

iENER'18

I Congreso Ingeniería Energética



La Analítica de Datos como clave en el Proceso de Transformación Digital

Sergio Gil López

Índice

- 1 Introducción: La Transformación Digital
- 2 TECNALIA: Digital Energy
- 3 Caso I: Sistemas de Control Predictivo (MPC)
- 4 Caso II: Sensores Virtuales
- 5 Conclusiones



ALGUNAS RESEÑAS HISTÓRICAS SOBRE ANALÍTICA DE DATOS...

J. W. Tukey, The future of Data Analysis, 1962.

THE FUTURE OF DATA ANALYSIS¹

By JOHN W. TUKEY

Princeton University and Bell Telephone Laboratories

I. General Considerations	2
1. Introduction	2
2. Special growth areas	3
3. How can new data analysis be initiated?	4
4. Sciences, mathematics, and the arts	5
5. Dangers of optimization	7
6. Why optimization?	8
7. The absence of judgment	9
8. The reflection of judgment upon theory	10
9. Teaching data analysis	11
10. Practicing data analysis	13
11. Facing uncertainty	13
II. Spotty Data	14
12. What is it?	14
13. An appropriate step forward	15
14. Trimming and Winsorizing samples	17
15. How soon should such techniques be put into service?	19
III. Spotty Data in More Complex Situations	21
16. Modified normal plotting	21
17. Automated examination	22
18. FUNOP	22
19. FUNOR-FUNOM in a two-way table	24
20. Example of use of FUNOR-FUNOM	27
IV. Multiple-Response Data	32
21. Where are we, and why?	32
22. The case of two samples	33
23. Factor analysis: the two parts	35
24. Factor analysis: regression	36
25. Factor analysis: the middle lines	36
26. Taxonomy; classification; incomplete data	38
V. Some Other Promising Areas	39
27. Stochastic-process data	39
28. Selection and screening problems	40
29. External, internal, and confounded estimates of error	41
30. The consequences of half-normal plotting	42
31. Heterogeneous data	42
32. Two samples with unequal variability	43
VI. Flexibility of Attack	44
33. Choice of modes of expression	44
34. Sizes, nomination, budgeting	45
35. A caveat about indications	46

Received July 1, 1961.

¹ Prepared in part in connection with research sponsored by the Army Research Office through Contract DA36-034-ORD-2297 with Princeton University. Reproduction in whole or part is permitted for any purpose of the United States Government.

Bases de Datos relacionales 1980.

Gregory Piatetsky-Shapiro organiza la 1ª Ed. Del Workshop KDD (Knowledge Discovery in Database), 1989.

Data Warehouse, 1990.

1996, aparece el perfil profesional del Data Scientist.

1997, aparece Google

1997, aparece el primer esquema de Deep Learning, LSTM.

2004, Google escribe el primer whitepaper sobre BIG DATA.

2005, se introduce el concepto Web2.0

2014, el acceso al dato es más grande desde dispositivos móviles que desde PCs

2015, se contrata el primer Data Scientist en la Casa Blanca.



Introducción

Factores claves de la Digitalización

En 2020 se estiman 50,1 billones de dispositivos conectados

En el sector industrial se genera 1 exabyte de datos anuales...

...se incrementará en un factor 20 en los próximos 10 años.

Cálculo distribuido.

Computación en la nube.

Internet of Things (IoT)

Analítica de Datos

Capacidad de almacenamiento

Capacidad de procesamiento

Capacidad de procesar Peta-bytes de datos.

En los últimos 10 años el volumen de datos almacenados se ha incrementado un 400%

105 millones de usuarios almacenan datos en la nube

El precio del almacenamiento ha disminuido más de un 50% en los últimos 10 años.



Introducción

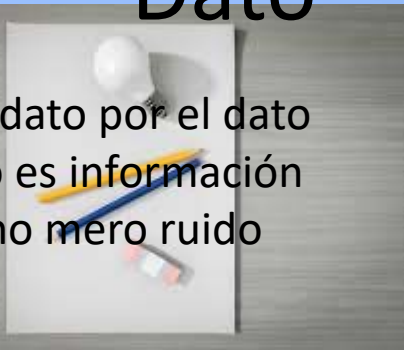
La relevancia de la Analítica de Datos



Aumento de la complejidad de los negocios

Dato

El dato por el dato no es información sino mero ruido



La analítica de datos
transforma el
datos en
información

Mayor
productos

Y de información en conocimiento que haga una realidad la transformación digital por diversidad de materias primas



Mayor diversidad de procesos productivos



Mayor complejidad de la operación y control



**TECNALIA TAMBIÉN SE PREPARA
PARA LA DIGITALIZACIÓN...**

Centro Tecnológico TECNALIA Research and Innovation

LAS PERSONAS DE TECNALIA

1.405 PERSONAS EN PLANTILLA

57% HOMBRES **43%** MUJERES

30 NACIONALIDADES DIFERENTES

42,4 AÑOS DE EDAD MEDIA

229 NÚMERO DE DOCTORES



TRABAJANDO POR UN OBJETIVO COMÚN: GENERAR OPORTUNIDADES DE NEGOCIO A TRAVÉS DE LA INVESTIGACIÓN APLICADA.

