

UNIVERSIDAD DE ALMERIA

ESCUELA SUPERIOR DE INGENIERÍA

**Desarrollo de modelos de predicción de
consumo de recursos en invernaderos**



Curso 2018/2019

Alumno/a:

Gabriel José Cantos Alcántara

Director/es:

Dr. D. Francisco Rodríguez Díaz
Dr. D. Manuel Berenguel Soria



AGRADECIMIENTOS

Quiero agradecer a todas aquellas personas que con su ayuda, han contribuido a la realización de este trabajo, ya fuese apoyo técnico o en el ámbito personal, especialmente:

- A los tutores Francisco Rodríguez Díaz Dirigida y Manuel Berenguel Soria, por la paciencia mostrada al dirigir este trabajo y sus aportaciones y correcciones al mismo.
- A todos los profesores e implicados que con su esfuerzo hicieron posible el curso 2017/18 del máster en Energía Solar de la Universidad de Almería.
- A mi familia y amigos que siempre se han interesado y me han dado su apoyo.

RESUMEN

La agricultura intensiva bajo plástico es una de las actividades de mayor impacto económico en la provincia Almería, la primera en el sector productivo. La rentabilidad económica de la actividad pasa por obtener el máximo beneficio de la venta de los productos en el mercado y la minimización de los costos de producción.

El cada vez más notable cambio climático tiene un efecto directo sobre la agricultura. Las variables climáticas suponen el factor determinante en el crecimiento y producción de los cultivos. La disponibilidad de recursos hidrológicos es también un problema sensible.

En este trabajo se aborda la predicción de la demanda eléctrica en invernaderos tecnificados en función de las variables climáticas exteriores y las condiciones climáticas requeridas al cultivo dentro del invernadero por parte del productor.

Para tal fin se aplican métodos de identificación de sistemas de caja negra lineales y no lineales. Antes de la generación de los modelos y su calibración, se analizan y tratan los datos de consumo, así como se estudian las variables a incluir en los modelos mediante los métodos de selección de variables.

Palabras clave: Agricultura, Identificación de sistemas, Demanda eléctrica

ABSTRACT

Intensive crops in greenhouses are one of the main economic activities of the province of Almería, the first in the productive sector. The economic profitability of the activity rely on obtaining the maximum incomes of the trade of products and the minimization of its production cost.

The more and more noteworthy climate change, also has a direct effect on agriculture. Climatic variables are the decisive factor in the growth and production of crops. The availability of hydrological resources is also a sensitive problem.

This work addresses the prediction of electrical demand in HiTech greenhouses, according to the external climatic variables and the climatic conditions required inside the greenhouse by the producer.

To do so, linear and non-linear black box system identification methods are applied. Before the creation and calibration of the models, consumption data are analyzed and process, as well as the variables to be included in the models are studied by variable selection methods.

Key words: Agriculture, System Identification, Electric demand

ÍNDICE GENERAL

1.	INTRODUCCIÓN	1
1.1.	Introducción	1
1.2.	Objetivos	2
1.3.	Contexto	2
1.4.	Resumen de resultados	3
1.5.	Fases de desarrollo	4
1.6.	Competencias del título utilizadas en el TFM	6
1.7.	Estructura de la memoria	7
2.	MATERIALES Y MÉTODOS	9
2.1.	Invernadero	9
2.2.	Softwares de desarrollo	16
2.3.	Técnicas de análisis de datos	18
2.4.	Técnicas de selección de variables de entradas	20
	2.4.1 Métodos visuales	20
	2.4.2 Métodos Basados en modelos	21
	2.4.3 Métodos sin modelos a partir de los datos observados	22
2.5.	Técnicas de modelado empírico	25
	2.5.1 modelos lineales	28
	2.5.2 Modelado no lineal	30
2.6.	Validación de modelos	34
3.	RESULTADOS	36
3.1.	Validación de los datos	36
3.2.	Selección de variables de entradas	61
	3.2.1 Análisis de correlación simple	61
	3.2.2 Análisis de componentes principales	64
	3.2.3 Correlación parcial	69
3.3.	Modelado empírico lineal	70
3.4.	Modelado empírico no lineal	80

3.5. Comparación de modelos	85
4. CONCLUSIONES Y FUTUROS TRABAJOS	92
BIBLIOGRAFÍA	92

ÍNDICE DE FIGURAS

1.1 Estructuras estudiadas en este trabajo y principales modelos obtenidos	3
1.2 Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción 5 minutos	4
1.3 Cronograma seguido para la ejecución del trabajo	5
2.1 Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].	10
2.2 Planta perfil y alzado de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].	11
2.3 De arriba abajo: Ventana cenital y ventana lateral de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].	11
2.4 Sistemas de calefacción por tuberías de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas. Bomba y tuberías a la izquierda y caldera de biomasa a la derecha [11].	12
2.5 Soplante de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].	13
2.6 Sistema de calefacción aeroterma de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas. Quemador diésel a la izquierda y ventilador de distribución a la derecha [11].	13
2.7 Máquina deshumificadora de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].	14
2.8 Captura de la interfaz gráfica de la herramienta de identificación de sistemas de Matlab, System Identification Toolbox	17
2.9 Red neuronal artificial estándar. Fuente: elaboración propia	31
2.10 Modelo de una neurona no lineal. Fuente: elaboración propia	32
3.1 Señal de radiación global horizontal (W/m^2) original y tratada durante el periodo de registro	37
3.2 Detalle de la señal de radiación global horizontal (W/m^2) original y tratada	37
3.3 Detalle 2 de la señal de radiación global horizontal (W/m^2) original y tratada	38
3.4 Señal original y tratada de la velocidad del viento (m/s) durante el periodo de registro	38

3.5	Detalle de la señal original y tratada de la velocidad del viento (m/s)	39
3.6	Señal original y tratada de la temperatura exterior (°C) durante el periodo de registro	39
3.7	Detalle de la señal original y tratada de la temperatura exterior (°C)	40
3.8	Señal original y tratada de la temperatura interior (°C) durante el periodo de registro	40
3.9	Detalle de la señal original y tratada de la temperatura exterior (°C)	41
3.10	Señal original y tratada de la temperatura del suelo (°C) durante el periodo de registro	41
3.11	Detalle de la señal original y tratada de la temperatura del suelo (°C)	42
3.12	Señal original y tratada de la humedad interior (%) durante el periodo de registro	42
3.13	Detalle de la señal original y tratada de la humedad interior (%)	43
3.14	Señal original y tratada de la humedad exterior (%) durante el periodo de registro	43
3.15	Detalle de la señal original y tratada de la humedad exterior (%)	44
3.16	Consigna de calefacción (°C) durante el periodo de registro	44
3.17	Zoom de la consigna de calefacción (°C) durante el periodo de registro	45
3.18	Detalle de la consigna de calefacción (°C) y señal lógica de actuación de la bomba calefacción	45
3.19	Señal de actuación de la bomba calefacción y consumo (kW) de la línea invernadero	46
3.20	Señal de actuación de la soplante durante el periodo de registro	46
3.21	Zoom de la señal de actuación de la soplante durante el periodo de registro	47
3.22	Señal de actuación del Aerotermo durante el periodo de registro	47
3.23	Zoom de la señal de actuación del Aerotermo durante el periodo de registro	48
3.24	Señal de actuación del Aerotermo y consumo de la línea invernadero (kW)	48
3.25	Consigna (°C) de la ventilación, establecida para la actuación automática de las ventanas durante el periodo de registro	49
3.26	Zoom de la consigna de la ventilación (°C), establecida para la actuación automática de las ventanas durante el periodo de registro	49
3.27	Señal de apertura (%) enviada a la ventilación cenital y lateral durante el periodo de registro	50
3.28	señales de actuación de las ventanas (p.u) y consumo de la línea invernadero (kW)	50
3.29	Señal de actuación de la máquina deshumificadora durante el periodo de registro	51

3.30 Zoom de la señal de actuación de la máquina deshumificadora durante el periodo de registro	51
3.31 Señal de actuación de la máquina deshumificadora y consumo de la línea invernadero (kW)	52
3.32 Señal original del consumo de la línea invernadero (kW) durante el periodo de registro	53
3.33 Detalle de la señal original del consumo de la línea invernadero (kW) durante el periodo de registro	53
3.34 Consumo línea invernadero (kW) una vez eliminados los valores negativos y las discontinuidades	54
3.35 Perfil tipo del consumo del equipo de medida de la línea invernadero (SCADA+sensores)	55
3.36 Perfil tipo del consumo de la bomba de calefacción (kW)	55
3.37 Perfil tipo del consumo (kW) de la máquina deshumificadora	56
3.38 Consumo Desconocido	57
3.39 Señal del consumo de la línea invernadero original y depurada de valores extraños	58
3.40 Detalle de la señal del consumo de la línea invernadero original y depurada de valores extraños	58
3.41 Señal original del consumo de la línea biomasa (kW) durante el periodo de registro	59
3.42 Zoom de la señal original del consumo de la línea biomasa (kW) durante el periodo de registro	59
3.43 Valores atípicos registrados en la señal consumo de la línea biomasa de la semana 14 a 30	60
3.44 Consumo línea biomasa Detalle	60
3.45 Varianza explicada por cada componente	65
3.46 Correlación de cada variable original con las componentes principales 1 y 2	66
3.47 Correlación de cada variable original con las componentes principales 3 y 4	66
3.48 Correlación de cada variable original con las componentes principales 5 y 6	67
3.49 Varianza explicada por cada componente	68
3.50 Correlación de cada variable original con las componentes principales 1 y 2	69
3.51 Conjunto de datos seleccionados para la estimación (naranja) y para la validación cruzada 1 (amarillo) y validación cruzada 2 (rojo)	70
3.52 Conjunto de datos para la estimación (rojo), validación cruzada 1 (azul) y validación cruzada 2 (morado). Par e-s, radiación-consumo	71

