecnológicos para abordar proyectos de Big Data"
LUCIANO SAEZ AYERRA ____ 40

10 La información de salud: motor de cambio de los sistemas sanitarios
JOSEP MARIA ARGIMON ____ 45

FARMACIA

11 Farmacia y Big Data
JESÚS AGUILAR SANTAMARIA ____ 51

12 "El volumen de información es crítico para el desarrollo de productos"
FEDERICO PLAZA ____ 55

13 Lo que aporta el Big Data a la industria farmacéutica
TERESA HERNANDO ____ 60

14 "Para el farmacéutico es clave conocer la situación de las ventas de productos"
CARLOS MOCHO ____ 64

TECNOLOGÍA

15 Proyectos. IBM Watson/ 23andme/PatientsLikeMe ____ 69

16 Píldoras. *Wearables* para monitorizar la salud ____ 73

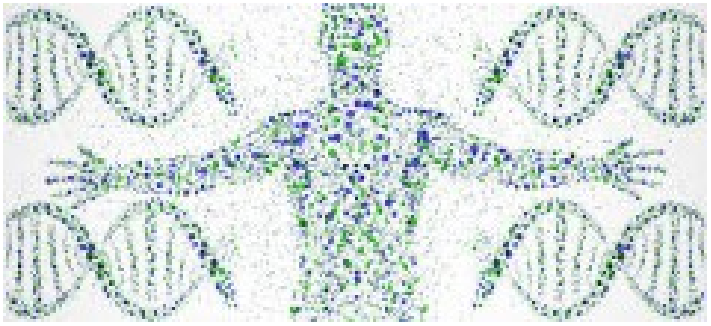
17 Mini diccionario ____ 76

El presente y futuro de la Medicina Traslacional

Gran parte de la investigación en el terreno biomédico se centra establecer puentes entre la investigación biomédica, la práctica clínica, y el tratamiento personalizado.

Existen dos tipos MERCADOS

- **Corriente arriba**: Empresas farmacéuticas y biomédicas.
- **Corriente abajo**: Hospitales y clínicas.



HOSPITALES, MINAS DE DATOS

Los hospitales son una gran mina de datos que se debe explotar para la mejora del tratamiento personalizado.

Para ello, es necesario:

1. Establecer un protocolo automático que posibilite el análisis de los datos de los pacientes (**pacientes presentes ayudan a pacientes futuros**)

2. Colaborar con equipos **médicos que crean en el big-data biomédico** para el **diagnóstico, optimización del tratamiento y la prognosis**.

3. Establecer métodos de **análisis claros, sencillos y transparentes** que proporcionen a los equipos médicos **variables pronóstico** y el **análisis de incertidumbre** en la toma de decisiones.

Consecuencias

- Optimización de recursos médicos (de modo iterativo).
- Mejora de la atención al paciente (política **zero-hero**).
- Impulso de la industria biomédica en nuestra región.

Importancia de la explotación de datos en salud

- **Mejorar la asistencia al paciente** dando **soporte a la toma de decisiones por análisis predictivos**
- **Instalar y Optimizar las políticas preventivas.**
- **Aumentar la calidad de la docencia.**
- **PERMITIR la INVESTIGACIÓN CLINICA.**
- **Diseñar acciones de salud pública.**
- **Facilitar la gestión de los servicios sanitarios.**
- **Optimización de recursos.**

* Carnicero J. y Rojas D. (Coordinadores). La explotación de datos de salud: Retos, oportunidades y límites. Pamplona: Sociedad Española de Informática de la Salud, 2016. http://www.seis.es

American College of Surgeons
Inspiring Quality: Highest Standards, Better Outcomes

ACS NSQIP®

Surgical Risk Calculator

Risk Calculator Homepage | About | FAQ | ACS Website | ACS NSQIP Website

Enter Patient and Surgical Information

Procedure

Clear

Begin by entering the procedure name or CPT code. One or more procedures will appear below the procedure box. You will need to click on the desired procedure to properly select it. You may also search using two words (or two partial words) by placing a '+' in between, for example: "cholecystectomy+cholangiography"

Reset All Selections

Are there other potential appropriate treatment options? ☐ Other Surgical Options ☐ Other Non-operative options ☐ None

Please enter as much of the following information as you can to receive the best risk estimates. A rough estimate will still be generated if you cannot provide all of the information below.

Age Group	Under 65 years	Diabetes	None
Sex	Female	Hypertension requiring medication	No
Functional status	Independent	Previous cardiac event	No
Emergency case	No	Congestive heart failure in 30 days prior to surgery	No
ASA class	I - Healthy patient		
Wound class	Clean	Dyspnea	None
Steroid use for chronic condition	No	Current smoker within 1 year	No
Ascites within 30 days prior to surgery	No	History of severe COPD	No
Systemic sepsis within 48 hours prior to surgery	None	Dialysis	No
		Acute Renal Failure	No
Ventilator dependent	No	BMI Calculation: Height (in)	
Disseminated cancer	No	Weight (lbs)	

Según la encuesta sobre el uso de Big Data del Instituto Vodafone*

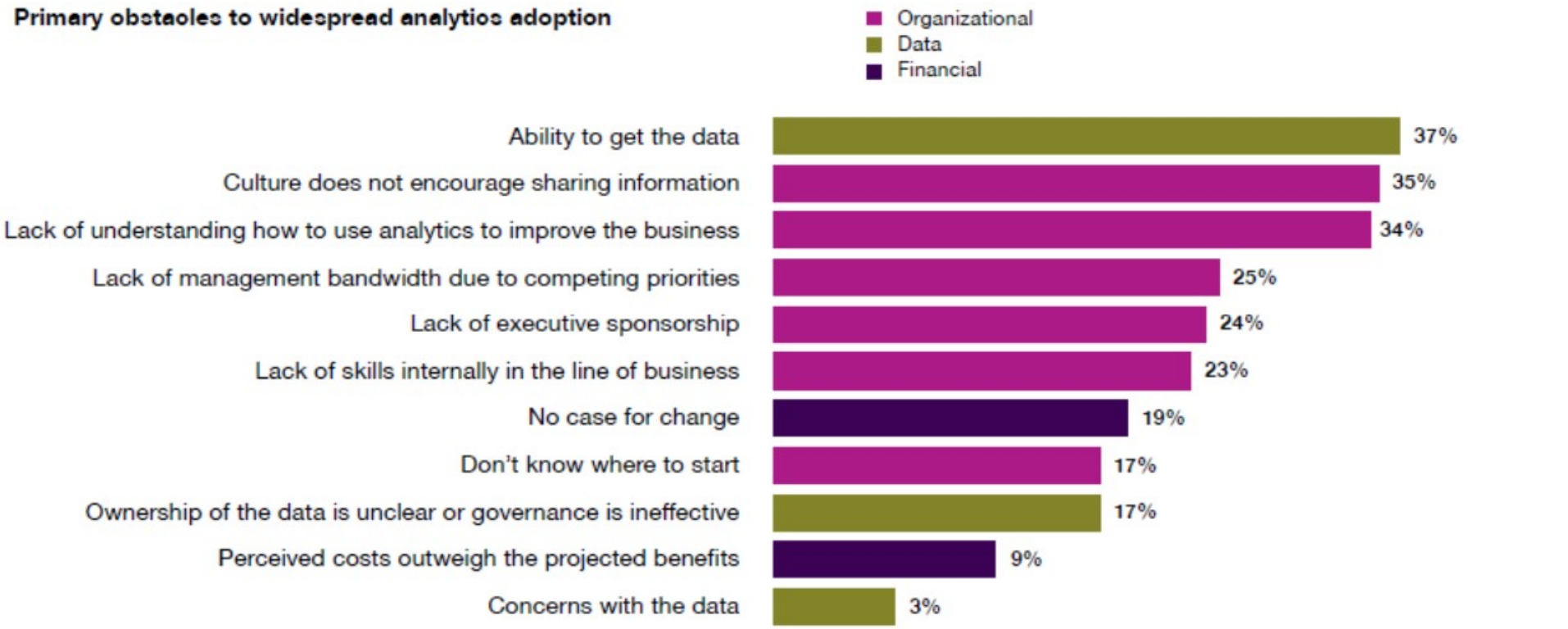
LOS CIUDADANOS QUIEREN COMPARTIR SUS DATOS

- El 86% de los ciudadanos españoles entrevistados se muestran de acuerdo con la recogida y análisis de información de salud de gran cantidad de personas

- El 69% de los encuestados en España darían, a estas instituciones sanitarias, acceso a sus datos de salud siempre que fuesen unidos a los de otros, anonimizados y usados solo para fines de investigación**.**

* Vodafone Institute "Big Data. A European survey on the opportunities and risks of data analytics".

LAS PRINCIPALES BARRERAS SON ORGANIZATIVAS



Source: Analytics: The New Path to Value, a joint *MIT Sloan Management Review* and IBM Institute of Business Value study. Copyright © Massachusetts Institute of Technology 2010.
Sample size Healthcare n= 116

Figure 4: Most of the top barriers to analytics adoption are organizational.

ALGUNAS HISTORIAS DE ÉXITO

Journal of Biomedical Informatics 60 (2016) 342–351

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

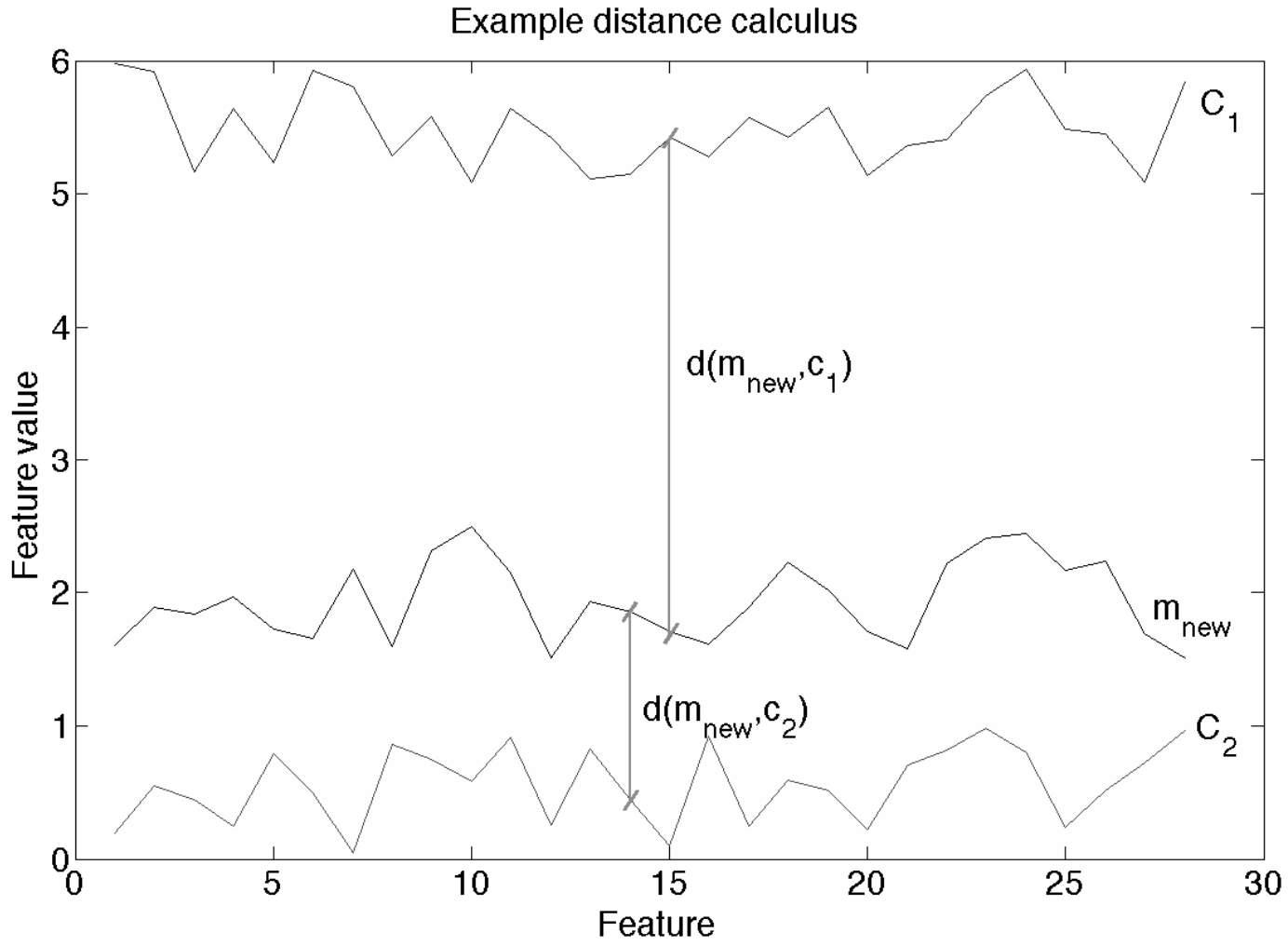
Analysis of clinical prognostic variables for Chronic Lymphocytic Leukemia decision-making problems

Enrique J. deAndrés-Galiana [a,d,1], Juan L. Fernández-Martínez [a,1,*], Oscar Luaces [d], Juan J. del Coz [d], Leticia Huergo-Zapico [c], Andrea Acebes-Huerta [c], Segundo González [c], Ana P. González-Rodríguez [b]

[a] Department of Mathematics, University of Oviedo, Spain
[b] Hematology Department, Hospital Central de Asturias, Oviedo, Spain
[c] Instituto Universitario Oncológico del Principado de Asturias (IUOPA), University of Oviedo, Spain
[d] Artificial Intelligence Center, University of Oviedo, Spain

- Análisis retrospectivo de 265 pacientes diagnosticados con Leucemia Linfocítica crónica entre 1997 y 2007 en el Hospital de Cabueñes (Dra. Ana González-Rodríguez)
- Predicción del desarrollo de la necesidad de quimioterapia y del desarrollo de enfermedades autoinmunes a la hora del diagnóstico.
- Implementación de la herramienta en una simple hoja Excel.