4 EJES DE UNA OFERTA AL SERVICIO DE LA EMPRESA



INGRESOS

102,1 MILLONES DE EUROS

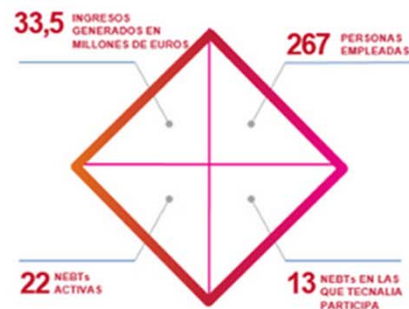
PATENTES

28+12 SOLICITADAS EN 2016 **116** FAMILIAS EN CARTERA **391** ACTIVAS CONTENIDAS EN ESAS FAMILIAS

CARTERA DE CLIENTES

4.050 EMPRESAS

NEBTs



Construcción Sostenible	Energía y Medio Ambiente	ICT
Industria y Transporte	Salud	Lab_services

Digital Energy en TECNALIA Research and Innovation

tecnalia Inspiring Business
 Digitalización al servicio de un sistema energético más seguro y eficiente, más inteligente y flexible, y preparado a los nuevos modelos de negocio con un rol más activo de los consumidores
 Desarrollamos la tecnología capaz de transformar el presente
www.tecnalia.com
 @tecnalia



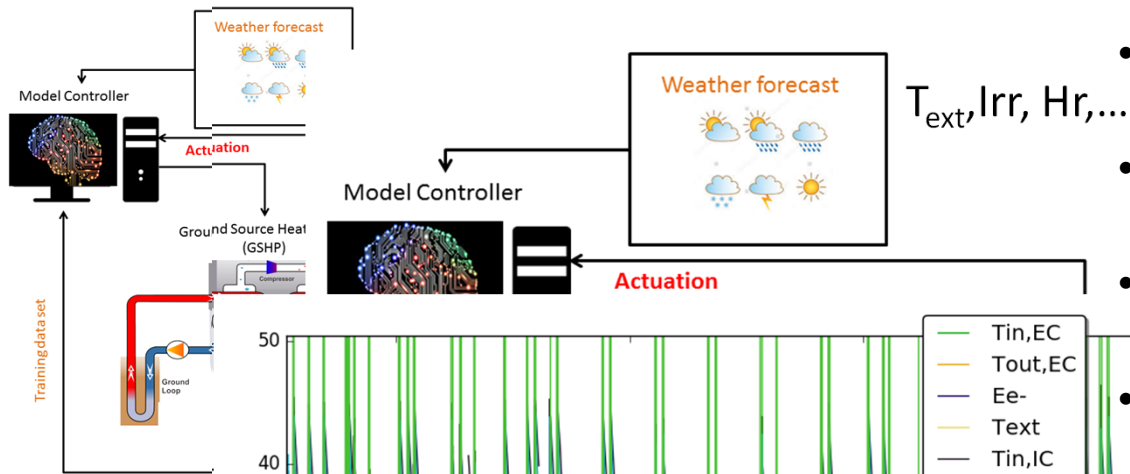


SISTEMA INTELIGENTE DE CONTROL DE LA CLIMATIZACIÓN...