3.53 Conjunto de datos para la estimación (naranja), validación cruzada 1 (amarillo) y validación cruzada 2 (rojo). Par e-s, bomba calefacción-consumo	71
3.54 Conjunto de datos para la estimación (naranja), validación cruzada 1 (amarillo) y validación cruzada 2 (rojo). Par e-s, Deshumificadora -consumo	72
3.55 Consumo real del invernadero (curva negra) y predicho por los modelos (Arx111 en verde y Arx721 en morado) basados en el estado de actuadores, para el conjunto de datos de estimación	74
3.56 Consumo real del invernadero (curva negra) y predicha por los modelos (Arx111 en verde y Arx721 en morado) basados en el estado de actuadores, para el conjunto de datos de validación	75
3.57 Consumo real del invernadero (curva negra) y predicho por los modelos (Arx111 en verde y Arx721 en morado) basados en el estado de actuadores, para el conjunto de datos de validación 2	76
3.58 Análisis residual. En la parte superior la autocorrelación. En la parte inferior la correlación cruzada entre radiación y el consumo	76
3.59 En la parte superior la autocorrelación. En la parte inferior la correlación cruzada entre temperatura exterior y el consumo	77
3.60 En la parte superior la autocorrelación. En la parte inferior la correlación cruzada entre la señal de la bomba de calefacción y el consumo	77
3.61 Consumo real del invernadero (curva negra) y predicho por los modelos (ArxC111 en rojo y ArxC510 en morado) basados en consignas, para el conjunto de datos de estimación	79
3.62 Consumo real del invernadero (curva negra) y predicho por los modelos (ArxC111 en rojo y ArxC510 en morado) basados en consignas, para el conjunto de datos de validación	79
3.63 Consumo real del invernadero (curva negra) y predicho por los modelos (ArxC111 en rojo y ArxC510 en morado) basados en consignas, para el conjunto de datos de validación 2	80
3.64 Captura de la selección del número de nodos a la derecha y valores pasados para cada regresor a la izquierda.	81
3.65 Consumo real del invernadero (curva negra) y predicha por los modelos no lineales basados en consignas (narx111 en gris y narx 510 en verde), para el conjunto de datos de estimación	81
3.66 Consumo real del invernadero (curva negra) y predicha por los modelos no lineales basados en consignas (narx111 en gris y narx 510 en verde), para el conjunto de datos de validación	82

3.67 Consumo real del invernadero (curva negra) y predicha por los modelos no lineales basados en consignas (narx111 en gris y narx 510 en verde), para el conjunto de datos de validación 2	82
3.68 Consumo real del invernadero (curva negra) y predicha por los modelos no lineales sin la variable tiempo (narx310ST en rojo, narx111ST en magenta y narx510ST en azul), para el conjunto de datos de estimación	84
3.69 Consumo real del invernadero (curva negra) y predicha por los modelos no lineales sin la variable tiempo (narx310ST en rojo, narx111ST en magenta y narx510ST en azul), para el conjunto de datos de Validación	84
3.70 Consumo real del invernadero (curva negra) y predicha por los modelos no lineales sin la variable tiempo (narx310ST en rojo, narx111ST en magenta y narx510ST en azul), para el conjunto de datos de Validación 2	85
3.71 Clasificación de las diferentes estructuras probadas y los mejores modelos de cada una	86
3.72 Consumo real y su predicción con los modelos indicados, aplicados a los datos de validación. Horizonte de predicción 30 segundos	86
3.73 Consumo real y su predicción con los modelos indicados aplicados a los datos de validación. Horizonte de predicción 5 minutos	87
3.74 Consumo real y su predicción con los modelos indicados aplicados a los datos de validación. Horizonte de predicción 10 minutos	88
3.75 Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción30 segundos	89
3.76 Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción 5 minutos	90
3.77 Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción 10 minutos	91

ÍNDICE DE TABLAS

1.1 Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 5 minutos, para un periodo de funcionamiento de calefacción y ventilación	4
1.2 Resumen de Resultados. Estructuras de los mejores modelos generadas	4
2.1 Consumo de los equipos instalados en el invernadero [15]	15
3.1 Conjunto de variables seleccionadas para el estudio	36
3.2 Variables estudiadas en la selección de variables	61
3.3 Coeficientes de correlación simple bivariada	62
3.4 Matriz de significancia	62
3.5 Matriz de coeficientes de determinación	64
3.6 Variabilidad aportada por cada componente	64
3.7 Matriz de coeficientes de correlación entre las variables originales y las nuevas componentes ortogonales	65
3.8 Variabilidad aportada por cada componente	68
3.9 Matriz de coeficientes de correlación entre las variables originales y las nuevas componentes ortogonales de la 2º iteración del análisis de componentes	68
3.10 Coeficientes lineales de correlación parcial entre y y el resto de variables	70
3.11 Variables usadas en la construcción de los modelos Arx basados en estados de actuadores	72
3.12 Combinaciones de estructuras probadas para el modelo Arx basado en actuadores y selección de las más óptimas	73
3.13 Ajustes y error de las estructuras seleccionadas para los modelos Arx basados en actuadores	73
3.14 Variables usadas en la construcción de los modelos Arx basados en consignas	78
3.15 Mejores estructuras generadas para los modelos Arx basados en consignas, sus ajustes y errores de predicción	78
3.16 Ajustes y errores de predicción de los modelos Narx basados en consignas y sin tiempo generados	83

3.17 Estadísticos de la diferencia entre los valores de consumo reales y estimados por los mejores modelos generados con un horizonte de predicción de 30 segundos, para los datos de validación.	87
3.18 Estadísticos de la diferencia entre los valores de consumo reales y estimados por los mejores modelos generados con un horizonte de predicción de 5 minutos, para los datos de validación.	88
3.19 Estadísticos de la diferencia entre los valores de consumo reales y estimados por los mejores modelos generados con un horizonte de predicción de 10 minutos, para los datos de validación.	88
3.20 Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 30s, para un periodo de funcionamiento de calefacción y ventilación	89
3.21 Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 5 minutos, para un periodo de funcionamiento de calefacción y ventilación	90
3.22 Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 10 minutos, para un periodo de funcionamiento de calefacción y ventilación	91

1. INTRODUCCIÓN

1.1. Introducción

El cultivo bajo plástico o en invernadero es una actividad económica que ha sufrido un gran desarrollo y expansión en los últimos años [1] [2]. Por ello se busca maximizar el rendimiento económico y la competitividad de las instalaciones, entre otras formas, incorporando las nuevas tecnologías [3].

El crecimiento de la población mundial también exige un aumento de la producción de alimentos [4], que supone un porcentaje alto del gasto de agua y energía a nivel mundial. Podemos decir entonces que el ahorro de energía en invernaderos es beneficioso tanto para productores como para la sociedad [5].

En este aspecto, existen dos variables que han de ser optimizadas. Por un lado la calidad en la producción, que trata de maximizar los beneficios obtenidos de la venta, produciendo productos de la mayor calidad posible así como su cantidad. Para ello se busca que el clima dentro del invernadero sea lo más adecuado posible. Por otro lado los costes de la producción y operación en el invernadero han de ser minimizados para obtener la mayor rentabilidad posible. Se trata pues de un compromiso entre el precio de venta del producto y el costo de su producción [6]. Para lograr ambos objetivos se necesita establecer y seguir unas consignas de las variables que influyen sobre el crecimiento y producción de los cultivos. Estas se eligen teniendo en cuenta entre otras, las condiciones meteorológicas, tipo y estado del cultivo o influencias del mercado [5], [6].

En este sentido, el invernadero al tratarse de un espacio cerrado permite controlar las variables climáticas y reducir las necesidades de riego y fertilizantes [7]. Para ello, los invernaderos se tecnifican, dotándolos de equipos capaces de controlar las variables climáticas, como ventilación natural bajo consigna, calefacción o enriquecimiento de CO₂. La actuación automática de estos equipos supone una demanda energética en función de las condiciones ambientales exteriores y la que se requiere en el invernadero. Esta demanda es la que en este trabajo se pretende predecir.

El control automático, con experiencia previa en muchas otras industrias, se incorpora a los cultivos bajo plástico permitiendo manipular calefacción, ventilación, mallas de sombra, luz, o inyección de CO₂ para obtener los valores deseados de temperatura, humedad interior, radiación sobre los cultivos o influenciar la fotosíntesis [6] [8]. En cultivos mediterráneos las variables manipulables que se utilizan suelen ser la ventilación, calefacción y mallas de sombreado. [1], [9]. La provincia de Almería presenta la mayor concentración de invernaderos del mundo con 35.489 hectáreas [10] y representa la principal actividad económica del sector productivo de la región. Por ello se hace indispensable la incorporación de la tecnología para garantizar la rentabilidad y la sostenibilidad de la actividad en un sector cada vez más competitivo y afectado

por el cambio climático [7].