Un clasificador sencillo



Group	Variable name	Samp. freq. (%)
Biochemical	ALB – Albumin (g/L)	98.49
	ALC – Absolute Lymphocyte Count (cells/microL)	100.00
	ALP – Alkaline phosphatase (U/L)	95.47
	B2 M – Beta 2 Microglobulin (mg/L)	93.58
	BU – Bilirubin (mg/dL)	96.23
	CR – Creatinine (mg/dL)	99.62
	GOT – Glutamic-Oxaloacetic Transaminase (U/L)	98.11
	GPT – Glutamic-Pyruvic Transaminase (U/L)	99.25
	HGB – Hemoglobin (g/dL)	100.00
	IgA – Immunoglobulin A (g/L)	96.60
	IgG – Immunoglobulin G (g/L)	96.60
	IgM – Immunoglobulin M (g/L)	96.60
	K – Potasium (mEq/L)	90.94
	LDH – Lactate Dehydrogenase (U/L)	96.98
	MBC – Monoclonal B cell Count (cells/microL)	90.94
	MCV – Mean Corpuscular Volume (fl)	100.00
	NA (mEq/L)- Sodium	90.57
	NCC – Natural killer Cell Count (cells/microL)	90.94
	PLT – Platelets (cells/microL)	100.00
	RET – Reticulocyte count (cells/microL)	75.47
	SNC – Segmented Neutrophils Count (cells/microL)	100.00
	T8C – CD8 T cell Count (cells/microL)	86.42
	TLC – Total Lymphocyte Count, CD8 + CD4 (cells/microL)	96.60
	UA – Uric acid (mg/dL)	97.36
	UR – Urea (mg/dL)	99.25
	WBC – White Blood cells Count (cells/microL)	100.00
CLL Specific	**CD38** – CD38 positive	81.51
	COOMBS – Coombs test	94.34
	LD – Time for duplication of the number of lymphocytes	96.98
	MOR – Morphology	98.49
	MP – Monoclonal Peak	98.87
	NLymph – Number of affected lymph nodes	99.62
	SMG – Splenomegaly	99.62
	ZAP70 -Zeta-chain-associated protein kinase 70 (%)	21.89
Personal	AGE – Age	100.00
	SEX – Sex	100.00

- Las variables pronóstico son de uso común en hematología.
- La frecuencia del muestreo no es del 100% en todos los casos.
- **ETAPAS DE LA MODELIZACIÓN**
  - **FILTRADO E IMPUTACIÓN**
  - **DEFINICIÓN DEL PODER PREDICTOR**
  - **ANÁLISIS DE LA MATRIZ DE CONFUSIÓN**
  - **INTERPRETACIONES MÉDICAS**

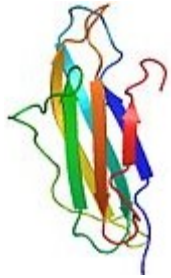
Necesidad de Quimioterapia

PLEASE, INTRODUCE THE VALUES FOR THE NEW SAMPLE

	Class 1: CT	Class 2: No CT	W	Pth	New Sample		Weighted distance to CT
B2M mg/L	4,2	2,2	99,991	0,47	3,8		1937,6
WBC cells/microL	61767,8	16783,1	0,034		50000,0		160083,6
ALC cells/microL	47555,3	11189,8	0,035		38000,0		111846,8
MBC cells/microL	40270,3	8391,7	0,021		40000,0		32,2
				Prediction	**Class 1: CT**	D1	523,4
				P1	0,76		
				P2	0,24		

L E G E N D		Prognostic variables for Chemotherapy Treatment prediction
		Class Data: Mean signatures for Chemotherapy Treatment and No Chemotherapy Treatment groups, the weight vector **W**, and the probal
		Input Data: New sample (m_{new}): **the user should introduce the values of the prognostic variables for the new sample to be diagnos**
		Intermediate Result: Individual Weighted distances
		Result: Total distance between the new sample and the corresponding class
		Result: probabilities to belong to class 1 or 2.
		Result: Class Prediction

- **B2M**: Beta2-Microglobulina: el incremento de ß2M es debido al aumento de linfocitos T citotóxicos. Indicador de mal pronóstico en leucemias (<2 **mg/l**)
- **WBC** (White Blood cell count): contaje de leucocitos (4000-11000 **cel/ml**). Defensa del sistema inmunitario contra infecciones.
- **ALC**: Absolute lymphocyte count (1300 and 4000 cel/ ml.): T-cells, B-cells, and NK cells.
- MBC: Monoclonal B-cell count



Desarrollo de enfermedades autoinmunes

PLEASE, INTRODUCE THE VALUES FOR THE NEW SAMPLE

	Class 1: AD	Class 2: No AD	W	Pth	New Sample		Weighted distance to AD		Weighted distance to No AD
PLT cells/microL	94985,0	202242,7	0,0003		190000,0		812,5		13,5
RET cells/microL	135730,0	69818,3	0,0005		75000,0		922,0		6,7
ALB g/L	40,4	37,4	5,7884		35,0		987,9		189,6
HGB g/dL	11,5	13,6	7,7686		15,0		756,3		116,6
BU mg/dL	1,1	0,6	16,461		0,8		18,3		10,6
UR mg/dL	64,1	46,7	1,3185		50,0		345,6		18,9
MCV fl	98,1	89,6	4,9456	0,5	92,0		904,2		136,5
NCC cells/microL	2250,7	741,1	0,0136		1000,0		289,3		12,4
K mEq/L	4,1	4,3	63,5621		4,4		388,3		18,0
WBC cells/microL	55999,0	24689,9	0,0002		26000,0		36,0		0,1
LD	398,1	343,4	0,1395		350,0		45,0		0,8
ALC cells/microL	42213,0	17780,3	0,0012		19000,0		775,9		2,1
MBC cells/microL	36333,4	14229,3	0,0005		16000,0		103,4		0,8
				Prediction	**Class 2: No AD**	D1	71,0	D2	22,5
				P1	0,24				
				P2	0,76				

Importancia del contaje de plaquetas, reticulocitos y células NK que son los principales objetivos de las enfermedades autoinmunes (anemia hemolitica y trombocitopenia).

Clin Transl Oncol
DOI 10.1007/s12094-015-1285-z

RESEARCH ARTICLE

On the prediction of Hodgkin lymphoma treatment response

E. J. deAndrés-Galiana[1,2] · J. L. Fernández-Martínez[1] · O. Luaces[2] · J. J. del Coz[2] · R. Fernández[3] · J. Solano[4] · E. A. Nogués[4] · Y. Zanabilli[5] · J. M. Alonso[6] · A. R. Payer[4] · J. M. Vicente[7] · J. Medina[5] · F. Taboada[8] · M. Vargas[9] · C. Alarcón[5] · M. Morán[5] · A. González-Ordóñez[5] · M. A. Palicio[9] · S. Ortiz[9] · C. Chamorro[10] · S. Gonzalez[11] · A. P. González-Rodríguez[4]

- Análisis retrospectivo de 263 pacientes diagnosticados con Linfoma de Hodgkin entre 2002 y 2012 en los servicios de hematología de los diferentes hospitales de la red de Hospitales públicos de Asturias coordinados por la Dr. A. González Rodríguez).
- Predicción de la respuesta al tratamiento a la hora del diagnóstico.
- Implementación de la herramienta en una simple hoja Excel.

- **IMPORTANCIA DE LA FERRITINA, ALANINA TRANSAMINASA (ALT) y de la FOSFATASA ALCALINA (ALP)**

Table 5 Best results for all the comparisons obtained after weight optimization

#	Comparisons	Base	Rate	TPR (%)	TNR (%)	FPR (%)	FNR (%)	FDR (%)	Acc (%)
1	CR (+) vs. PR and PD (−)	SF	67.2818	98.73	23.08	76.92	1.27	7.87	91.2548
		GPT	41.0086						
2.1	CR and PR (+) vs. PD (−)	SF	75.2264	98.43	33.33	66.67	1.57	2.34	96.1977
2.2	CR (+) vs. PR (−)	SF	57.7157	99.58	11.76	88.24	0.42	5.98	93.7008
		ALT	41.3166						
		ALP	38.9228						

Rate is the maximum percentile distance rate, *TPR* true-positive rate, *TNR* true-negative rate, *FPR* false-positive rate, *FNR* false-negative rate and *Acc* final accuracy of the prediction. Signs (+) and (−) represent the positive and negatives groups, respectively

Table 6 Mean values of the true positives, true negatives, false positives and false negatives and optimized weights (ω_a) of the optimum NN classifier after weight optimization, for all the comparisons

#	Comparisons	Base	TP	TN	FP	FN	ω_a
1	CR (+) vs. PR and PD (−)	SF	254.7	**2309.3**	338.6	3007.0	0.0016
		GPT	24.4	**99.2**	23.2	55.3	0.0319
2.1	CR and PR (+) vs. PD (−)	SF	275.4	**2796.7**	225.5	2669.5	0.0020
2.2	CR (+) vs. PR (−)	SF	276.7	**2401.0**	405.5	3330.0	0.0026
		ALT	**24.3**	18.0	44.2	140.0	0.0663
		ALP	123.2	**376.0**	163.5	1059.0	0.0051