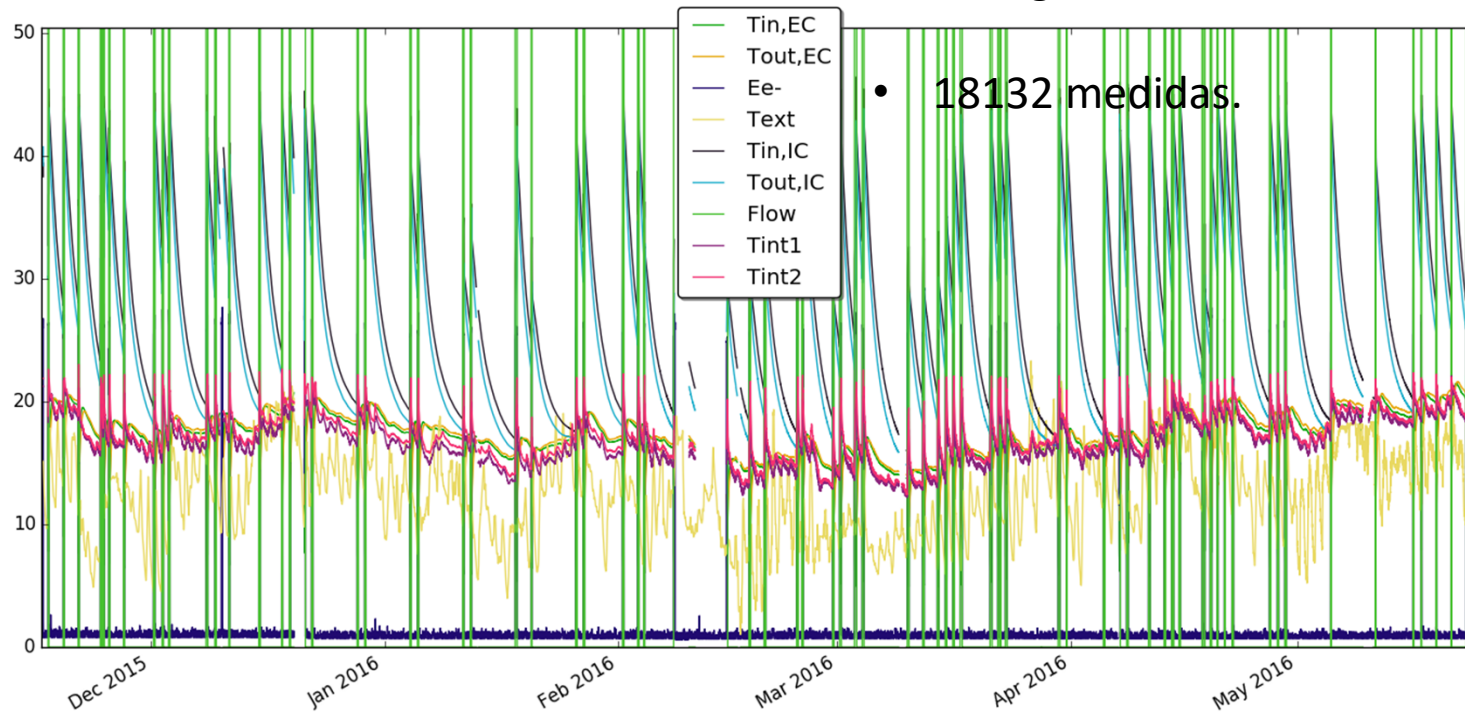
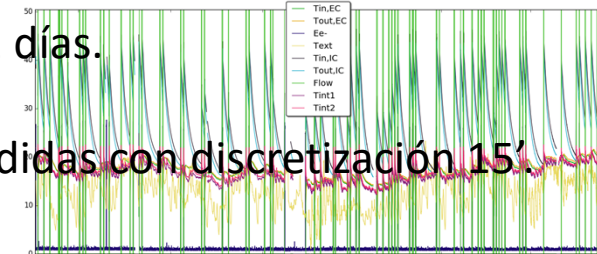
Sistemas de Control Predictivo (MPC)

- Caso de Uso: Optimización de los arranques y paros de un Sistema de Geotermia (GSHP):
 - Objetivo: Modelización energética de un sistema de climatización y su relación con las variables de entorno, con el fin de minimizar su consumo sin afectar al confort (condiciones interiores).
 - Problema Multi-paramétrico no-lineal.
 - Modelos Físicos.
 - Modelos Híbridos. Ej. Aproximación RC.
 - Modelos “caja negra” o “data driven”. Analítica de Datos.
 - Ventajas: presencia de decalajes temporales entre estímulo y respuesta (inercias térmicas); efectos procedentes de instantes de tiempo previos (no linealidad temporal) → controlar desde la anticipación (Model Predictive Control, MPC).
 - Clave de aplicación: Automatización de la solución.
 - Proyectos: GEOTECH, FHP, S. Gil-Lopez, E. Zabala , E. Perea.

Sistemas de Control Predictivo (MPC)



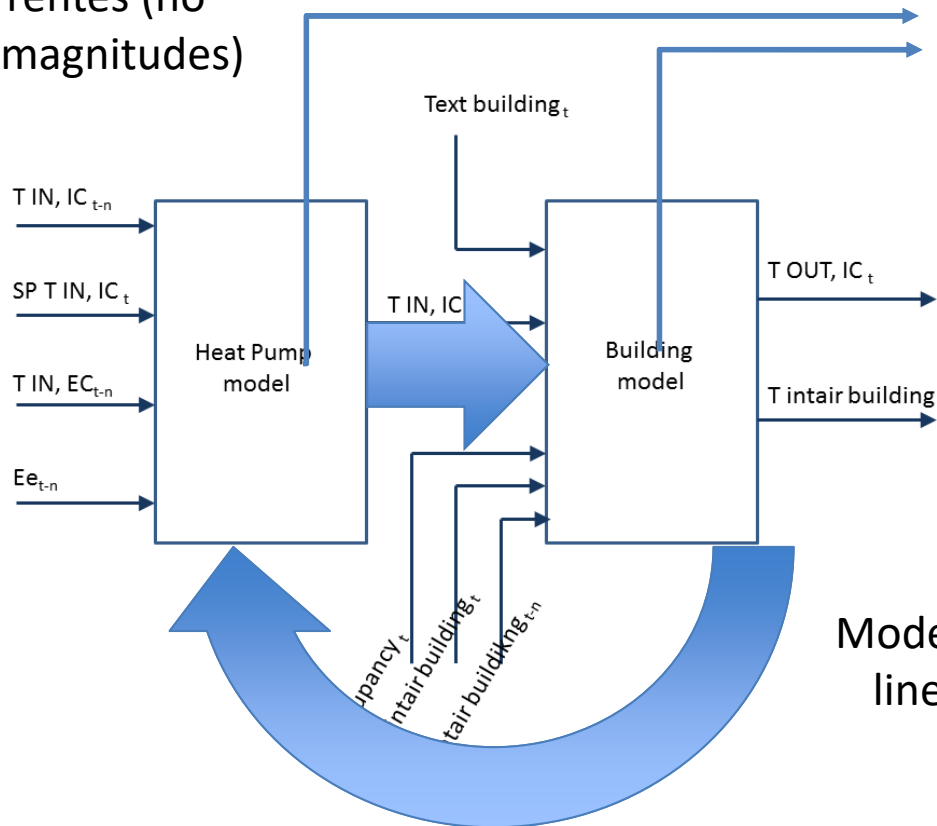
- 192 días.
- Medidas con discretización 15'.
- 9 magnitudes físicas relacionadas.