1.2. Objetivos

El objetivo principal de este trabajo es diseñar e implementar un modelo predictivo de la demanda de energía eléctrica a corto y medio plazo de un invernadero tecnificado. A tal fin se usará como referencia un invernadero piloto existente situado en la estación experimental de la fundación Cajamar en El Ejido denominada “Las Palmerillas”. Los subobjetivos para llevar a cabo y lograr el objetivo principal serán:

- Tratamiento de los datos y división en conjuntos de entrenamiento, validación y prueba. Para ello se realizará un filtrado para eliminar las falsas medidas y el ruido, así como una normalización de los datos.
- Realizar un análisis del consumo energético de un invernadero tecnificado y establecer los patrones de consumo de la instalación sobre la base de los datos experimentales existentes y, a efectos de generalización, sobre la base de las especificaciones de elementos y sistemas y de las pautas de operación de los mismos en las distintas fases del cultivo.
- Selección de variables de entrada para los modelos. Previamente se preseleccionará un conjunto de variables de entrada candidatas, de las que se seleccionarán las variables de entrada finales a lo largo del proceso de desarrollo del modelo.
- Selección del tipo y arquitectura del modelo, así como su calibración y validación. Por último se comparan los modelos generados.

1.3. Contexto

Este trabajo se enmarca dentro del proyecto CHROMAE Project (DPI2017-85007-R) [11]., que busca mejorar la gestión y eficiencia del consumo de recursos en el ámbito de la agricultura industrial, haciendo uso para ello de las herramientas y técnicas de la ingeniería de sistemas y control automático. En dicho proyecto participa el grupo de investigación ARM (TEP-197) del departamento de informática de la universidad de Almería.

Uno de los objetivos principales del proyecto es identificar las necesidades de demanda, así como las posibilidades de producción de los recursos que intervienen en áreas agro-industriales, como Almería. De esta forma se busca crear sinergias que permitan una mejor gestión y aprovechamiento de los recursos, con la subsecuente reducción de coste a la vez que se mejora la sostenibilidad de la actividad [12].

En este aspecto, el proyecto recoge y relaciona las actividades de diferentes sectores como:

- Agricultura en invernaderos con diferentes grados de tecnificación

- Desalación de agua
- Producción de energía eléctrica a partir del recurso solar
- Gestión inteligente en la edificación y transporte
- Aprovechamiento de los residuos agrícolas

Este trabajo concretamente, se centra en la previsión de la demanda eléctrica en invernaderos tecnificados. Para ello hace uso del invernadero nº 7 de la estación experimental "Las Palmerillas". Este invernadero de 877 m² es una de las instalaciones dentro del proyecto CHROMAE, debido a que se encuentra perfectamente monitorizado, registrándose tanto las variables climáticas, actuación de los controladores y consumo eléctrico. En este trabajo se hace uso de estos datos para la creación de los modelos predictivos de demanda eléctrica en invernaderos.

1.4. Resumen de resultados

Este trabajo se compone de diferentes etapas, siendo el resultado de cada etapa necesario para la ejecución de la siguiente. Si bien los modelos obtenidos para la predicción suponen los resultados más importantes, el tratamiento de datos así como la selección de variables también suponen aportaciones importante.

En la figura 1.1 se muestra un diagrama con las principales estructuras de modelos predictivos consideradas en este trabajo así como los mejores modelos obtenidos a partir de ellas.

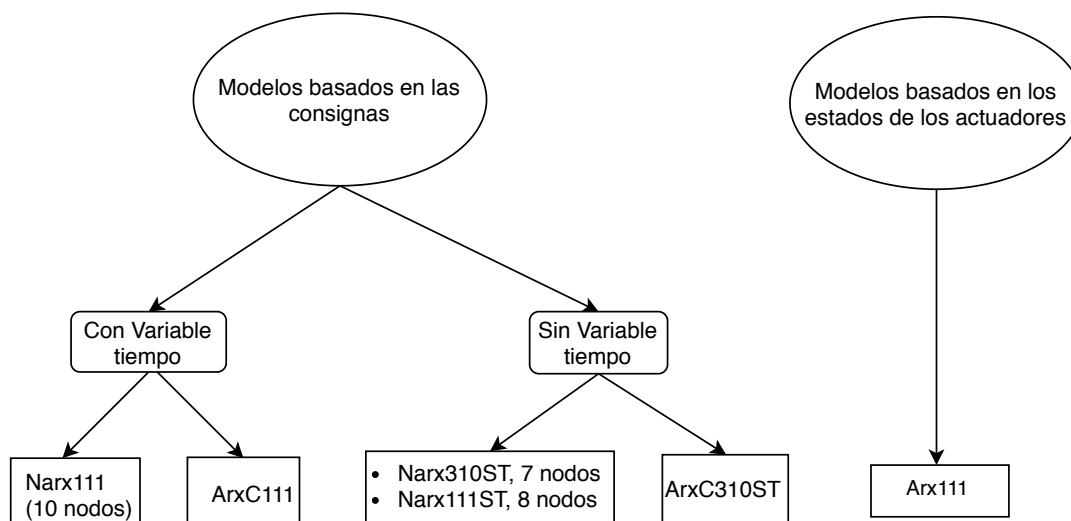


Fig. 1.1. Estructuras estudiadas en este trabajo y principales modelos obtenidos

En la gráfica 1.2 se muestra el resultado de la predicción con los modelos basados en consignas para un horizonte de predicción de 5 minutos en un periodo con calefacción y ventilación.

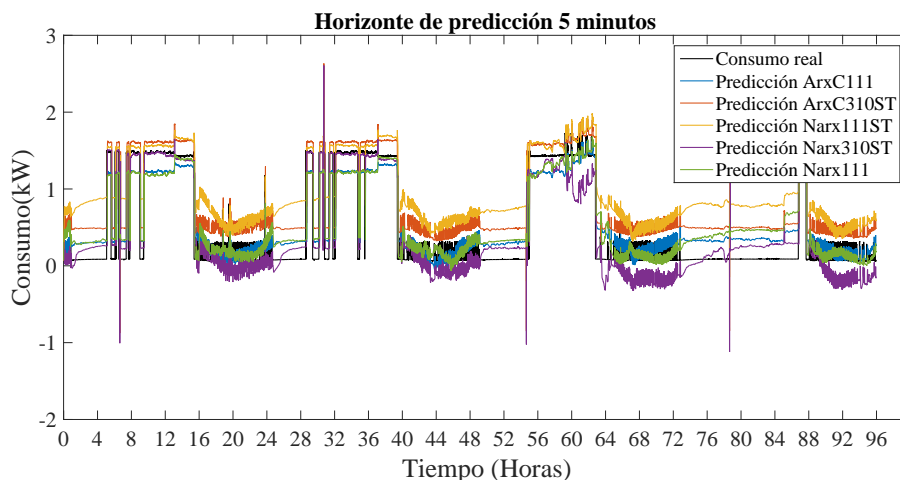


Fig. 1.2. Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción 5 minutos

Cuyos datos cuantitativos para la comparación de los estadísticos de los residuos se muestran en la siguiente tabla:

TABLA 1.1. Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 5 minutos, para un periodo de funcionamiento de calefacción y ventilación

	Ajuste	Máx. error	Media error	Varianza	Desviación estándar
ArxC111	56.40	1.335	0.205	0.0274	0.1656
ArxC310ST	36.47	1.596	0.3351	0.0355	0.1884
Narx111ST	10.05	1.751	0.474	0.0719	0.268
Narx310ST	53.55	1.494	0.188	0.0438	0.2093
Narx111	53.01	1.511	0.246	0.0342	0.1849

Los resultados de las diferentes comparaciones llevadas a cabo indican que los mejores resultados se obtienen con modelos de ordenes bajos, los cuales se muestran en la tabla:

TABLA 1.2. Resumen de Resultados. Estructuras de los mejores modelos generadas

	N° de nodos	na	nb	nk
ArxC111	-	1	1 2 1 1 1 1 1 1 2	1 1 0 1 0 1 0 0 0
Narx111	11	1	1 2 1 1 1 1 1 1 2	1 1 0 1 0 1 0 0 0

1.5. Fases de desarrollo

El punto de partida del trabajo es la tenencia de una serie histórica de datos de consumo de diferentes recursos en un invernadero. Así como las condiciones ambientales, hora, época del

año, tipo y estado del cultivo. A partir de estos datos se realizara el proceso de identificación para obtener un modelo del sistema que satisfaga los objetivos planteados. Para la consecución de dichos objetivos se seguirán los siguientes pasos:

- Filtrado de datos. Observación visual de los datos medidos con la intención de eliminar medidas erróneas, posible ruido o reconstruir señales incompletas.
- Selección de variables. Hay que escoger un conjunto adecuado de datos entrada-salida, que permita el desarrollo y entrenamiento del modelo, donde las entradas pertenecen a un tipo de variable. Por tanto deben elegirse aquellas variables que afecten a la dinámica del proceso que se desea predecir y desechar las variables que no afectan al sistema o aquellas variables redundantes debido a que están fuertemente relacionadas entre sí.
- Selección de la estructura y los modelos: se seleccionaran los modelos que se usaran para construir el predictivo. Modelos lineales y no lineales serán propuestos y evaluados.
- Ajuste de parámetros y validación: consiste en usar un conjunto de datos reales del sistema, para ajustar los parámetros del modelo elegido, hasta conseguir que la salida calculada reproduzca la salida real. Lugo se valida el modelo evaluando si la salida calculada por el modelo es igual a la salida real con cierto error aceptable. Los datos con los que se compara la salida del modelo serán diferentes del conjunto de datos usados en el ajuste.
- Elaboración de la memoria del Trabajo Fin de Máster.