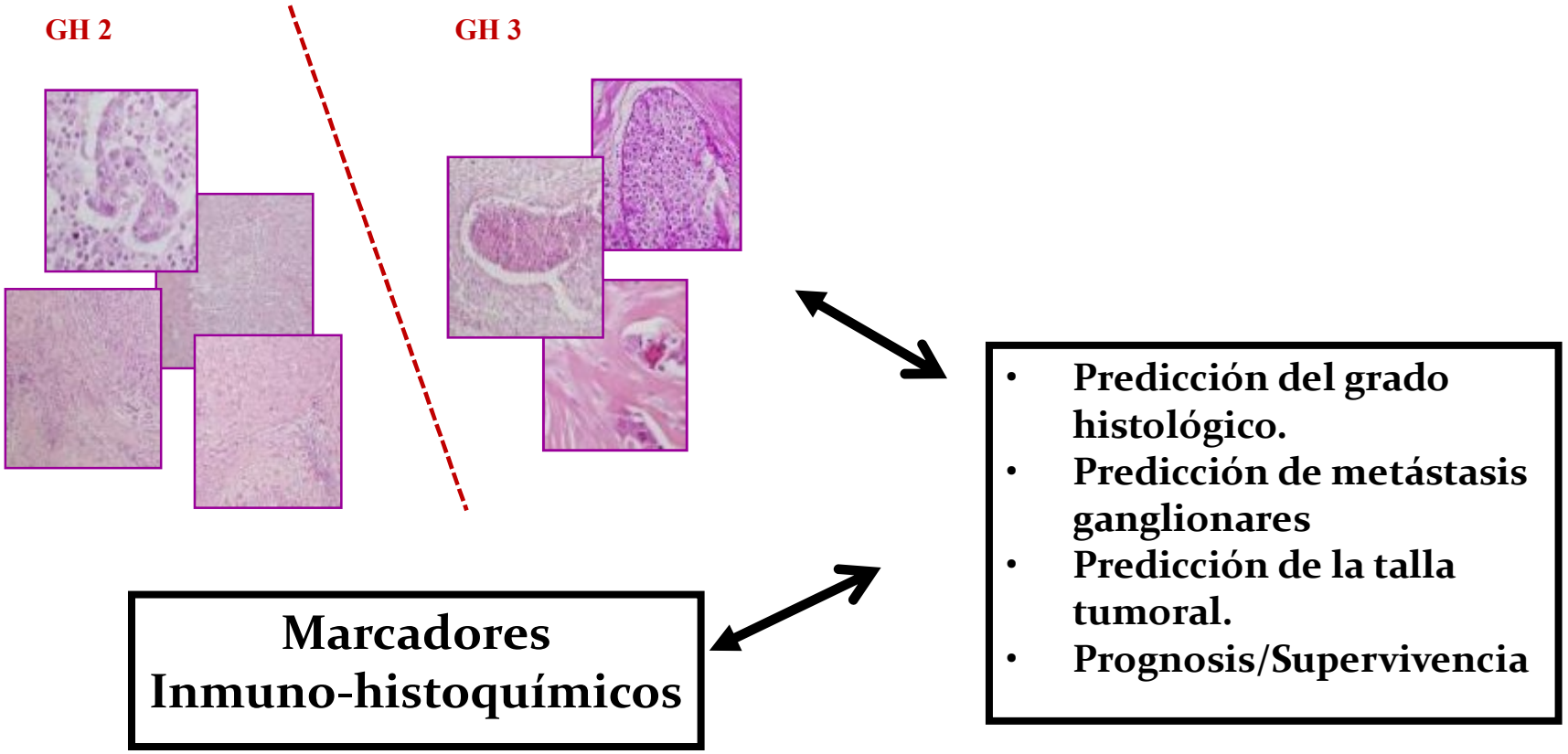
Signs (+) and (-) represent the positive and negatives groups, respectively. Boldfaces indicate the highest value for each prognostic variable. Normal bounds for the decision correspond to the TP and TN groups

ALT Alanine aminotransferase (U/L)

ALP Alkaline phosphatase (U/L)

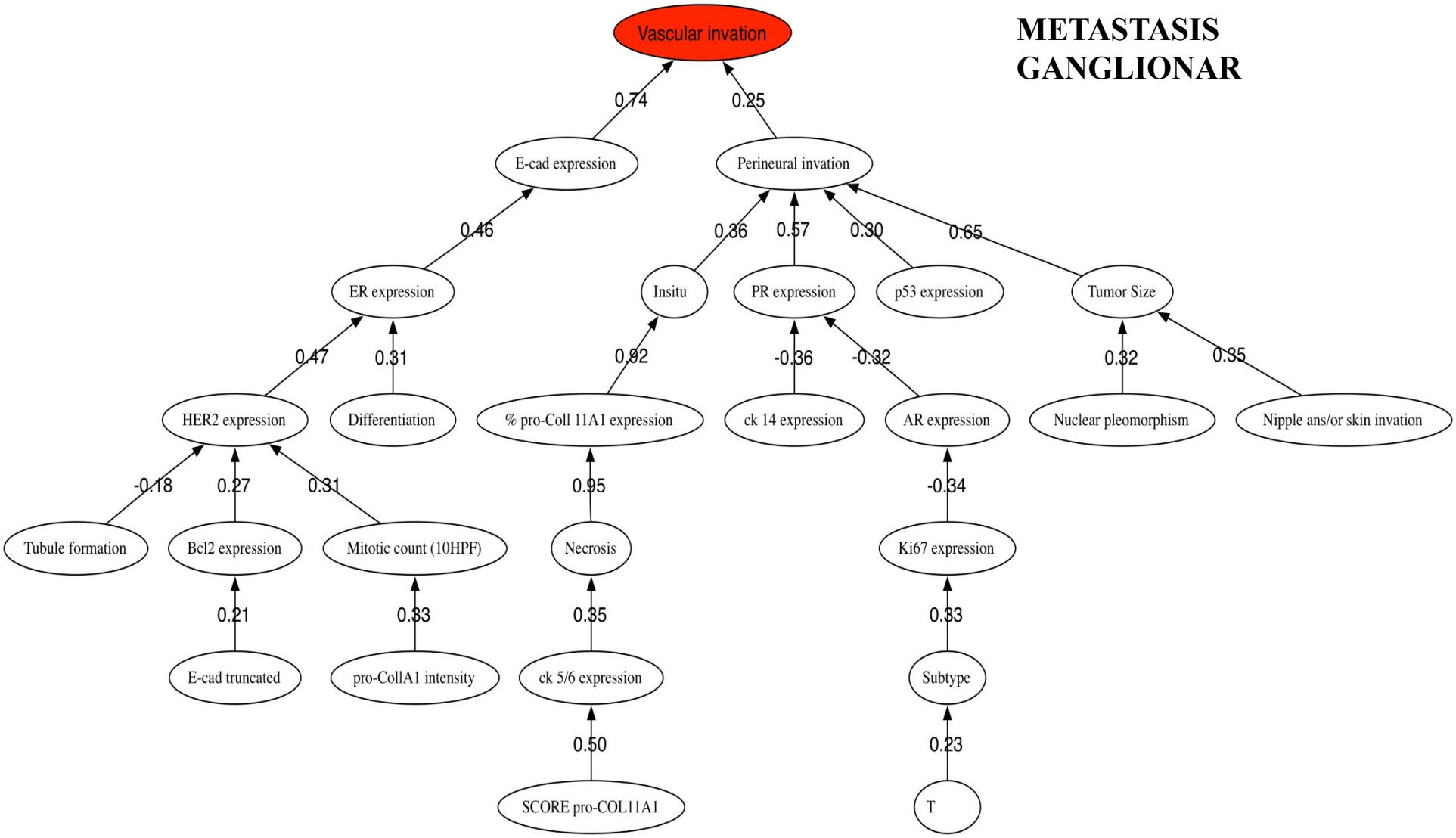
SF Serum ferritin (ng/mL)

Canceres de mama triplemente negativos



- Análisis retrospectivo de 105 pacientes diagnosticados con TBNC en el HUCA (Drs. Carmen García-Pravia y Primitiva Menéndez).

METASTASIS GANGLIONAR



Haciendo el Big-data pequeño

FINISTERRAE

FINISTERRAE

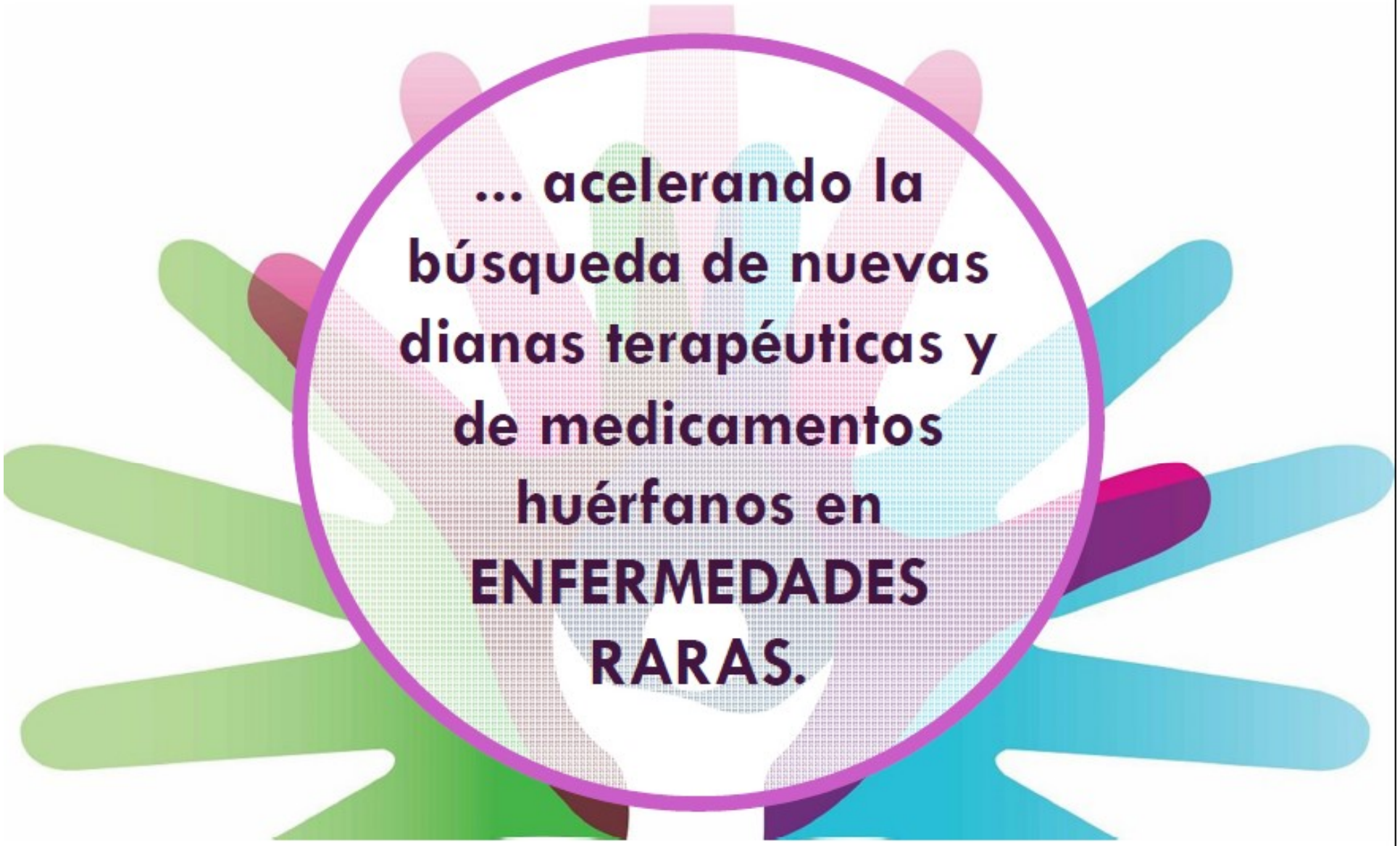
FINISTERRAE

BigData para la optimización de fármacos en enfermedades raras, neurodegenerativas y cáncer.

"Esta generación tiene la oportunidad histórica y la responsabilidad de transformar la Medicina utilizando enfoques sistemáticos en Ciencias Biológicas para acelerar drásticamente la comprensión y el tratamiento de enfermedades"
The Eli and Edythe L. Broad Institute
Of MIT and Harvard

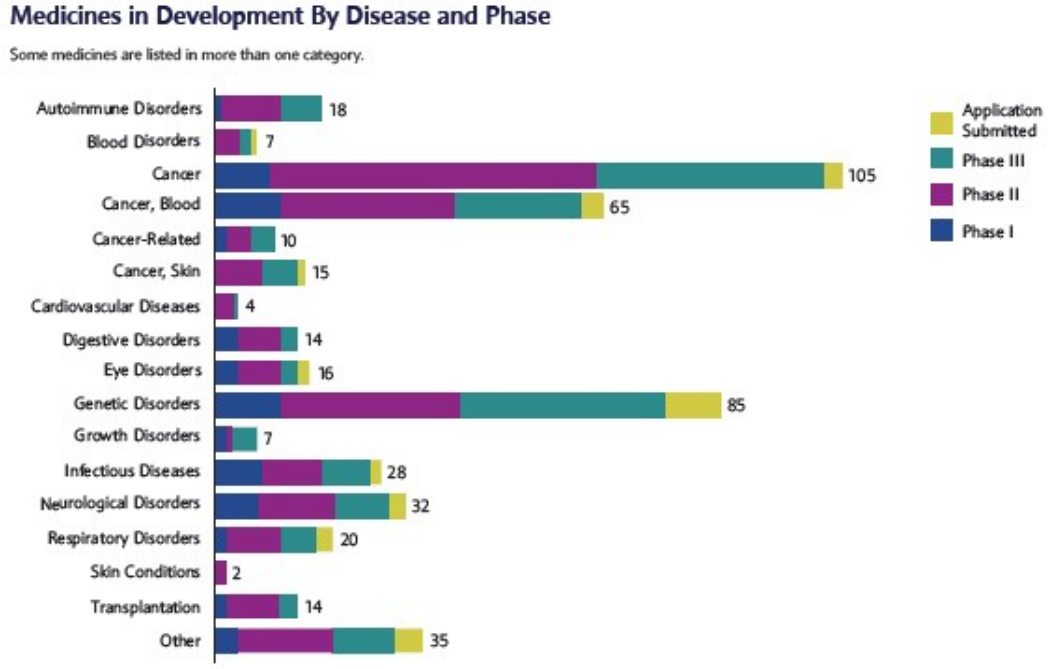
Etimológicamente,
FINISTERRAE
es el fin de la tierra; en nuestro caso, se refiere al principio de la genómica y del **BIG DATA** biomédico con fines traslacionales.