- 18132 medidas.

Sistemas de Control Predictivo (MPC)

Modelos recurrentes (no linealidad entre magnitudes)

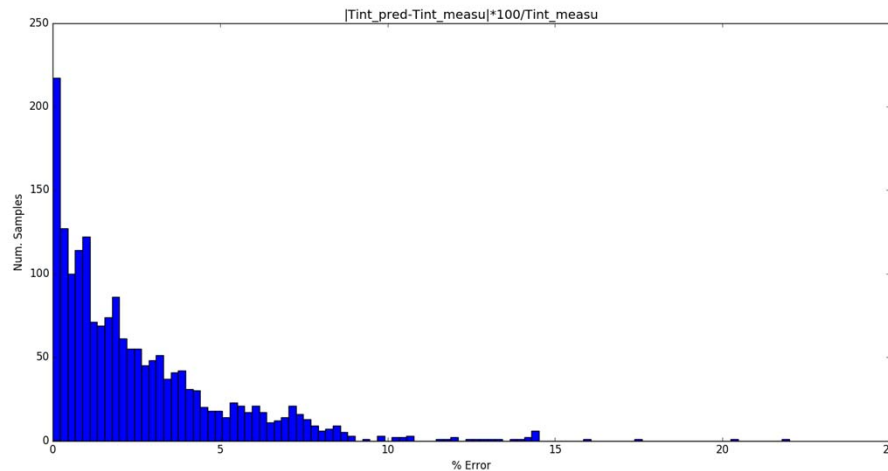


Sistemas de aprendizaje máquina supervisado basados en ensamblaje: Random Forest

Esquema de Cross-validation LOOCV ("Leave One Out Cross Validation")

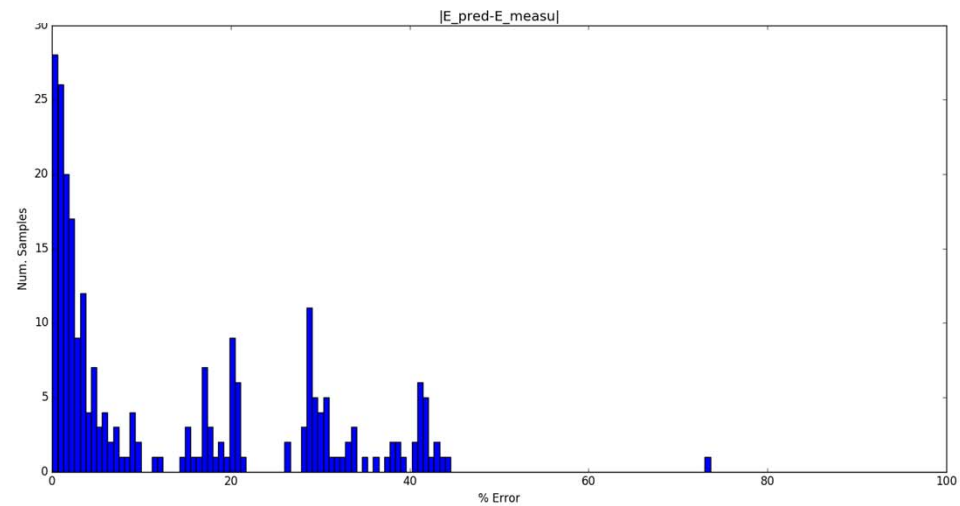
Modelos recursivos (no-linealidad temporal)

Sistemas de Control Predictivo (MPC)