En la figura 1.3 se muestra además el cronograma asociado a la elaboración del TFM.

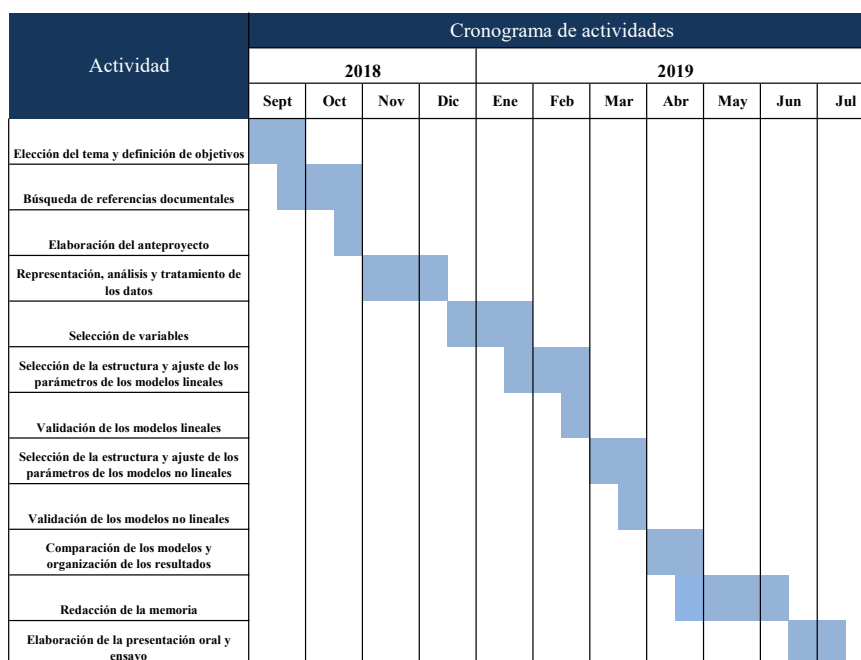


Fig. 1.3. Cronograma seguido para la ejecución del trabajo

El proyecto ha tenido una duración de 11 meses. En los cuales se han dedicado 12 horas semanales de trabajo, lo que supone un total de 528 horas dedicadas.

1.6. Competencias del título utilizadas en el TFM

Gran parte de las competencias adquiridas en la realización de los estudios del Máster en energía solar han sido empleadas en este trabajo, y como consecuencia reforzadas. A continuación se muestran aquellas que han sido de aplicación directa:

Competencias básicas y generales:

- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.

El carácter de este trabajo es de investigación. Aunque el número de créditos asignados dificulta el empleo del tiempo necesario para desarrollar una labor de investigación, esta sería imposible sin los conocimientos base que han permitido adquirir y aplicar los nuevos conocimientos necesarios.

- CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

La realización de este trabajo ha implicado la extensión de los conocimientos adquiridos en la asignatura de 'energía solar y agricultura' a otros campos de estudio como la estadística y los modelos predictivos. Campos que son nuevos para el autor, lo que resulta en mayor complejidad a la vez que se adquieren nuevos conocimientos y aplican otros ya adquiridos.

- CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.

Sin duda esta competencia juega un papel fundamental. Si bien es cierto que se dispone de un conjunto de datos reales a partir de las mediciones, el hecho de no haber estado presente en la experimentación dificulta la interpretación. En primer lugar el autor se familiarizó con las variables disponibles, nomenclatura y periodos de registro. Tras esto resulta difícil conocer que equipos operaron, si de forma manual u automática, sin la ayuda por ejemplo de un cuaderno de notas. Como consecuencia se han tomado decisiones, que aunque simplificando, han permitido avanzar y obtener un resultado lógico y adecuado.

- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Tanto en la realización de la memoria como de la presentación oral, el autor intenta exponer de manera simple y clara todos los conceptos necesarios, mediante ejemplos, representaciones gráficas o cálculos.

- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

En gran medida este trabajo ha sido realizado de manera autónoma recurriendo a las referencias y a la documentación en todo momento, también requiriendo información o guía a los tutores cuando fue necesario.

Competencias específicas:

- CE4 - Aprender el conocimiento y desarrollar la capacidad para la aplicación de energía solar en agricultura.

El conocimiento previo de la dinámica dentro del invernadero ha facilitado tanto la interpretación de los datos, así como la selección de variables y creación de modelos.

- CE13 - Aprender el conocimiento y desarrollar la capacidad para realizar una labor de investigación dentro del campo de la energía solar.

El desarrollo de este trabajo ha implicado la aplicación de técnicas de investigación dentro del campo de la energía solar. Tales conocimientos van desde la búsqueda bibliográfica, uso de software adecuados, obtención de conclusiones o redacción de la memoria.

- CE15 - Realizar, presentar y defender un ejercicio original realizado individualmente ante un tribunal formado por profesores del máster, consistente en un proyecto de naturaleza profesional o un trabajo de investigación donde se sinteticen las competencias adquiridas en las enseñanzas.

Todo el trabajo aquí expuesto, junto con las presentaciones y defensa oral de la memoria ha sido desarrollada de forma individual. La toma de decisiones se hace indispensable ante las dificultades y diferentes caminos que se ofrecen en su desarrollo. El trabajo autónomo junto con las guías de los tutores permiten el desarrollo y finalización satisfactoria del trabajo.

1.7. Estructura de la memoria

La estructura de la memoria presenta diferentes partes que se organizan para facilitar su comprensión.

De esta forma en el capítulo I se realiza una breve introducción del problema que se aborda y el contexto, así como se establecen los objetivos del trabajo. También se muestra un resumen de los resultados, la planificación seguida en la realización del trabajo y las competencias utilizadas.

En el capítulo 2 se introducen los materiales empleados en el trabajo, tanto para la adquisición de los datos como para su posterior análisis, procesado y obtención de resultados. Además se introducen los conceptos teóricos de análisis de datos, selección de variables y de modelado empíricos.

Estos conceptos son aplicados en el apartado 3 cuyos resultados son mostrados gráfica y numéricamente para los diferentes modelos creados. En la última parte del capítulo se muestra la comparación entre los diferentes modelos creados.

En el capítulo 4 se exponen las principales conclusiones del trabajo y se proponen futuras actuaciones para complementar y ampliar el desarrollo de este trabajo.

En el último apartado se muestran las referencias bibliográficas empleadas.

2. MATERIALES Y MÉTODOS

Los medios materiales adquieren en este trabajo un carácter especial. Debido a la utilización de datos reales, se presenta de forma detallada los equipos utilizados para su adquisición. Posteriormente se describen las técnicas, métodos y software empleados para alcanzar los objetivos planteados en el apartado 1.0.2. a partir de los datos brutos obtenidos empíricamente.

Para la realización del Trabajo Fin de Máster han sido necesarios los siguientes materiales y métodos.

2.1. Invernadero

El invernadero es la instalación principal, donde se encuentran la mayoría de los equipos empleados en este estudio. Esta instalación no es un invernadero al uso. Se trata de un centro de investigación experimental, por lo que se encuentra altamente tecnificado, dotado con múltiples equipos para mejorar la producción, así como los sensores necesarios para su monitorización y evaluación.

Concretamente el invernadero se encuentra situado en la Estación Experimental de la Fundación Cajamar “Las Palmerillas”, en el municipio Almeriense de El Ejido [13]. Dicha estación está dedicada a la mejora de la producción de los cultivos bajo plásticos en climas mediterráneos, como el Almeriense, mediante técnicas respetuosas con el medio ambiente y haciendo uso de recursos locales. Esto permite poner en valor recursos locales que de otra forma no son aprovechados, reduciendo así el consumo de otros recursos como combustibles o electricidad. De esta forma se busca incrementar el beneficio económico mejorando la productividad a la vez que se es medioambiental mente sostenible. Todo esto bajo la máxima de que las técnicas y conocimientos desarrollados logren la implementación comercial en el sector de los invernaderos siendo aceptadas y factible para los productores [13].

Aunque la estación cuenta con múltiples invernaderos con diferentes cultivos, en este trabajo se ha hecho uso del invernadero número 7. El motivo es que el grupo de investigación de Automática, Robótica y Mecatrónica de la universidad de Almería, TEP-197, tiene instalados allí un conjunto de sensores y equipo de adquisición de datos. Estos equipos registran cada 30 segundos las condiciones climáticas en el interior de la nave así como el consumo eléctrico de cada uno de los equipos destinados a la mejora de la productividad de los cultivos mediante el control del clima del invernadero. Los equipos destinados a controlar las variables climáticas dentro del invernadero, pueden funcionar manualmente, mediante ensayos, o programados de forma automática. De esta forma se registra una serie histórica de datos que permite conocer con gran exactitud la operación de los equipos y las condiciones climáticas en cada momento del día. El banco de datos usado se ha obtenido durante la campaña de 2015 comenzando en Enero de 2015 y finaliza en diciembre de ese mismo año.