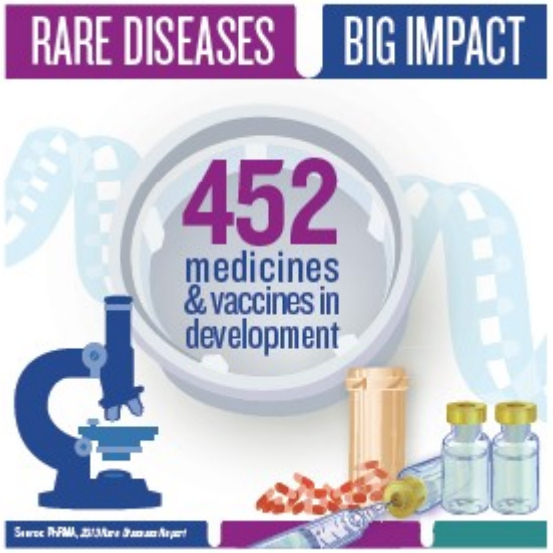
El objetivo es hacer el big data pequeño, reduciendo al máximo la complejidad.


... acelerando la
búsqueda de nuevas
dianas terapéuticas y
de medicamentos
huérfanos en
ENFERMEDADES
RARAS.

...y también en
enfermedades
neurodegenerativas
(Parkinson,
Alzheimer, ELA,...)
y Cáncer...

ITS NOW OR NEVER (E.P., 1960)

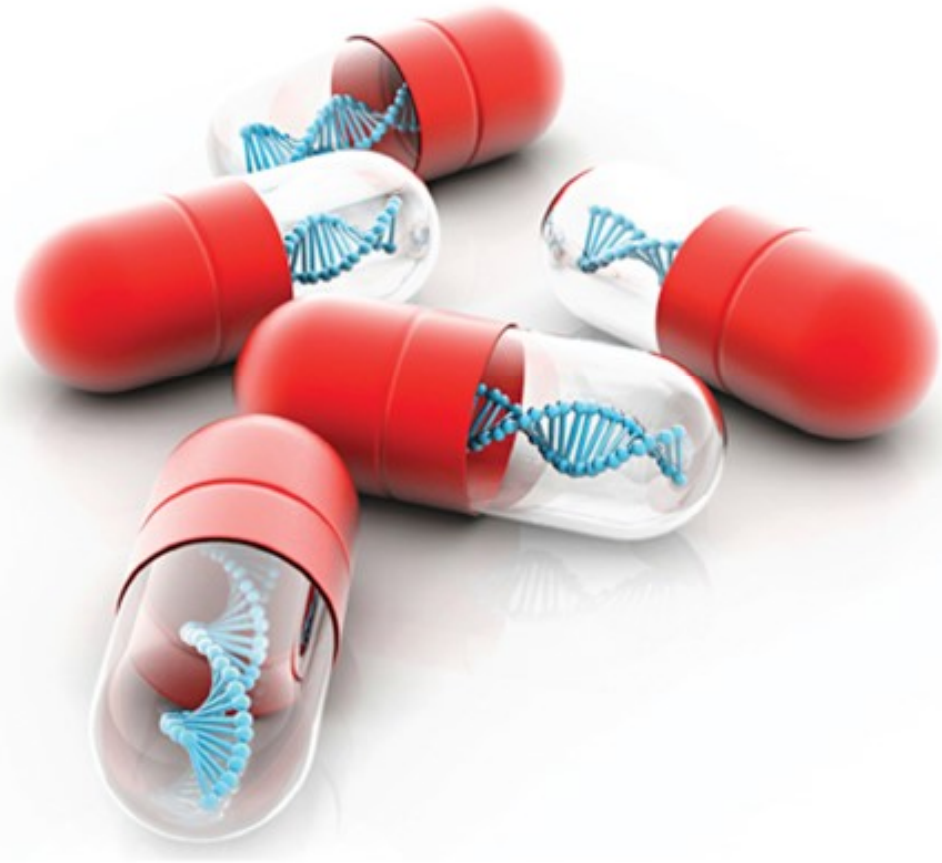




(Fuente: informe sobre enfermedades raras 2013).

FINISTERRAE

Así mismo, se trata de mejorar el diagnóstico precoz de diferentes tipos de cáncer, y posibilitar la OPTIMIZACIÓN DEL TRATAMIENTO PERSONALIZADO (medicina de precisión).



PLOS ONE

RESEARCH ARTICLE

A Novel Peptide for Simultaneously Enhanced Treatment of Head and Neck Cancer and Mitigation of Oral Mucositis

Peili Chen[1]*, Maria Mancini[2], Stephen T. Sonis[2,3], Juan Fernandez-Martinez[2,4], Jing Liu[1¤a], Ezra E. W. Cohen[1¤b], F. Gary Toback[1]*

1 Department of Medicine, University of Chicago, Chicago, Illinois, 60637, United States of America, **2** Biomodels, LLC, Watertown, Massachusetts, 02472, United States of America, **3** Brigham and Women's Hospital, Boston, Massachusetts, 02115, United States of America, **4** Mathematics Department, Universidad de Oviedo, Asturias, Spain

¤a Current address: Department of Pediatrics, University of Texas—Southwestern Medical Center, Dallas, Texas 75390, United States of America
¤b Current address: Moores Cancer Center, University of California, San Diego, California 92093, United States of America
* pchen@medicine.bsd.uchicago.edu (PC); gtoback@medicine.bsd.uchicago.edu (FGT)

CrossMark
click for updates

Además, la plataforma FINISTERRAE contribuirá a la OPTIMIZACIÓN DE DECISIONES MÉDICAS y a la valorización del dato hospitalario

Dentro de la **OPTIMIZACIÓN DE FÁRMACOS,** cobra especial interés la minimización de efectos secundarios (toxicidades) con objeto de evitar sufrimientos innecesarios en los pacientes

Libertas Academica
FREEDOM TO RESEARCH

Open Access: Full open access to this and thousands of other papers at http://www.la-press.com.

Cancer Informatics

Supervised Classification by Filter Methods and Recursive Feature Elimination Predicts Risk of Radiotherapy-Related Fatigue in Patients with Prostate Cancer

Leorey N. Saligan[1], Juan Luis Fernández-Martínez[2], Enrique J. deAndrés-Galiana[2] and Stephen Sonis[3]

[1]National Institute of Nursing Research, National Institutes of Health, Bethesda, Maryland, USA. [2]Universidad de Oviedo, Spain. [3]Biomodels, LLC, Watertown, MA, USA.

Relationship of Mitochondrial Enzymes to Fatigue Intensity in Men With Prostate Cancer Receiving External Beam Radiation Therapy

Biological Research for Nursing
1-7
© The Author(s) 2015
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1099800415617848
brn.sagepub.com
SAGE

Kristin Filler, PhD, RN, OCN[1], Debra Lyon, PhD, RN, FAAN[2], Nancy McCain, PhD, RN, FAAN[3], James Bennett Jr., MD, PhD[4], Juan Luis Fernández-Martínez, PhD[5], Enrique Juan deAndrés-Galiana, PhD[5], R. K. Elswick Jr., PhD[6], Nada Lukkahatai, PhD, RN[7], and Leorey Saligan, PhD, RN, CRNP, FAAN[1]

Libertas Academica
FREEDOM TO RESEARCH

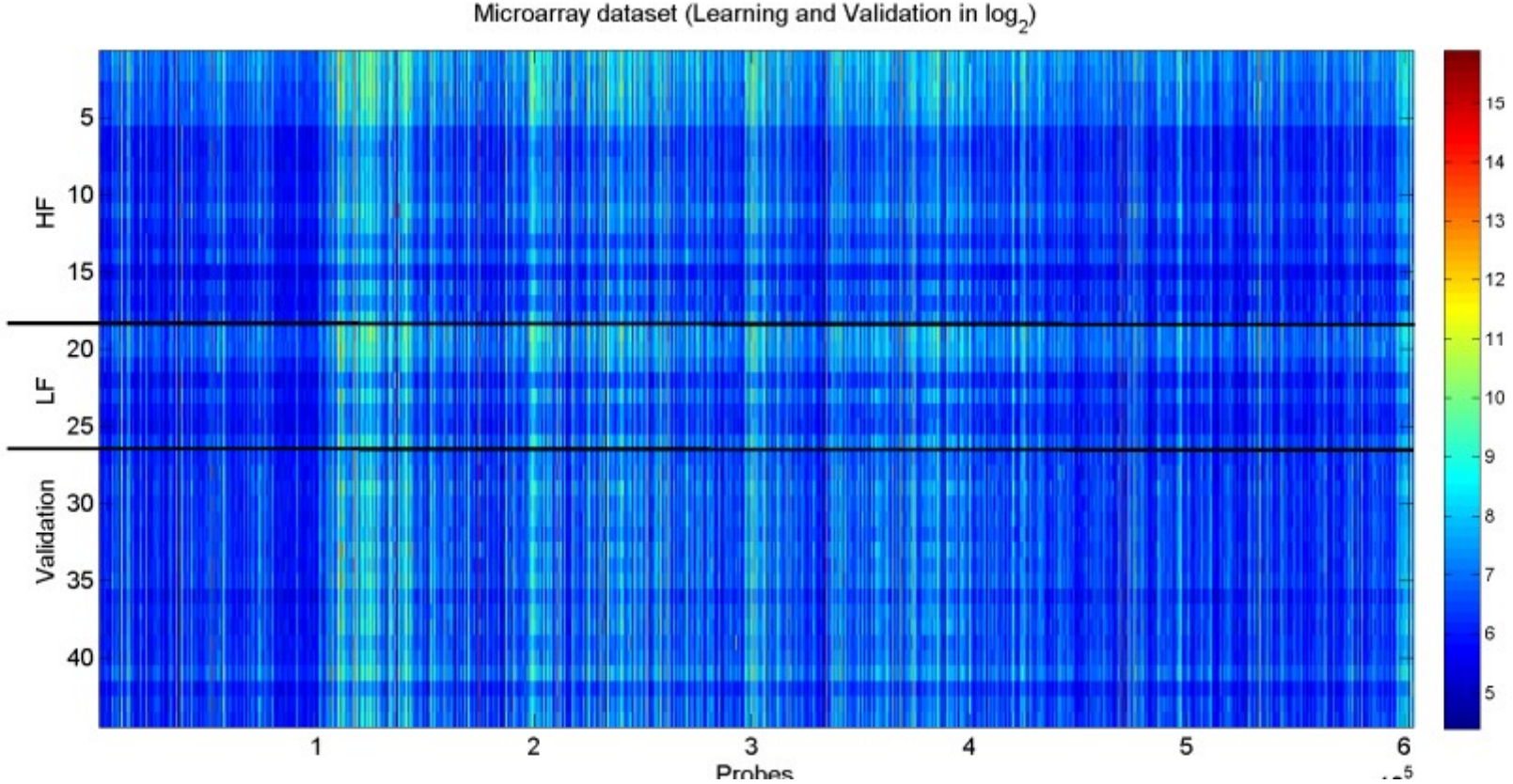
Open Access: Full open access to this and thousands of other papers at http://www.la-press.com.