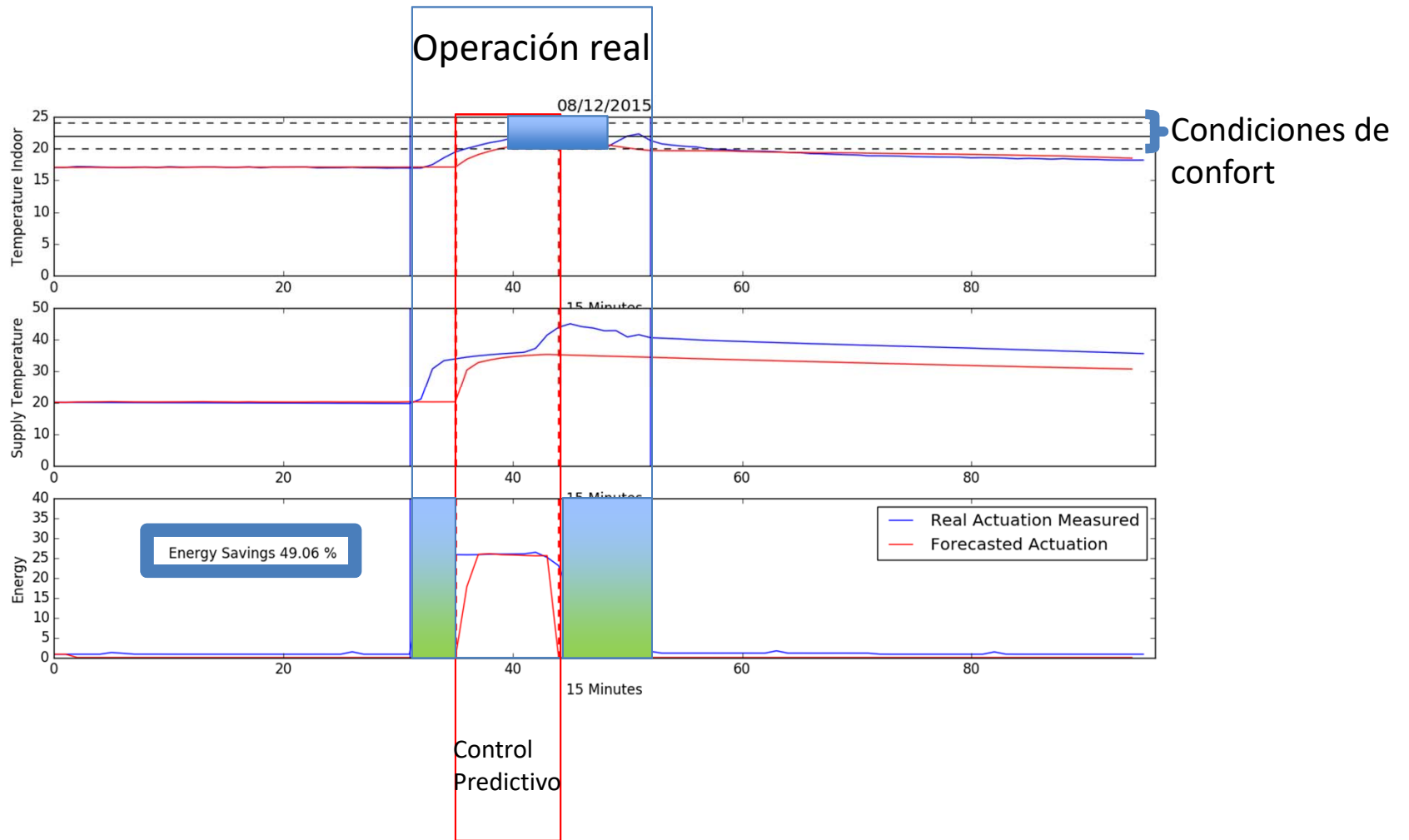


Temperatura Interior:
MAPE = 0.5°C (2,5%)

Consumo Eléctrico
MAPE = 3,3 J (20,3%)



Sistemas de Control Predictivo (MPC)