La nave 7 se trata de un invernadero tipo raspa y amagado formado por cinco capillas, con una superficie de 877 metros cuadrados con una altura de 4,4 m en la capilla y 3,6 m en el amagado. Está orientado Norte-Sur. En la figura 2.1 se muestra una captura del exterior de este invernadero. Durante la campaña de recogida de datos el tipo de cultivo en el interior del invernadero era tomate.



Fig. 2.1. Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].

Como se comentaba anteriormente el invernadero esta equipado con una serie de sistemas encargados de mejorar la producción del cultivo a la vez que se minimizan los costes para obtener la mayor rentabilidad posible. Para lograr ambos objetivos se necesita establecer y seguir unas consignas de las variables que influyen sobre el crecimiento y producción de los cultivos [4], [6]. Estas consignas responden a variables, principalmente climáticas, que deben estar dentro de unos límites para asegurar la calidad y productividad. En el invernadero de estudio existen los siguientes sistemas de control:

- Sistema de ventilación natural: el invernadero posee un sistema de ventilación natural consistente en aperturas motorizadas, de forma que su operación es controlada de acuerdo a la programación. El objetivo principal de la ventilación es procurar que la temperatura interior del invernadero no alcance valores elevados. Para conseguir tal fin, los motores de las ventanas regulan su apertura de forma automática conforme a la consigna de temperatura interior establecida. La ventilación también permite evitar valores elevados de humedad en el interior así como renovar el aire interior igualando los niveles de CO_2 con el exterior.

El invernadero cuenta con una apertura cenital en cada una de las cinco capillas, con apertura hacia el oeste y dos ventanas laterales, una en la cara norte y la otra en la cara sur.

En la figura 2.2 se muestra la planta perfil y alzado de la nave 7, donde se ilustra la disposición de las ventanas. Por otra parte en la figura 2.3 se muestra una captura real de las ventanas.

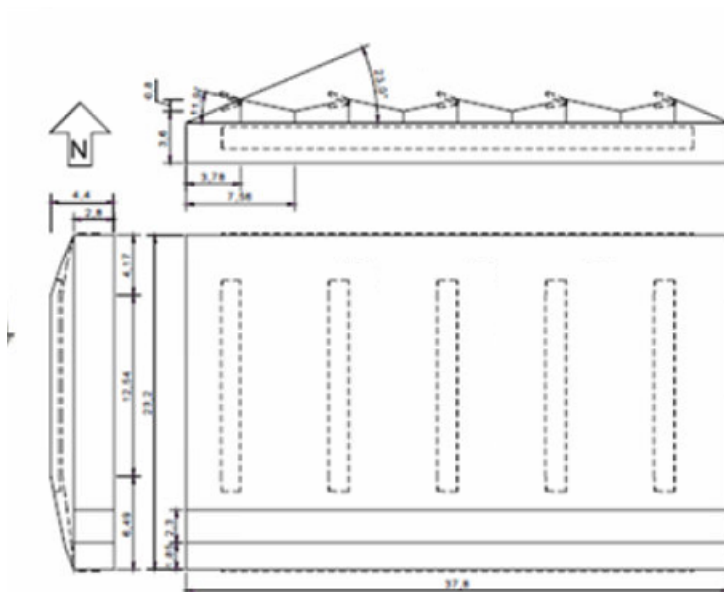


Fig. 2.2. Planta perfil y alzado de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11]



Fig. 2.3. De arriba abajo: Ventana cenital y ventana lateral de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].

- Sistema de calefacción por tuberías: este sistema se encarga de que la temperatura del cultivo no caiga por debajo de cierto límite. Para ello cuenta con una red de tuberías que recorre el interior del invernadero transportando agua caliente, que por convección calienta el aire interior de la nave. El agua se calienta en una caldera de biomasa y es recirculada mediante una bomba eléctrica.

Este sistema adquiere especial importancia en las noches de invierno. Momentos en los que la temperatura interior puede bajar de los 10 °C, valor al que se suele fijar la consigna de activación. Por debajo de esta temperatura el crecimiento del cultivo de tomate se ve afectado negativamente [7]. Experiencias en la estación de las palmerillas muestran aumentos en la producción de hasta el 24 % con el uso de calefacción [14].

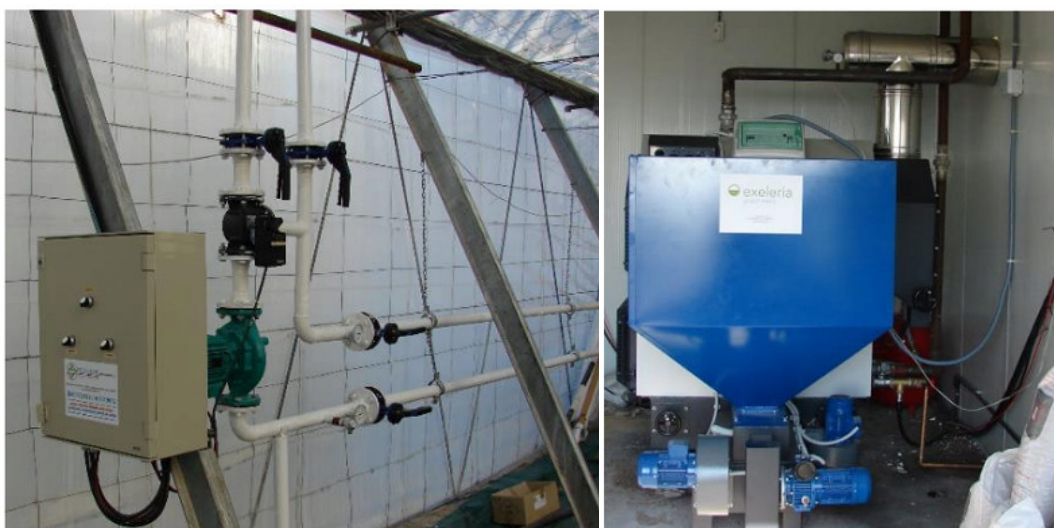


Fig. 2.4. Sistemas de calefacción por tuberías de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas. Bomba y tuberías a la izquierda y caldera de biomasa a la derecha [11].

- Soplante: se trata de un sistema que recoge los gases de combustión de la caldera de biomasa de la calefacción y los almacena en un tanque a una presión de 1.9 bar. Estos gases de combustión son ricos en CO_2 y son usados para el enriquecimiento carbónico del cultivo, factor que puede mejorar el rendimiento de la producción notablemente. Experiencias en las palmerillas han demostrado que existe una sinergia cuando se realiza calefacción y enriquecimiento carbónico. En estas situaciones la adición de CO_2 puede elevar la producción hasta un 26 % [14].

Por tanto la soplante funciona, si es necesario, cuando lo hace la calefacción mediante tuberías. En la figura 2.5 se muestra una fotografía de la soplante instalada en el invernadero.



Fig. 2.5. Soplante de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11].

- **Calefacción aeroterma:** al igual que la calefacción mediante tuberías sirve para calefactar el interior del invernadero. En este caso se trata de un sistema de calefacción por aire, donde un quemador diésel situado dentro del invernadero quema el combustible pulverizado. Los gases de combustión son expulsados directamente en el interior del invernaderos y repartidos mediante un ventilador de distribución. Estos gases de combustión, además de calefactar producen enriquecimiento carbónico del invernadero.



Fig. 2.6. Sistema de calefacción aeroterma de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas. Quemador diésel a la izquierda y ventilador de distribución a la derecha [11].

- **Sistema de deshumificación:** se trata de una máquina deshumificadora que se encarga de

eliminar humedad dentro del invernadero mediante su condensación. La humedad excesiva disminuye la transpiración de las plantas y aumenta el riesgo de proliferación de hongos.

La humedad en el interior se ve afectada por la temperatura interior, así como el grado de infiltración del invernadero y la transpiración del cultivo. Se debe evitar la condensación en formas de gotas en las superficies interiores del invernadero. Son especialmente delicados los amaneceres y puestas de sol de invierno. La ventilación natural es efectiva contra el exceso de humedad, sin embargo, mantener la consigna de temperatura interior tiene prioridad. En la figura 2.7 se muestra una imagen real de la máquina deshumidificadora.



Fig. 2.7. Máquina deshumidificadora de la Nave 7 de la Estación Experimental de la Fundación Cajamar “Las Palmerillas” [11]

- Sistema de seguimiento y monitorización: se trata de un conjunto de sensores instalados en diferentes puntos del invernadero, destinados a medir todas las variables climáticas que afectan al invernadero. En el exterior se encuentra una estación meteorológica completa, puede verse en la figura 2.1. En el interior se hayen otro conjunto de sensores destinados a medir variables del interior del invernadero, como humedad, radiación par, temperatura o concentración de CO_2 . Estos sensores transforman las variables físicas en vatios, los cuales son medidos por vatímetros tipo SINEAX M563. Los vatímetros están conectados a un sistema de adquisición y control, SCADA, el cual lee cada variable. El sistema SCADA, se encuentra conectado al ordenador de control donde quedan registrado los datos.

El consumo eléctrico de los sistemas anteriormente nombrados, se mide mediante dos líneas de consumo. Estas líneas registran por tanto el consumo total de los equipos asociados a ella.