Cancer Informatics

Supervised Classification by Filter Methods and Recursive Feature Elimination Predicts Risk of Radiotherapy-Related Fatigue in Patients with Prostate Cancer

Leorey N. Saligan[1], Juan Luis Fernández-Martínez[2], Enrique J. deAndrés-Galiana[2] and Stephen Sonis[3]

[1]National Institute of Nursing Research, National Institutes of Health, Bethesda, Maryland, USA. [2]Universidad de Oviedo, Spain. [3]Biomodels, LLC, Watertown, MA, USA.



JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 23, Number 8, 2016
© Mary Ann Liebert, Inc.
Pp. 678–692
DOI: 10.1089/cmb.2016.0008

Design of Biomedical Robots for Phenotype Prediction Problems

ENRIQUE J. DEANDRÉS-GALIANA,[1]
JUAN LUIS FERNÁNDEZ-MARTÍNEZ,[2] and STEPHEN T. SONIS[3]

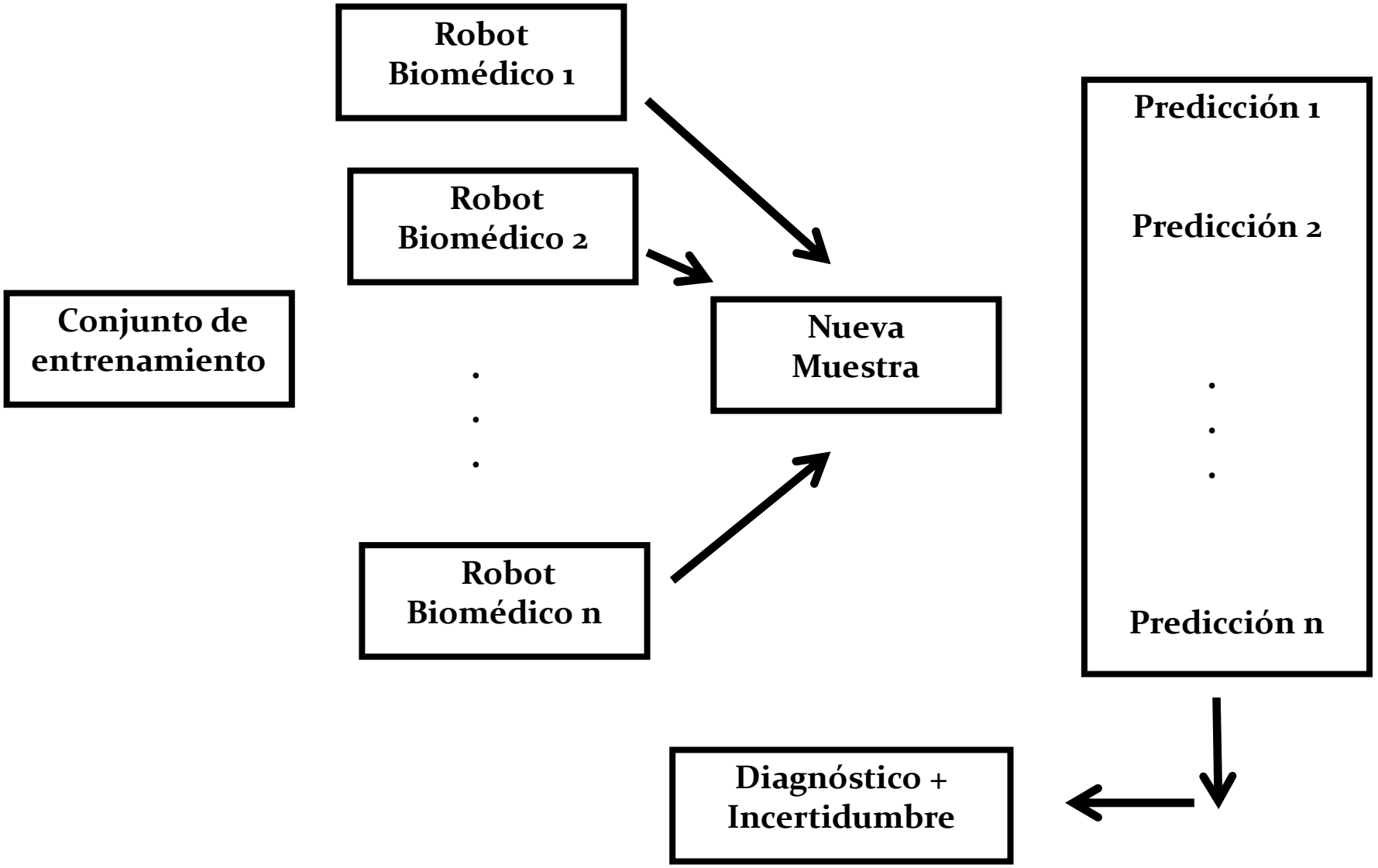
JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 00, Number 0, 2016
© Mary Ann Liebert, Inc.
Pp. 1–12
DOI: 10.1089/cmb.2016.0042

Research Article

Impact of Microarray Preprocessing Techniques in Unraveling Biological Pathways

ENRIQUE J. DEANDRÉS-GALIANA,[1] JUAN LUIS FERNÁNDEZ-MARTÍNEZ,[1]
LEOREY N. SALIGAN,[2] and STEPHEN T. SONIS[3]

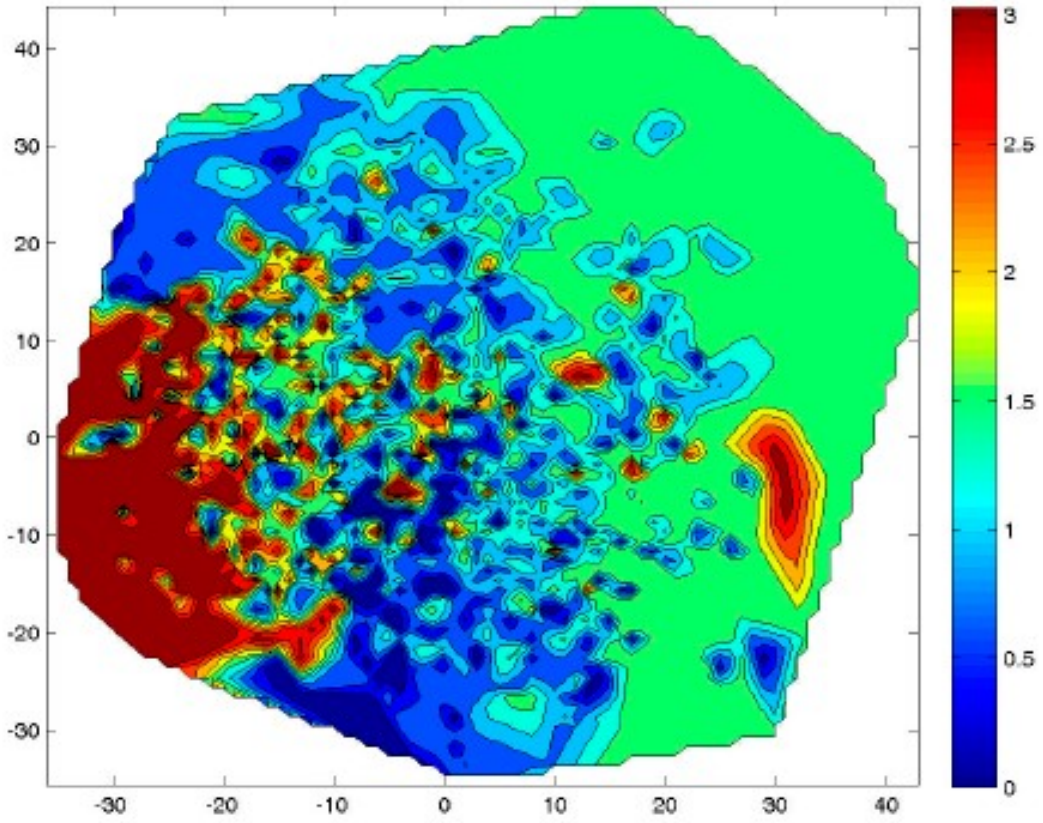
Robots Biomédicos



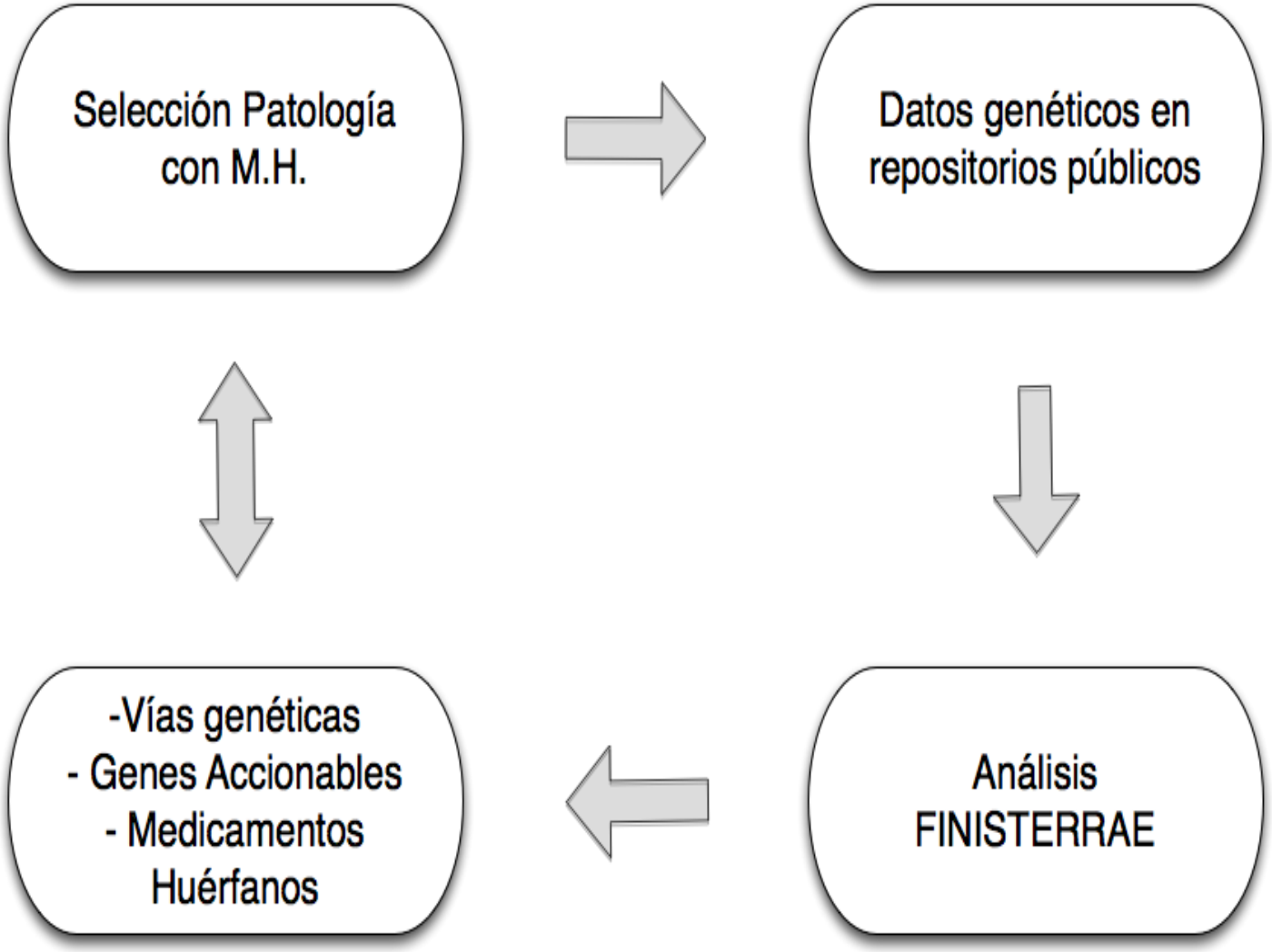
The Uncertainty Space in a biometry prediction problem (face recognition)

$$M_{tol} = \left\{ \mathbf{g} \in \mathbb{R}^{s} : \frac{\left\| \mathbf{L}^{*}(\mathbf{g}) - \mathbf{c}^{obs} \right\|_{p}}{\left\| \mathbf{c}^{obs} \right\|_{p}} \leq tol \right\}$$

Uncertainty Region



FINISTERRAE: APRENDIENDO CONOCIMIENTO



FINISTERRAE

Técnicas de selección de genes de acuerdo a su poder discriminador del fenotipo

Diseño de firmas genéticas a pequeña escala y análisis de estabilidad en la predicción