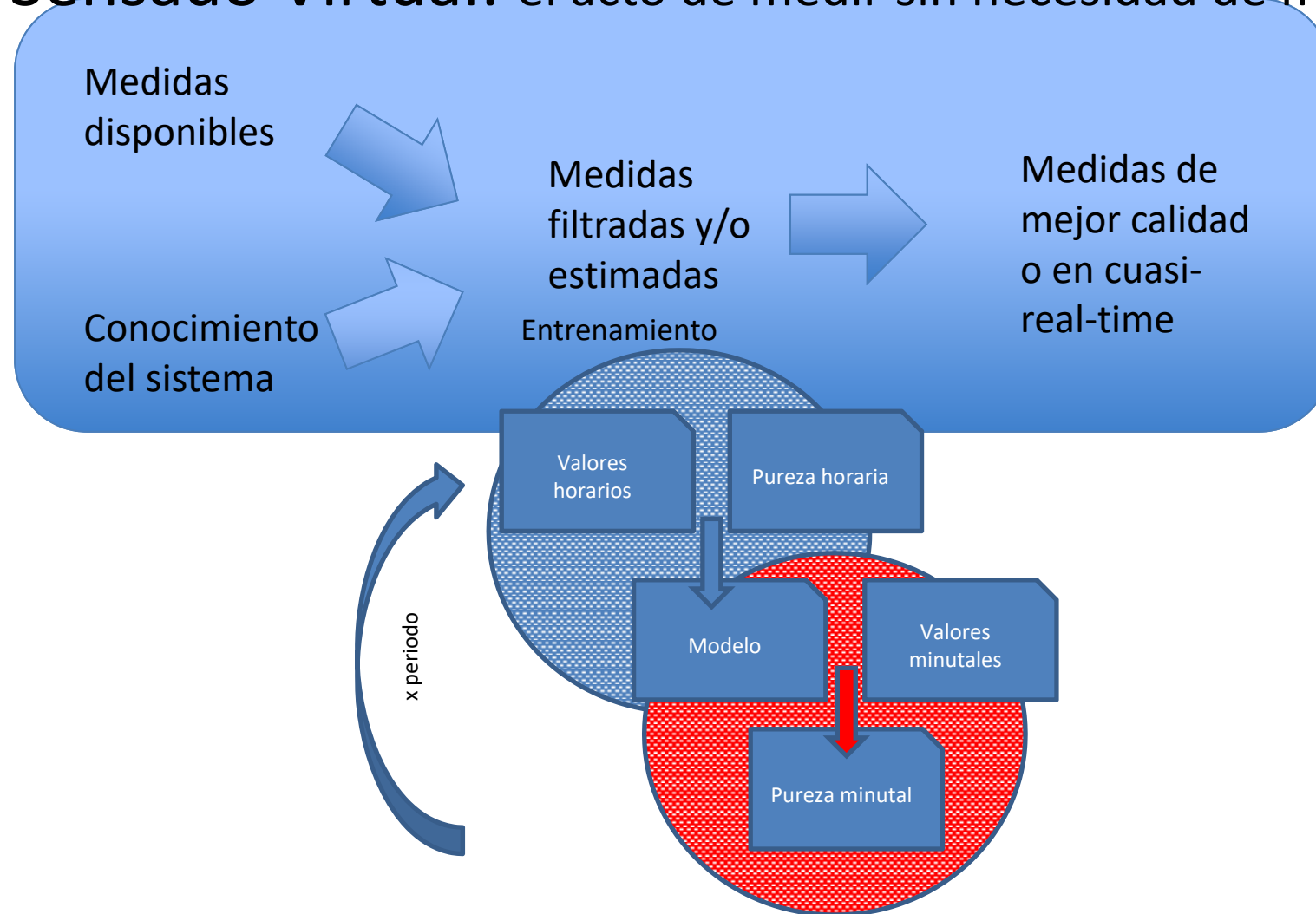


SENSORES VIRTUALES...

Sensado Virtual: el acto de conocer sin necesidad de medir

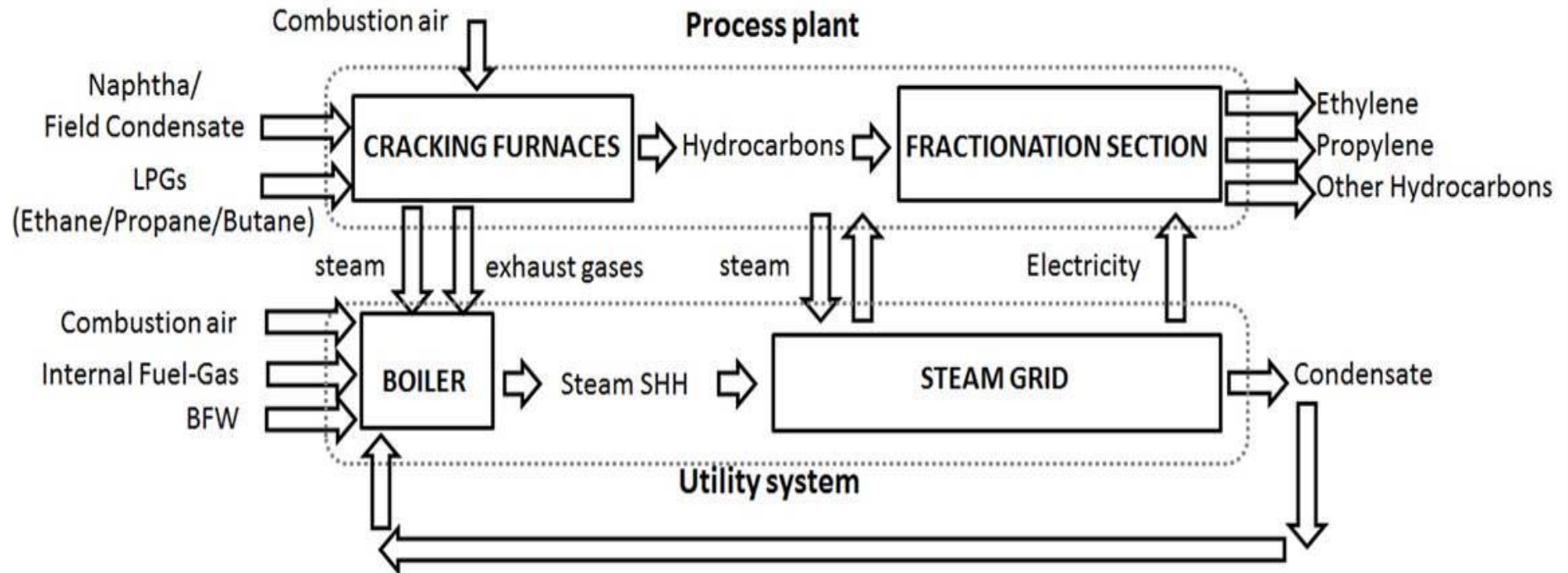
- Caso de Uso: desarrollo de sensores virtuales, ¿cómo inferir un valor a partir de medidas indirectas?
 - Objetivo: Modelización de la relación de causalidad existente entre un conjunto de magnitudes independientes (fáciles de medir) y una magnitud dependiente.
 - Problema Multi-paramétrico no-lineal que permite:
 - Modelos Físicos.
 - Modelos Híbridos.
 - Modelos “caja negra” o “data driven”. Analítica de Datos.
 - Ventajas: permite paliar los errores en la medida de sensores sometidos a condiciones extremas; permite paliar el decalaje temporal entre la toma de muestras y la obtención de la medida; y permite reducir la frecuencia entre muestras de una medida.
 - Clave de aplicación: control en real-time.
 - Proyecto: TOPREF, Soft Sensing, S. Gil-Lopez, I. Mendia, Z. Lizuain, F. Boto