Las dos líneas de consumo existentes son las siguientes:

Línea invernadero. Con los siguientes equipos:

- Motores AC de la ventilación natural
- Máquina deshumificadora
- Sistema de seguimiento y monitorización
- Bomba de calefacción por tuberías
- Aerotermo

Línea Biomasa. Con los siguientes equipos:

- Soplante
- Sistemas auxiliares de la caldera de biomasa
- Sistema de seguimiento y monitorización

En la tabla 2.1 se muestran las potencias nominales de cada equipo:

TABLA 2.1. Consumo de los equipos instalados en el invernadero [15]

Equipo	Potencia nominal (W)
Soplante	1800
Sistemas auxiliares caldera biomasa	210
Sensores y SCADA	140
Aerotermo	1220
Bomba Calefacción	1685
Máquina deshumificadora ¹	8500
Motor ventana cenital	188.5
Dos motores Ventanas laterales	2x84.22
Sensores y SCADA	147
Total línea invernadero	12084
Total línea biomasa	11720

¹Se trata de un consumo promedio. La máquina consume 7 kW para una temperatura de 10°C y 10 kW para 27°C.

2.2. Softwares de desarrollo

Para el desarrollo de este trabajo se han usado una serie de programas informáticos, sin los cuales no habría sido posible manejar un volumen tan grande de datos, representarlos, así como realizar cálculos sobre ellos y como fin último la obtención de los modelos.

Estos programas ofrecen una serie de herramientas, las cuales con el suficiente conocimiento permite la realización de las diferentes tareas que se han llevado a cabo en este TFM.

Los principales programas informáticos usados han sido:

Microsoft Excel [16]

Se trata de la famosa hoja de cálculo del paquete Microsoft office. Aunque los datos vienen en ficheros tipo .txt la hoja de cálculo permite abrir estos ficheros y organizar los datos por columnas eligiendo el tipo de delimitador. Una vez abierto podemos identificar cada columna con una variable, así como realizar representaciones gráficas de la base de datos original sin temor de que hayamos cometido algún error en la exportación, ya que no se ha realizado ninguna.

Matlab [17]

Matlab es un entorno completo de cálculo y programación con su propio lenguaje, llamado *m*. Matlab posee una gran capacidad de cálculo, así como de tratamiento matricial, representación gráfica y simulación. Además tiene diferentes toolboxes que permiten la realización de tareas especializadas en diferentes áreas de conocimientos. En este trabajo se han empleado al menos dos de ellas:

- Import Data app. Este paquete de herramientas ha permitido abrir y leer de forma sencilla los ficheros excel con los datos de las mediciones. Se ha podido seleccionar aquellas columnas que resulta de interés leer, descartando el resto. De esta forma se cargan los datos en Matlab para su posterior procesamiento. Esta aplicación permite generar un código de lectura de forma que tras abrir los datos con la aplicación una vez y generar el código, no es necesario volver a hacerlo y se puede usar directamente el código. Esto facilita y ahorra tiempo debido a la gran cantidad de datos con los que se trabaja.
- System Identification Toolbox [18]. Se trata de un paquete de herramientas destinado a la identificación de sistemas y creación de modelos dinámicos e-s tanto lineales como no lineales. Tiene una sección destinada al pre tratamiento de los datos, con funciones de interpolación y reconstrucción, filtrado, selección de señales y subconjunto de datos. Permite la representación gráfica en el dominio del tiempo y la frecuencia de varias variables a la vez. La sección de identificación permite crear diferentes estructuras, tanto para modelos lineales como no lineales. Además, permite elegir aquella estructura que optimiza la predicción mediante diversos criterios para minimizar el error cometido. Por último, la herramienta permite el ajuste de los parámetros del modelo, y validar los mo-

delos mediante validación cruzada, análisis residual, respuesta del modelo ante diferentes entradas, etc. En la figura 2.8 se muestra una captura de pantalla de esta herramienta.

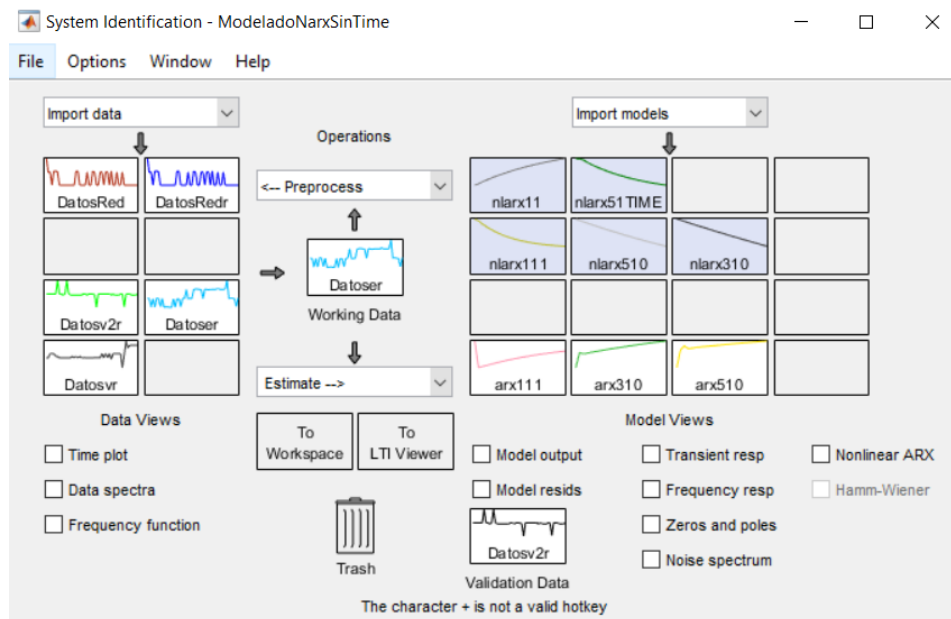


Fig. 2.8. Captura de la interfaz gráfica de la herramienta de identificación de sistemas de Matlab, System Identification Toolbox

También ha sido importante el uso del entorno gráfico de Matlab. Este ha permitido la representación gráfica de las variables de interés, facilitando el análisis. Algunos de los comandos más usados han sido

- *plot(x, y, 'marcador')*: permite la representación gráfica de y en función de x . Si queremos una gráfica de dispersión escogemos algún marcador para que no una la nube de puntos. Aunque para este tipo de representaciones llamados diagramas de dispersión esta el comando específico *scatter*.
- *hold on*: permite dibujar sobre una misma gráfica varias curvas, siempre que compartan el mismo eje x .
- *subplot*: Permite representar diferentes gráficas conjuntamente en un espacio dividido en tantas partes como gráficas.

Otros comandos importantes usados cuya aplicación se detalla más adelante son:

- *filloutliers*: esta función permite obtener los valores atípicos de la serie histórica que serán considerados como falsas medidas. Esta clasificación es subjetiva en cuanto que habrá valores que claramente podrán ser clasificados como falsas medidas, por ser físicamente imposible o diferenciarse mucho del resto, mientras que otros valores son más difícil de

ser encontrados y el éxito de clarificarlos correctamente depende del método y los parámetros usados. Por ejemplo, si nos basamos en la media de los datos para clarificarlos, no será lo mismo usar el periodo completo del cultivo donde la variabilidad será mayor que si usamos una media local calculada sobre una semana o un mes. En nuestro caso se ha usado la opción *movmedian* que clasifica como valores atípicos los valores que se diferencian de la mediana más de tres veces la desviación media absoluta. Para el cálculo de la desviación y la mediana no se usa toda la serie histórica, sino que se le impone una ventana móvil alrededor de los valores estudiados, obteniéndose valores locales. El tamaño de dicha ventana influye considerablemente en los resultados obtenidos, y se ha escogido individualmente para cada variable obteniendo el resultado deseado. Se ha usado el método de la mediana móvil, ya que resulta más robusto que otros métodos como el de la media móvil a la hora de detectar valores atípicos, y ha sido usado en trabajos previos similares [19]. Los valores clasificados como atípicos han sido sustituidos con el valor disponible inmediatamente anterior.

- *fillmissings*: Esta función permite encontrar las discontinuidades en las medidas, y reemplazarlas mediante algún método. En este caso se ha elegido la opción *previous* que completa el la serie histórica con el valor disponible inmediatamente anterior a la discontinuidad.
- *zscore*: permite normalizar los datos, obteniendo para el conjunto una media nula y varianza unidad.
- *corrcoef*: permite obtener los coeficientes de correlación lineal entre dos variables.
- *partialcorri*: permite obtener los coeficientes de correlación parcial de un conjunto de variables.
- *PCA*: permite el análisis de componentes principales.

2.3. Técnicas de análisis de datos

Puesto que vamos a construir modelos predictivos en base a datos reales, es indispensable partir de datos fiables, que no estén contaminados, no contengan falsas medidas y representen de forma segura las variables medidas. Por ello se debe realizar un exhaustivo análisis de los datos con los que se va a trabajar. Aunque no existe un procedimiento completamente automático, ya que cada variable es singular, existen procedimientos y técnicas que nos ayudan a analizar las series históricas [2].