Técnicas de muestreo

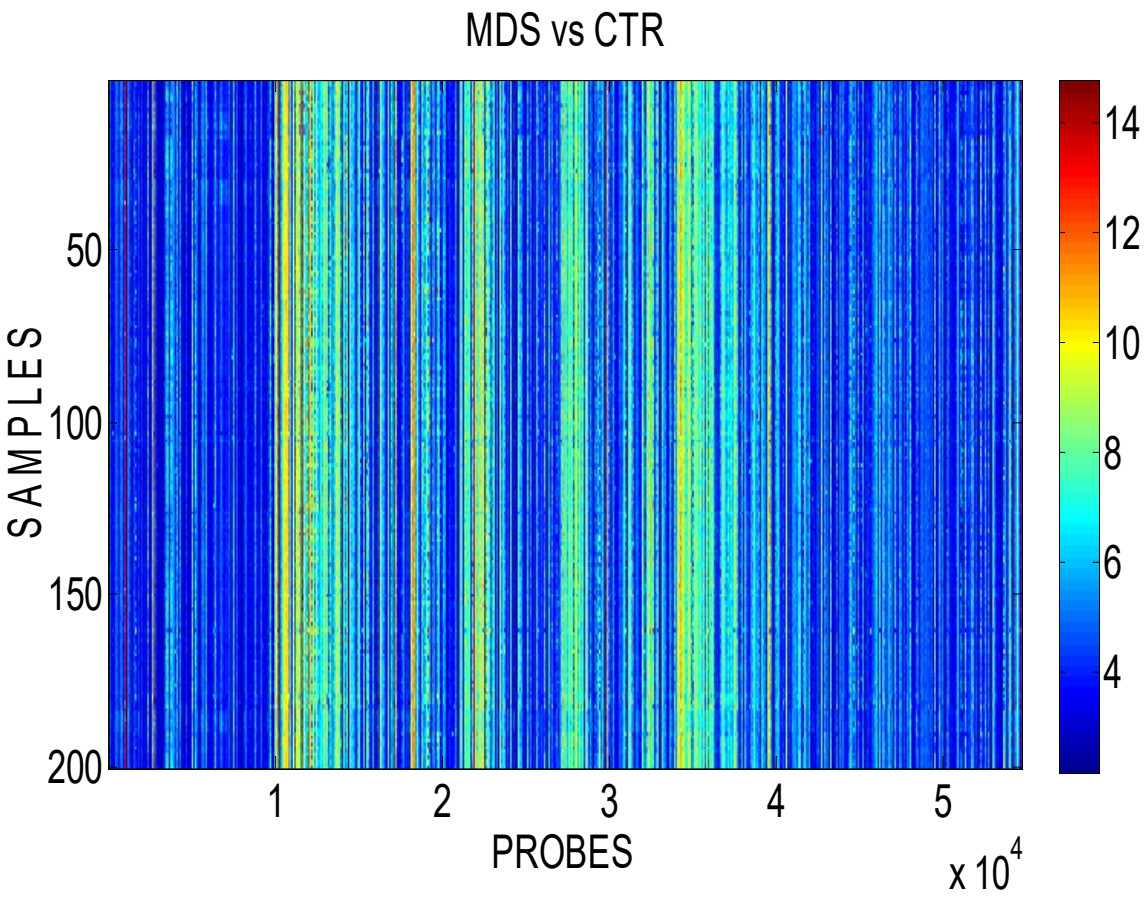
de firmas genéticas equivalentes y redes de correlación

Análisis de Vías Genéticas Defectivas

Genes Accionables

Compuestos y Fármacos

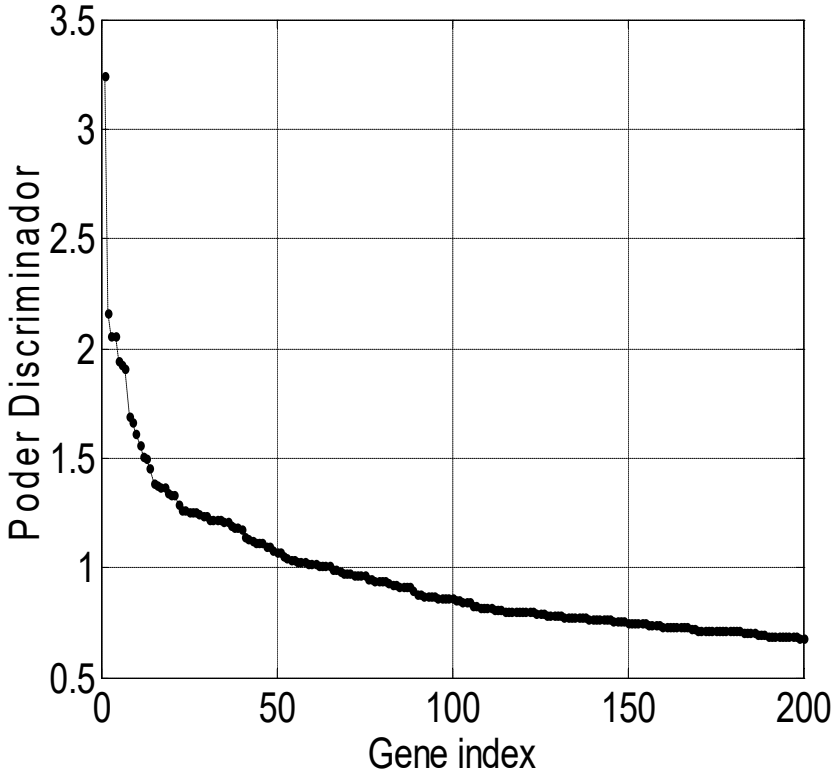
Síndrome Mielodisplásicos

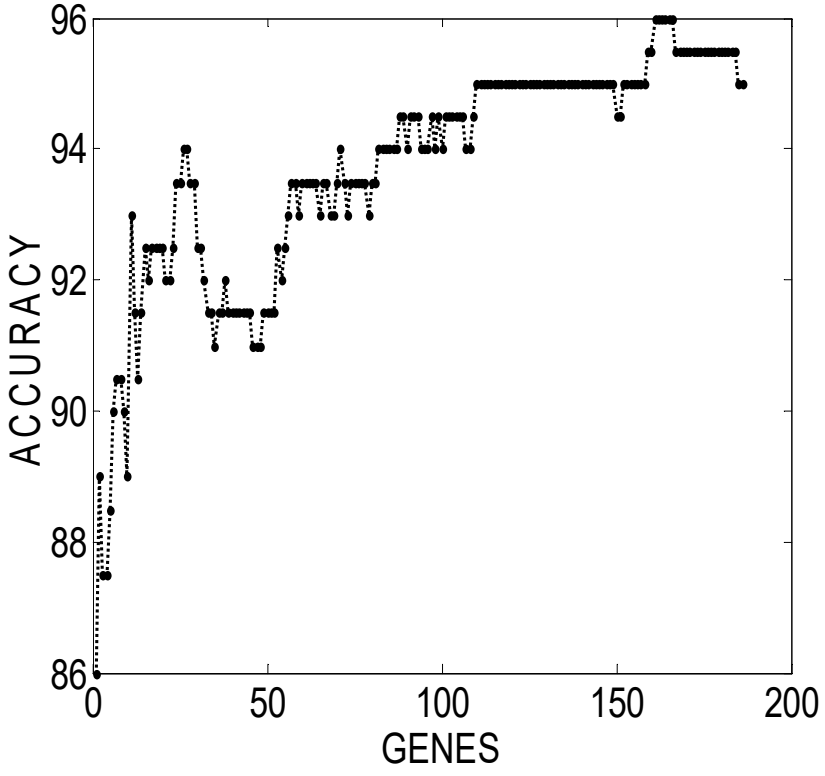


183 patients with MDS patients and 17 healthy controls were included in the study.

Bone marrow samples were obtained and CD34+ cells isolated from MDS patients and healthy controls. Samples were hybridized to Affymetrix GeneChip Human Genome U133 Plus 2.0 arrays

Pellagatti A, Cazzola M, Giagounidis A, Perry J et al. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* 2010 Apr;24(4):756-64.

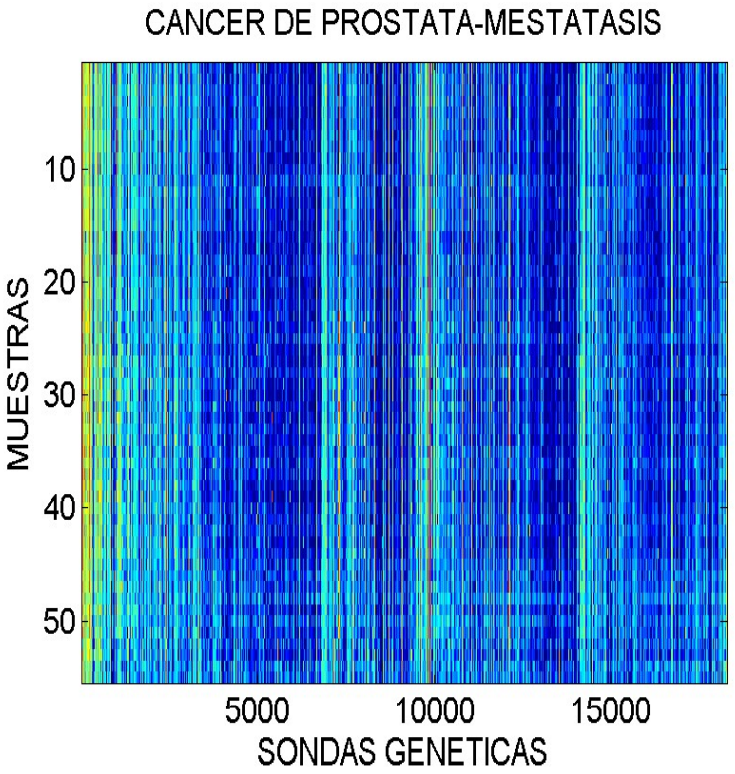

Poder Discriminador
Gene index
3.5
3
2.5
2
1.5
1
0.5
0
50
100
150
200


ACCURACY
GENES
96
94
92
90
88
86
0
50
100
150
200

Análisis Finisterrae

- Predicción del fenotipo MDS vs Control con 96% de exactitud con una firma que contiene 161 genes.

- Las vías genéticas implicadas más importantes detectadas por nuestro análisis son: B Cell Development Pathways, NF-kappaB Signaling, Hematopoietic Cell Lineage, Primary Immunodeficiency, and DNA Topological Changes.



Metástasis en cáncer de próstata

VIAS GENÉTICAS DEFECTIVAS

PAK Pathway
PEDF Induced Signaling
Stem Cell Differentiation Pathways
P70S6 Kinase Signaling

MEDICAMENTOS HUÉRFANOS

- **NADPH : The NADPH system is responsible for generating free radicals in immune cells to destroy pathogens.**
- **TAMOXIFENO: Selective estrogen-receptor modulator (SERM) used for breast cancer prevention and treatment. It is on the World Health Organization's List of Essential Medicines. Tamoxifen is available as a generic medication.**

¿Por qué sabemos que la metodología FINISTERRAE funciona?

Porque se han realizado pruebas de concepto sobre enfermedades raras para los cuales se han hallado recientemente medicamentos, habiendo encontrado sus principios activos entre las vías genéticas implicadas detectadas por FINISTERRAE.