Sensado Virtual: el acto de medir sin necesidad de medir

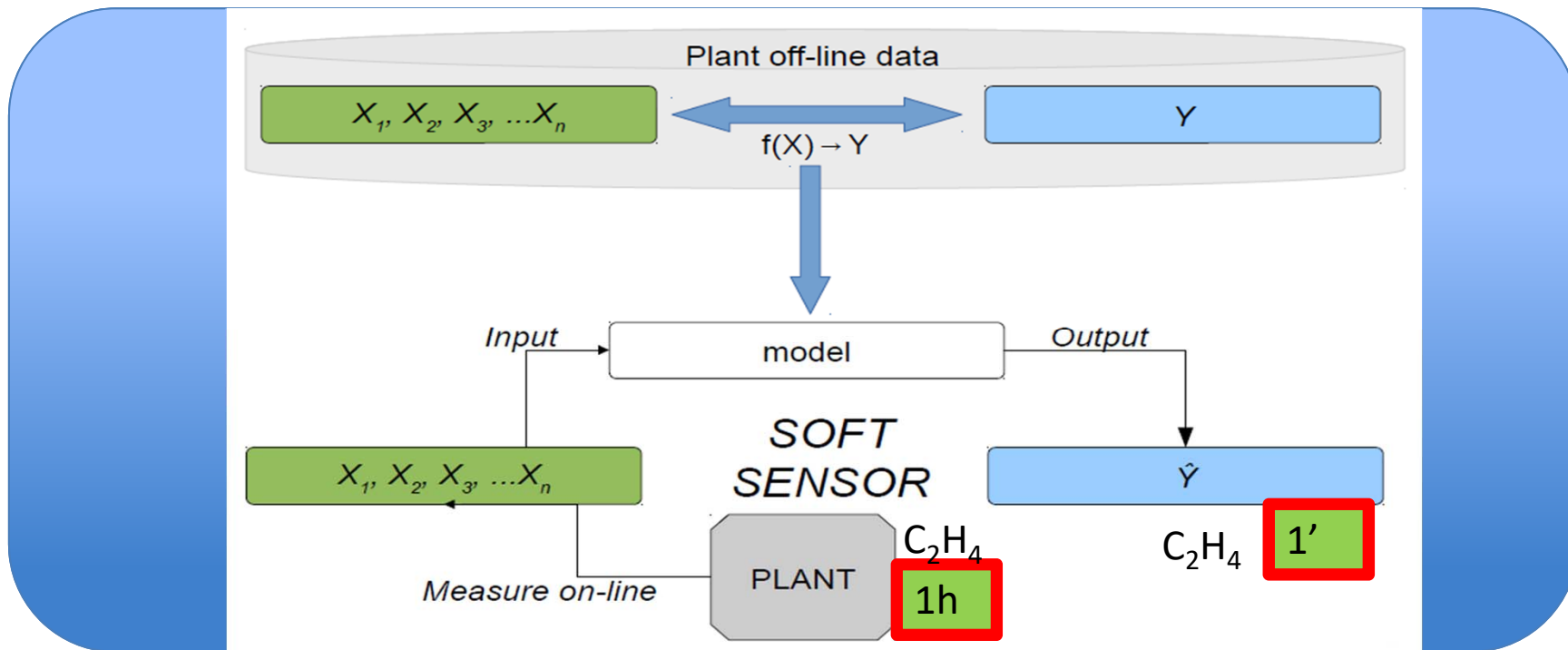


Sensado Virtual: el acto de medir sin necesidad de medir

Horno de cracking del nafta

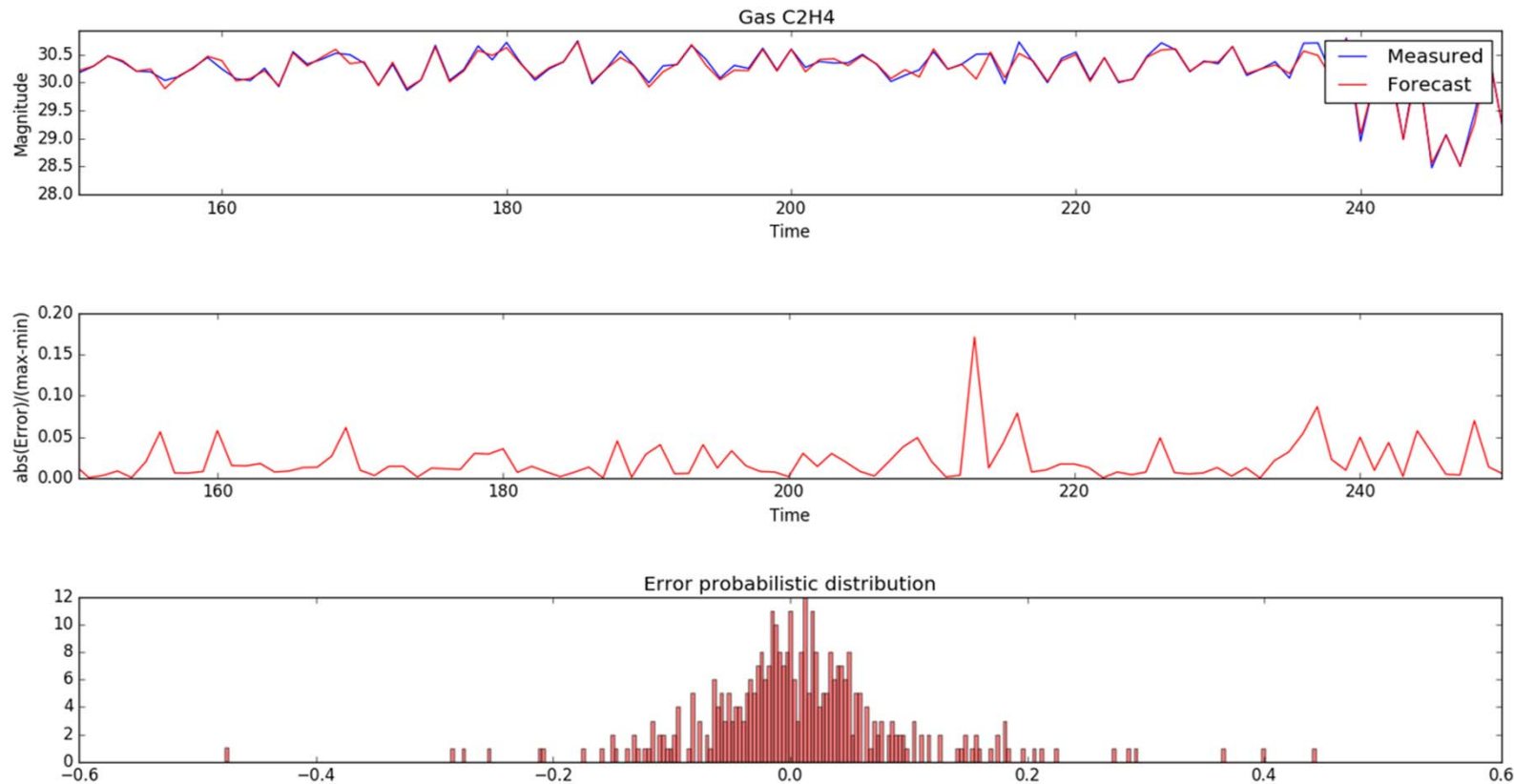


Sensado Virtual: el acto de medir sin necesidad de medir



Sensado Virtual: el acto de medir sin necesidad de medir

H ₂	CH ₄	C ₂ H ₄	C ₃ H ₆
0.015	0.008	0.017	0.007





CONCLUSIONES.



Conclusiones

- Estamos en pleno proceso de transformación digital.
- La Analítica de Datos será clave en los próximos años.
- Caso de uso I: Sistema inteligente de control de clima en edificios.
 - Los modelos “data driven” ofrecen mejor capacidad de generalización (otros edificios) que otro tipo de modelos.
 - Los modelos “data driven” ofrecen calidad en las magnitudes simuladas ($T_{\text{int}} \pm 2,5\%$; $E_e \pm 20\%$).
 - La Analítica de Datos permite optimizar el funcionamiento del clima de un edificio adaptándolo a las condiciones cambiantes sin afectar al confort.
 - Permiten la automatización del proceso y su interacción con los BEMs.
- Caso de uso II: Sensores virtuales.
 - La Analítica de Datos puede reducir un conjunto de datos en sus componentes fundamentales, eliminando información redundante.
 - El sensado virtual permite “medir” de manera indirecta magnitudes físicas a partir de medidas indirectas de otras magnitudes relacionadas.
 - Permite casar en tiempo la monitorización y el control, ofreciendo información en “real time” de magnitudes claves en el control de los procesos industriales.
 - Mejorará la eficiencia de los procesos industriales.
 - Permite reducir costes en la monitorización de determinados procesos.
 - La calidad de las estimaciones de los sensores virtuales es suficiente como para considerarlos en los procesos de control de las plantas industriales del futuro.

iENER'18

I Congreso Ingeniería Energética



GRACIAS POR SU ATENCIÓN

sergio.gil@tecnalia.com