Lo primero es realizar representaciones gráficas de los datos. Estas pueden ser en función del tiempo o de otra variable medida. Esta segunda opción nos permite además identificar posibles relaciones entre variables, como se explica en el apartado 2.4. Como se comento anteriormente Matlab ha sido el principal software empleado para el desarrollo de este trabajo. Matlab posee un potente paquete gráfico que permite controlar amplios parámetros de las gráficas. Estas

gráficas deben ser detenidamente analizadas, intentando detectar cualquier anomalía. Las irregularidades más frecuentes que podemos encontrar son:

- Falsas medidas y valores atípicos: se trata de medidas erróneas que no han ocurrido realmente. Son los llamados Outliers, en inglés. Pueden deberse a la desconexión de un sensor, caída en la tensión de alimentación o a una perturbación que afecte a la medida [7]. Para detectarlos debemos establecer el rango de valores físicamente posible que puede adquirir una variable. Su detección se facilita en la representación gráfica ya que no siguen la distribución característica del resto. Estas falsas medidas deben ser eliminadas para que no afecten negativamente en la construcción del predictivo. Si las eliminamos sin más, estamos creando una discontinuidad en la serie histórica. Existen distintos métodos como la interpolación [20] para sustituir estos valores por otros próximos a la realidad. En este trabajo se ha usado la función filloutliers, explicada anteriormente, sustituyendo los valores atípicos por el valor disponible inmediatamente anterior en la serie histórica.
- Ruido: el ruido es una señal aleatoria que no guarda correlación con el tiempo. Se trata de una distorsión que aparece cuando medimos las señales. Si la amplitud de dicho ruido es significativa sobre la señal que queremos medir, este debe ser eliminado. Este ruido suele estar caracterizado por su frecuencia. En este trabajo por ejemplo, la medición de viento aparece con ruido, y ha sido eliminada mediante la aplicación de un filtro Savitzky-Golay [21]. Este filtro permite suavizar los picos de la serie histórica causados por el ruido, preservando las características de la distribución original de la variable medida. Para ello un filtro tipo Savitzky-Golay aproxima cada punto de la serie histórica, mediante una regresión local con un polinomio de orden k , usando $k+1$ puntos su entorno.
- Discontinuidades: puede haber en las series históricas de las variables, instantes o periodos en los que no se ha medido y no existe registro de esa variable. Esto puede ser debido a la rotura de un sensor o su indisponibilidad temporal. Además como las variables medidas son muy numerosas, difícilmente todos los sensores han registrado durante el mismo tiempo y de manera continua en el tiempo. Al igual que las falsas medidas, se han sustituido los huecos en las mediciones con el valor disponible inmediatamente anterior. La función fillmissings anteriormente explicada ha sido usada para tal fin.

Con los comandos realizados hasta ahora hemos obtenido para cada variable medida, una sucesión de datos limpia y depurada sin valores erróneos ni huecos. De ellos podemos obtener información gráfica y estadística. Sin embargo, para que el pre procesamiento de los datos este completo debe realizarse además una normalización.

Cada variable posee una escala y su variabilidad puede ser muy diferente si las comparamos entre sí. Por ejemplo la radiación solar varía de valores usuales de 500 W/m^2 durante el día a 0 W/m^2 en la noche. Mientras que las variaciones en la temperatura externa es como mucho de decenas de grados centígrados. Esto puede causar problemas numéricos a la hora de estimar los modelos, dificultando su convergencia. Por ello es conveniente normalizar los datos. Cuando normalizamos hacemos que la media de un vector de datos resulte nula y su varianza sea la

unidad. Matlab dispone de la función $zscore(x)$ que devuelve la normalización de la matriz x . no recuerdo que ejecuta una normalización sobre los datos introducidos. Una vez normalizados los datos estamos en disposición de trabajar con ellos y aplicar los métodos que se describen en los próximos apartados.

2.4. Técnicas de selección de variables de entradas

La construcción de modelos predictivos autoregresivos implica estimar una salida concreta del sistema, el consumo eléctrico en este caso, a partir de sus valores pasados y una, ninguna o varias variables de entradas. La elección de las variables a usar como entradas en los modelos predictivos no es trivial. Las variables usadas en los modelos deben aportar información suficiente que permita predecir con el mínimo error las futuras salidas del sistema. Esto podría hacernos pensar en un primer momento en incluir todas las variables disponibles. Sin embargo, esto fácilmente podría llevarnos a usar variables redundantes, que comparten gran parte de la información sin incluir ninguna nueva. Como consecuencia estaríamos sobre parametrizando el modelo, incurriendo en una predicción mas costosa y menos precisa [2].

Por tanto deben elegirse aquellas variables que afecten a la dinámica del proceso que se desea predecir y desechar las variables que no afectan al sistema o aquellas variables redundantes debido a que están fuertemente relacionada con otra variable ya usada. Dentro de cada variable se escogerá un conjunto de datos que aporte la mayor información posible, seleccionando por ejemplo datos de distintos puntos de operación del sistema. La idea es que los datos sean lo menos parecido posible entre sí. También puede darse la situación contraria, dejando fuera variables que influyen considerablemente en la salida. En tal caso, el modelo resultante producirá predicciones pobres [2].

Por ello tras limpiar y filtrar las medidas se procede a una selección de variables, que nos de una idea de que variables usar y cuales desechar. Cuando dos variables esta relacionadas entre sí, se dice que existe colinealidad o que están correlacionadas. Esta correlación puede ser lineal (colinealidad), cuando una variable es combinación lineal de otras. Es decir es proporcional a las otras. También pueden darse correlaciones no lineales, cuando una variable esta relacionada con otra de forma cuadrática, exponencial, logarítmica etc. A continuación se exponen los métodos más empleados, centrándonos en aquellos que han sido usados en este trabajo.

2.4.1. Métodos visuales

Se trata de representar gráficamente una variable en función de otra. Esto permite observar gráficamente relaciones entre variables que no tienen por que ser lineales. La función 'Scatter' de Matlab permite esta representación. Este método, aunque no cuantifica las relaciones, puede ofrecernos información cualitativa de las relaciones de forma rápida, sencilla y bastante intuitiva.

2.4.2. Métodos Basados en modelos

Estos métodos se basan en la creación de modelos predictivos con diferentes entradas y la comparación de sus bondades. Para que se entienda: se crean modelos y se le van añadiendo o eliminando paulatinamente variables de entrada. Se compara si el resultado obtenido mejora o no. Esto implica la construcción de un modelo, para su posterior tratamiento, sin conocer a priori las variables que se han de usar. Esto se presenta como una desventaja, sobretodo en casos donde el número de entradas disponibles es alto. Además este tipo de métodos implican un alto número de ensayos de prueba y error debido al gran número de combinaciones posibles que pueden tomar las entradas.

Por contra, tienen la ventaja de que la eficacia del modelo usado está garantizada pues el criterio de selección se basa en la mejoría del resultado obtenido a través de este.

Estos métodos son llamados heurísticos, y aunque no garantizan el óptimo, obtienen soluciones razonables en tiempos aceptables. Existen diferentes algoritmos de selección y discriminación de variables. Algunos de los más conocidos y usados son:

- Selección hacia delante (o forward selection). Se parte de un modelo inicial con una sola variable. A partir de aquí se usa la correlación parcial para ir añadiendo variables al modelo en cada iteración. Este algoritmo tiene la ventaja de que necesita menos capacidad de cálculo, ya que empieza trabajando con pocas variables. Sin embargo tiene inconvenientes que hacen que su aplicación sea reducida. Por ejemplo no es capaz de eliminar variables que se han vuelto innecesarias tras la inclusión de otra.
- Selección hacia atrás (o backward selection). A diferencia del método anterior, aquí se parte de un modelo con todas las variables. Se usa algún método para excluir una a una variables que no aporten información al modelo. Este modelo tiene la desventaja de que necesita mayor capacidad de cálculo, ya que se parte de un modelo con todas las variables. Esto mismo resulta en una ventaja, al no dejar fuera ninguna variable significativa para explicar la dinámica, lo que resulta muy útil, especialmente cuando las variables no son demasiado numerosas.
- Selección paso a paso (o stepwise selection). Este método es una combinación de los dos anteriores. En cada iteración se añade o elimina una variable. Por lo que existen dos criterios, uno para eliminar una entrada y otro para añadir una entrada. El proceso termina cuando el resultado no mejora significativamente al añadir o eliminar una variable. Este método es el más usado de los tres.

Existen otros algoritmos más sofisticados como los algoritmos genéticos (AG) o los algoritmos meméticos (AM). Los algoritmos genéticos por ejemplo, se basan en la premisa de que diferentes partes de la solución óptima pueden ser descubiertas independientemente y luego ser combinadas para formar mejores soluciones. Bajo esta suposición construyen diferentes modelos que poseen características de la solución óptima. En este trabajo no se han usado métodos basados en modelos, por lo que no se entrará en más detalles. Si se quiere ampliar información,

en [22] y [23] se explica de forma detallada el uso de estos algoritmos, en [2] se presenta una tabla con ejemplos y casos de referencias bibliográfica donde se usan diferentes métodos basados en modelos.

2.4.3. Métodos sin modelos a partir de los datos observados

Existen otras técnicas de selección de variables que estudian, que señales tienen más influencia sobre la variable a predecir a través de los datos que las componen. Para ellos se usan métodos y criterios estadísticos. Esto implica la necesidad de tener un gran número de datos registrados. Además este método exige una gran capacidad de cálculo cuando se trabaja con un elevado número de variables. A su favor, estas técnicas no necesitan la construcción de un modelo a previo. Si el número de variables no es muy elevado, estas técnicas permiten obtener la solución óptima, proporcionando las variables más influyentes a usar, independientemente del modelo que se escoja luego para estimar la salida. Por ello se denominan métodos exactos u óptimos [23]. Las características de este trabajo, en el que se dispone un gran número de datos y las posibles variables a usar no son excesivas, permiten aplicar este tipo de técnicas. Existen en la literatura múltiples métodos de selección de variables para modelos lineales clásicos. Cuando el problema es no lineal, la complejidad aumenta de forma considerable. En este trabajo se han empleado varios métodos que permiten conocer la relación lineal entre variables, así como un método para conocer la información no lineal.