Ejemplo: Esclerosis Sistémica:

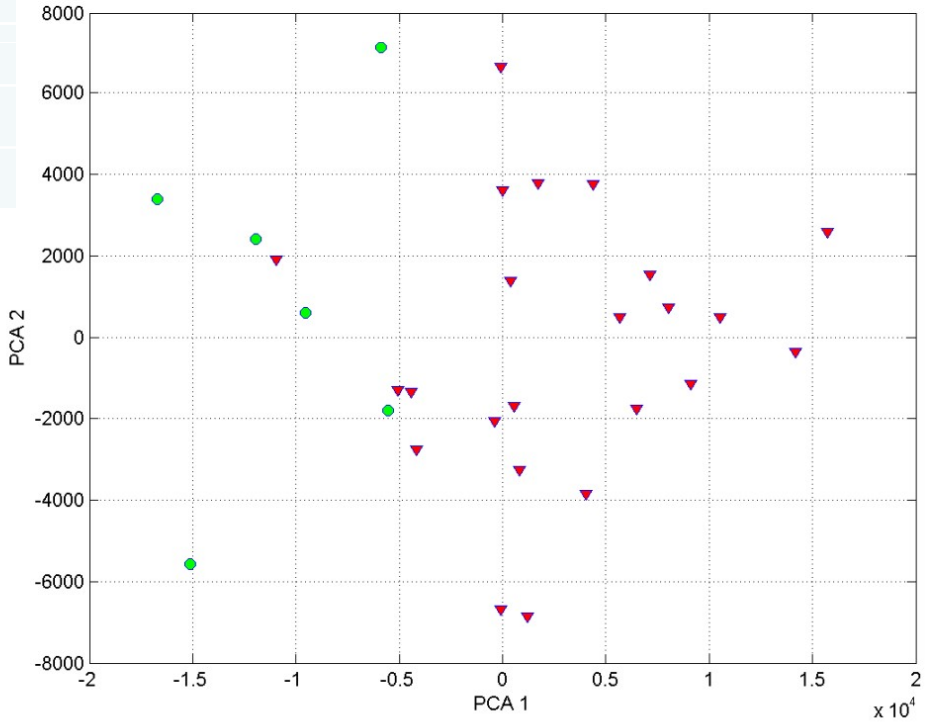
- Genentech ha desarrollado Tocilizumab, un anticuerpo monoclonal que se une al receptor celular de la interleucina-6.
- Celgene ha desarrollado recientemente Pomalyst que es una droga anti-angiogénica que actúa como inmunomodulador provocando la sobreregulación de IL-2/IL-10 y la subregulación de IL-6.

Porque utilizamos técnicas de validación cruzada para analizar la estabilidad de la firmas genéticas halladas.

MIOSITIS-POLIMIOSITIS

Allograft Rejection
Influenza A
Class I MHC Mediated Antigen Presentation
Staphylococcus Aureus Infection
Interferon Signaling
Immune Response IFN Alpha/beta Signaling Pathway
Phagosome
Tuberculosis
Cell Adhesion Molecules (CAMs)
Epstein-Barr Virus Infection
TNF Signaling

Hemos identificado las posibles infección por virus y/o bacterias responsables de su desarrollo, identificado las vías genéticas involucradas, y un conjunto de genes defectivos que se expresan sólo en el músculo.



Esclerosis Múltiple, ELA, Parkinson y Alzheimer

- Estudio de las bases moleculares en el desarrollo de diferentes enfermedades neurodegenerativas, análisis de vías genéticas involucradas, firmas genéticas a pequeña escala para diagnóstico precoz.

- Hemos mostrado la importancia de las **caspasas** en el **ELA**, que controlan el mecanismo de apoptosis. Se ha hallado una relación de las vías genéticas afectadas con el virus de **Epstein-Barr** (herpesvirus).

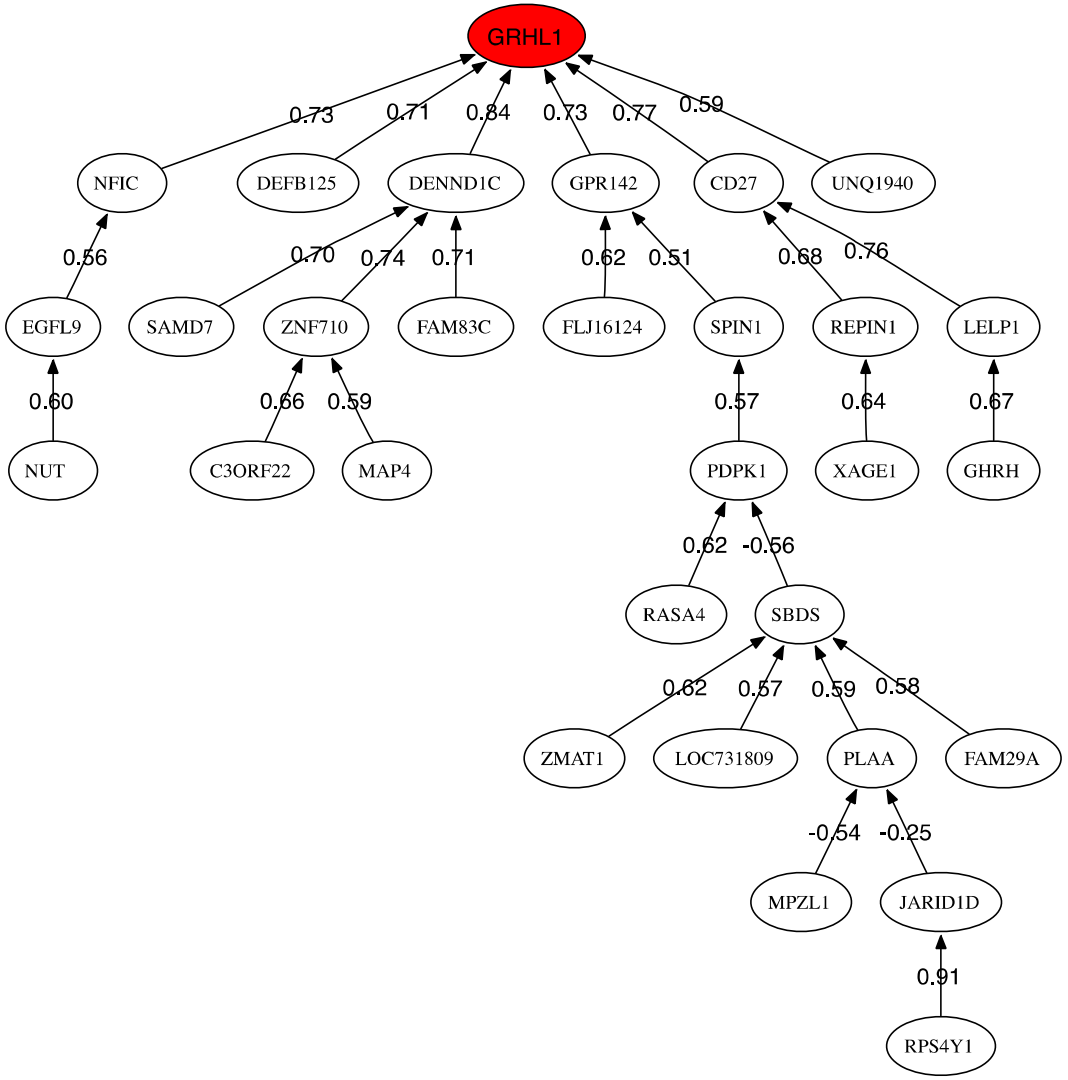
- En el caso de la **esclerosis múltiple** las vías afectadas están relacionadas con **Artritis reumatoide, CD28 y Células Presentadoras de antígeno**.

- En el caso del **Parkinson** hemos visto la importancia de "pathways" de hipoxia (**HIF-1 signaling) , transcripción (FOXA1 transcription factor) y señalización (P-53 signaling**).

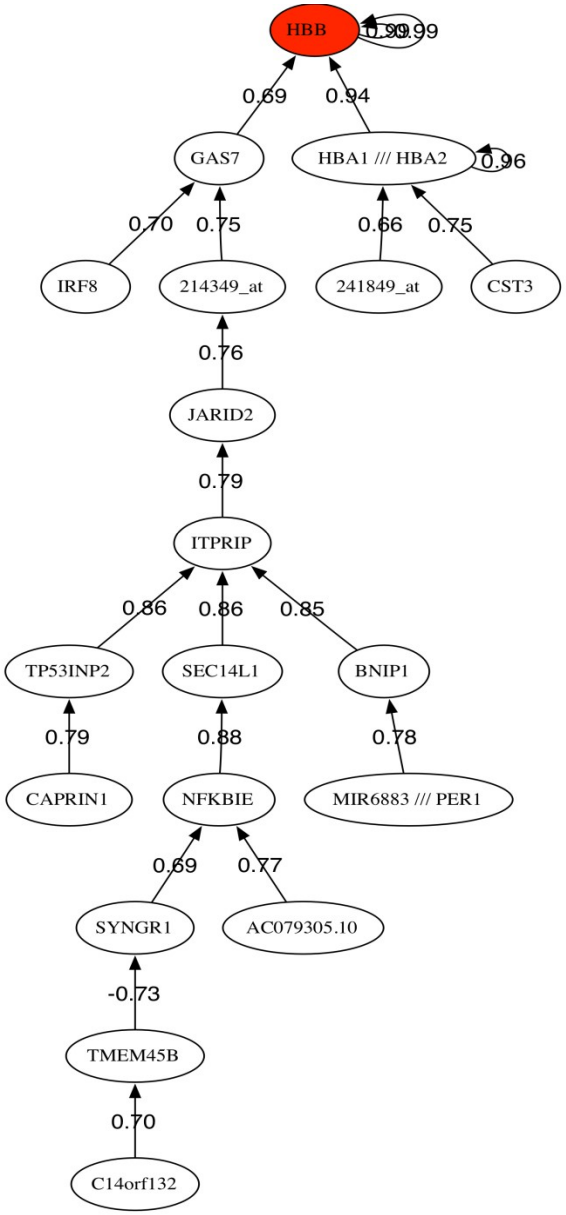
- En el caso del **Alzheimer y DCL** los resultados son muy sorprendentes e implican **procesos de traslación del MRNA viral**.

En todos los casos se han encontrado medicamentos huérfanos que desearíamos que se probasen clínicamente.

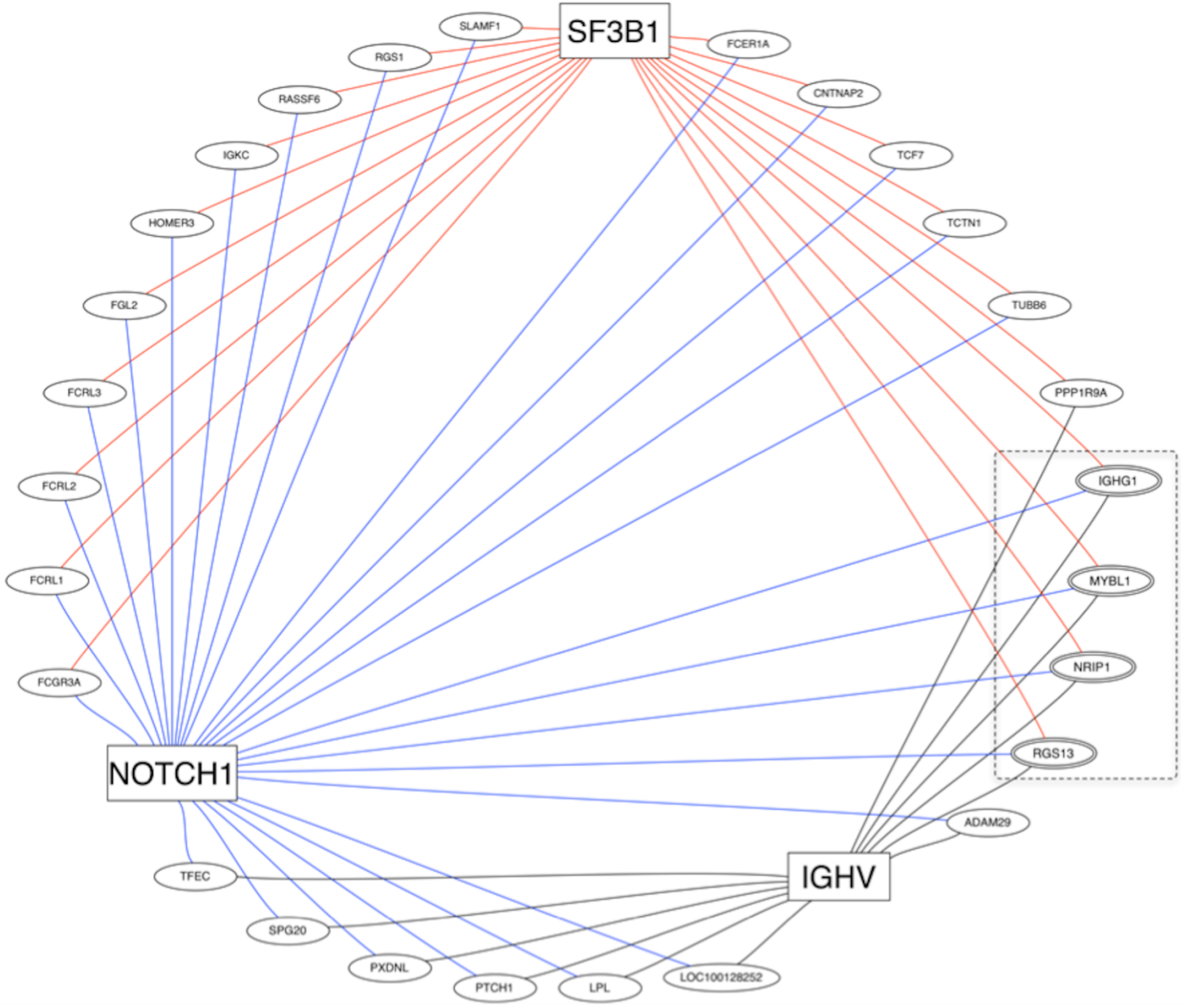
Parkinson



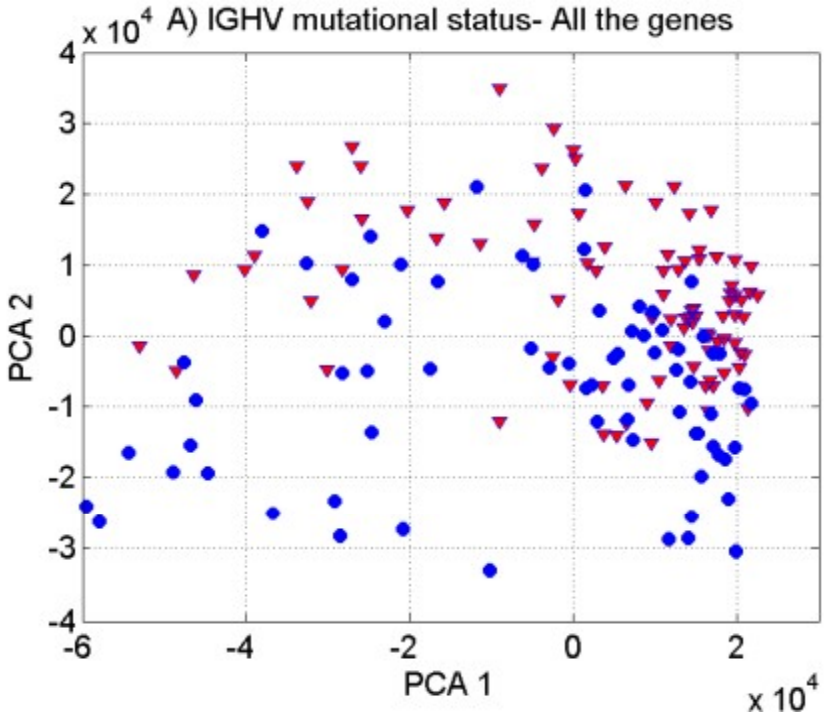
Esclerosis

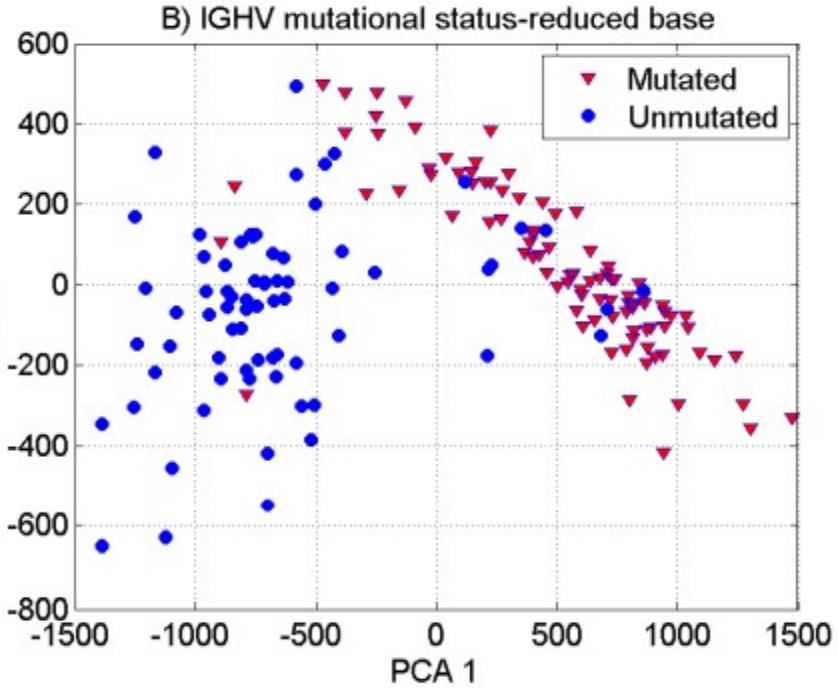


¿Cómo afectan las mutaciones en los pacientes con leucemia?



PARSIMONIA





- **Cada problema de aprendizaje tiene su propio alfabeto**
- **Cuando se encuentran los alfabetos de un problema, la clasificación se convierte en linealmente separable y el clasificador es capaz de generalizar con alta precisión.**

A MODO DE CONCLUSION

- **En mi opinión, no existe diferencia entre la investigación básica y la aplicada, sólo para aquellos que quieren justificar su ineficiencia.**
- **Como educadores, tenemos la responsabilidad de transmitir a nuestros alumnos el gusto por la disciplina y el conocimiento.**
- **Las matemáticas BIEN EXPLICADAS tienen que ser la base del lenguaje de todos los científicos e incluso humanistas.**
- **Las matemáticas son DEMASIADO IMPORTANTES para el futuro tecnológico de nuestro PAIS.**
- **Tenemos que educar KNOWMADS. Para ello el conocimiento de idiomas y de las técnicas del estado del arte de la tecnología es fundamental.**
- **A veces MENOS es MAS. Centrémonos en enseñar a nuestros alumnos a pensar, la autoría y la resolución efectiva de problemas.**
- **Si se persigue el éxito debemos reforzar positivamente el fracaso, especialmente cuando detrás hay esfuerzo.**

"Parsimony is the key in genetic data study"

Simplifying what is complex

Genetics stepping ahead as Spain-based researchers develop biomedical robots that can help in the study of rare diseases, neurodegenerative disorders and cancer, as well as the creation of orphan drugs

Report: Mélisande Rouger

Predicting disease development and anticipating adequate treatment may appear to be as uncertain as gazing into a crystal ball. But, what if there was a way to sort out and classify the myriad available genetic information to help doctors to do so?

A multidisciplinary team from Oviedo University is beginning to demonstrate that possibility by using biomedical robots that can adapt and improve predictions as they unravel phenotype – how genetic information expresses itself in a given environment.

'Searching for classifications that can optimally explain gene expression is a little bit like trying to open a safe deposit box with multiple combinations. We are working with 50,000 genetic probes in cohorts of 100 or 200 patients at a time; this means we are dealing with an extremely high level of indetermination.

'We need tools that take into account the degree of uncertainty contained in phenotype prediction and understand that genetic networks have their own structure,' explained Professor Juan Luis Fernández-Martínez, Head of Inverted Problems at the Mathematics Department in Oviedo University.

Combining knowledge from applied mathematics, statistics and computational science, since 2010 Fernández-Martínez and team have been designing biomedical robots with Biomodels, a company headed by Dr Steve Sonis from Harvard University and Brigham and Women Hospital in Boston.

Their robots are able to learn from complex data sets and help doctors make decisions and plan personalised treatment according to a patient's needs; in a word, they enable personalised medicine.

The principle behind designing such machines is simpler than one might have thought. 'People believed they needed highly complex computers to treat so much data, but we actually did it with a pretty simple model. Basically, we applied the concept of parsimony. Uncertainty comes from ambiguity in knowledge, so it's interesting to look at things through that angle and simplify what is complex,' Fernández added.

The engineer and his colleagues believe the machines will not only help in the study of rare and neurodegenerative disease and cancer, but also in designing orphan drugs and new treatments, and tailoring therapy to each patient. The team has just published a study in the *Journal of Computational Biology*, in which they show robotic potential in chronic lymphocytic leukaemia, inclusion body myositis and amyotrophic lateral sclerosis.

In leukaemia, the researchers have highlighted the relevance of Lipoprotein lipase, an enzyme that decomposes triglycerides in free fatty acids and glycerol. The team has also discovered the importance of certain genes and how viruses and bacteria influence the genesis of myositis, an autoimmune disease affecting the muscular system, by showing how the immune system mistakes the organism's proteins with those of the invader.

Biomedical robots have also been able to underline the importance of caspases, a group of proteins controlling programmed cell death and which have been related to mutations in gene SOD1, a cause for amyotrophic lateral sclerosis.

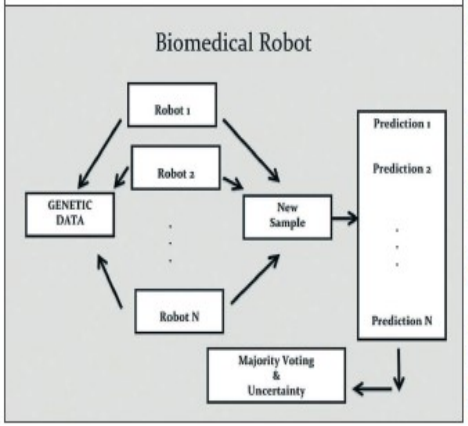
All these findings still need clinical validation, Fernández-Martínez stressed.

Meanwhile, the Oviedo team has developed tools that can help differentiate between genes called headers, which are responsible for most of the phenotype, and helpers, adjuvant genes that bring further detailed information on discrimination. Researchers have applied this method to a toxicity study and will publish results soon on Parkinson and Alzheimer's disease from the so-called Finisterrae project (http://www.slideshare.net/ifalberti/effective-analysis-of-biomedical-big-data-for-the-optimisation-of-drugs-for-rare-and-neurodegenerative-diseases-and-cancer).

The Oviedo researchers also recently published a paper questioning most of what has been said so far on microarray pre-processing techniques by showing their impact in biological pathways (Impact of Microarray Pre-processing Techniques in Unravelling Biological Pathways, Journal of Computational Biology, http://online.liebertpub.com/doi/abs/10.1089/cmb.2016.0042).

'Genetics is still an understudied field and gene function is often misunderstood. Our long-term goal is to make a genetic test as common as a blood test,' Fernández-Martínez said. 'We will multiply cooperation projects with the medical field to do so.'

Professor Fernández-Martínez (left) with Dr Enrique J Andrés Galiana outside the Geology Department at Oviedo University.



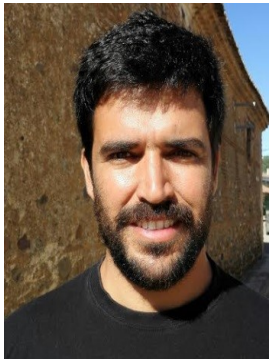
Trained as a petroleum engineer in Paris (1988) and London (1989), and after years as an IT software engineer in France, in 1994 **Juan Luis Fernández-Martínez** gained a PhD in mining engineering from the University of Oviedo, in Spain. He joined the university's mathematics department and became Professor of applied mathematics. During 2008-2010 he was a visiting and research professor at UC Berkeley-Lawrence Berkeley Laboratories and Stanford University. His expertise includes cooperative global optimisation methods, with applications in oil and gas, biometry, finance and biomedicine – in which he aims to design biomedical robots in translational medicine for diagnosis, prognosis, and treatment optimisation. The Finisterrae project particularly aims to find effective solutions for rare and neurodegenerative diseases, as well as various types of cancer.

MUCHAS GRACIAS POR SU ATENCIÓN

ESTOY A SU DISPOSICIÓN PARA RESPONDER A SUS PREGUNTAS

FINISTERRAE

EL EQUIPO



Enrique J. De Andrés Galiana
Doctor en Matemática Aplicada.

Juan Luis Fdez.-Martínez
Director del Grupo de Problemas Inversos e IP.

Zulima Fernández-Muñiz
Doctor en Matemática Aplicada.

Ana Cernea-Cobernau
Doctor en Matemática Aplicada

COLABORADORES MEDICOS

- Dr. Stephen T. Sonis (Harvard University, Brigham and Women Hospital, Dana Farber Cancer Institute).
- Dr. Segundo González (Universidad de Oviedo, IUOPA).
- Dr. Manuel Menéndez (HUCA, Universidad de Oviedo).

Ignacio Fdez.-Alberti
RI & Comunicación