Los primeros son los llamados coeficientes de correlación. El método no lineal empleado es el llamado análisis de componentes principales.

Coefficientes de correlación

El primero de estos coeficientes es el coeficiente de correlación lineal simple, r_{xy} (correlación bivariada simple). Se trata del análisis lineal más básico y sencillo. Este mide como varía de forma lineal una variable cuando lo hace otra [24]. O dicho de otra forma: es un indicador de cuanto se aproxima la relación entre dos variables a una línea recta. Se calcula como: [25]

$$r_{x,y} = \frac{Cov(x,y)}{s_x \cdot s_y} \quad (2.1)$$

Donde s_x y s_y son las desviaciones típicas muestrales de las variables x e y , respectivamente.

Este coeficiente toma valores entre -1 y 1, de forma que:

- Mientras más se acerque a 1 en valor absoluto mayor será la fuerza de la correlación entre los valores de las variables observadas.
- Si es positivo, la correlación es directa. Esto quiere decir que cuando aumentan los valores observados en una variable, aumentan los los valores observados en la otra variable de estudio.

- Si es negativo, la correlación entre los datos observados es inversa. Cuando aumentan los valores de una variable disminuyen los valores de la otra.
- Cuando no existe relación lineal entre variables, el coeficiente de correlación toma valores cercanos a 0.

Este coeficiente nos permitirá por tanto evaluar que relación lineal guarda el consumo medido del invernadero con cada una de las variables disponibles. Nos dará de esta forma una idea de que variables son más influyentes. Sin embargo no nos aporta ninguna información acerca de la colinealidad entre variables. Una variable puede compartir mucha información con la variable a predecir y que sin embargo esa información ya haya sido aportada por otra/as variables más influyente. Para ello habrá que recurrir a otros métodos.

Otro aspecto importante a la hora de calcular este coeficiente es el índice de significancia. Este índice de significancia nos permite rechazar la hipótesis nula, en este caso la hipótesis de que el coeficiente de correlación es nulo y no existe correlación. Si denotamos al índice de significancia como p , entonces nos permiten afirmar que el coeficiente de correlación es diferente de cero con una probabilidad de equivocarnos de p (en tanto por uno) [17].

La función `corrcoef()` de Matlab nos permite obtener la matriz de coeficientes de correlación de varias variables a la vez. También nos permite especificar el índice de significancia que queremos para los resultados obtenidos.

Coefficiente de determinación

El siguiente de estos coeficientes es el coeficiente de determinación. Este coeficiente, denotado como R^2_{xy} , mide que porcentaje de la variabilidad lineal de y es explicada por el regresor x . Por tanto $1 - R^2$ nos da la información que no es común entre ambas. El coeficiente de determinación se calcula a partir del coeficiente de correlación lineal como

$$R^2_{x,y} = r^2_{x,y} \quad (2.2)$$

Correlación parcial

Los dos coeficientes propuestos hasta ahora nos dan idea de la influencia de las variables con la variable a predecir. Sin embargo puede haber problemas de multicolinealidad, es decir dos variables independientes ambas fuertemente relacionadas con la variable dependiente, pero que aportan la misma información. Para detectar esto usamos la correlación parcial. Mediante la matriz de correlación podemos tener una idea de si dos variables comparten información. Ahora bien, no sabemos si la información compartida afecta a nuestro modelo o no. Por ello es importante la correlación parcial, que se define como la correlación entre dos variables si las demás variables no varían, es decir, los valores de las demás variables son fijos [26]. Esto en la práctica permite analizar la correlación entre la variable a predecir, consumo, con una de las variables independientes, controlando el resto de las variables. Con esto nos aseguramos que la

correlación calculada se debe únicamente a variaciones en la variable independiente y no una tercera, siendo esta información lo nuevo que aportaría al modelo. Matemáticamente:

$$\Delta R^2 = R_{y,123}^2 - R_{y,12}^2 = R_{y(3,12)}^2 \quad (2.3)$$

Donde

- ΔR^2 : variabilidad de la salida y explicada exclusivamente por la variable 3.
- $R_{y,123}^2$: variabilidad explicada por las variables 1,2 y 3.
- $R_{y,12}^2$: Variabilidad explicada por las variables 1 y 2,
- $R_{y(3,12)}^2$: correlación parcial entre y y la variable 3.

El comando *partialcorri(Y,X)* de Matlab permite obtener las correlación entre Y y cada componente de X, controlando el resto de componentes contenidas en X.

Componentes principales

Las técnicas descritas hasta ahora permiten conocer la relación lineal entre variables. Existen otras técnicas más potentes, capaces analizar otro tipo de relaciones. Es el caso del análisis de componentes principales, PCA, por sus siglas en ingles.

Este método se basa en el estudio de los datos y sus características, sin prestar atención si están relacionados o no con la variable a predecir. El análisis de componentes principales es una técnica clásica de reducción de dimensión [27]. El objetivo es transformar un conjunto de datos en otro de menor dimensión. Se busca entre todas la variables originales, un subconjunto de variables que represente los datos iniciales con la menor pérdida de información posible. Para ello, el análisis de componentes principales busca unas nuevas variables 'virtuales' ortogonales entre sí en las cuales proyectar los datos. La primera componente está dada por la dirección que maximiza la varianza de los datos proyectados. La segunda componente se busca, siguiendo el mismo criterio, en el espacio ortogonal a la primera componente, y así sucesivamente. Cada una de estas componentes esta formada por componentes lineales de las variables originales con un peso relativo a su participación. Estas nuevas componentes principales no se corresponden con variables físicas reales por lo que su interpretación se complica.

Una vez realizada la descomposición en componentes principales se pueden adoptar diferentes criterios para eliminar variables. En principio, en este trabajo bastará con la representación gráfica sobre los nuevos ejes.

Este método se ve bastante afectado cuando las variables están en diferentes unidades y escalas. Por ello el análisis suele realizarse sobre la matriz de correlación. Matlab posee varios comandos para realizar este análisis. *svd(R)* realiza la descomposición a partir de la matriz de correlación. *PCA()* en cambio realiza la descomposición a partir de los datos originales. En este trabajo se usó el comando PCA con los datos normalizados.

Existen otras técnicas como los falsos vecinos, información mutua o correlación canónica que analiza la relación no lineal entre variables sin necesidad de construir un modelo previo. Si se quiere ampliar información, en [2] se presenta una tabla con ejemplos y casos de referencias bibliográfica donde se usan estos métodos.

2.5. Técnicas de modelado empírico

Las técnicas de modelado empírico, identificación o regresión, permiten el modelado de sistemas a partir de datos reales de entrada y salida. Por ello también se conocen como modelos e-s. A diferencia de los modelos basados en primeros principios, el modelado empírico permite predecir y modelar sistemas sin conocer las leyes físicas que rigen dichos sistemas [2]. Para predecir las salidas del sistema se usan datos de entrada de variables que afectan al sistema y los valores pasados de la propia salida que se quiere predecir, por lo que reciben el nombre de regresivos. Si bien estos modelos presentan ciertas desventajas como la pérdida de parte de la dinámica del invernadero o la imposibilidad de conocer y relacionar las ecuaciones que gobiernan el invernadero, presentan una serie de ventajas, entre las que destaca la relativa simplicidad de obtención del modelo.

Para obtener dicho modelo, se parte de la idea de que los sistemas pueden ser modelados mediante una familia de estructuras, seleccionando aquel modelo de la familia más adecuado para aproximar los datos observados. Una vez seleccionada la estructura del modelo se procede a calibrarlo mediante el ajuste de sus parámetros.

Las familias de estructura las podemos clasificar en dos grandes grupos. Modelos lineales y modelos no lineales.

Los modelos lineales fueron los primeros en desarrollarse y usarse debido a su sencillez matemática y a que los resultados eran aceptables. Los primeros análisis estadísticos de series históricas fueron desarrollados por Yule [28] aplicándolos a la predicción de manchas solares. Más tarde, en 1970 Box y Jenkins publicaron su trabajo 'Series Analysis: Forecasting and Control' [29] que supuso un importante hito en el análisis y predicción de series históricas. En su trabajo se presentaba un método de aproximación que permitía la aplicación de la teoría de análisis de series de forma sistemática en aplicaciones de predicción. Esto popularizó los modelos ARIMA.

Las principales familias de modelos lineales se resumen a continuación, y de forma general son estructuras formadas por dos funciones de transferencias una dedicada a las entradas del sistema y otras a las perturbaciones o ruido aleatorio. Dichas funciones de transferencias están formadas por polinomios, en los que cada término representa un valor pasado de una componente que afecta al sistema y los coeficientes permiten el ajuste del modelo.

- AR: modelo autoregresivo. El valor de la salida es función de sus valores pasados y de un término de ruido blanco, que supone error en la predicción. Es el modelo lineal más simple y general que cuenta con un único polinomio que modela la relación entre la salida

y sus valores pasados.

- **Arx:** Modelo autoregresivo con entradas exógenas. Partiendo del modelo anterior, se le añaden valores pasados de entradas externas que afectan al modelo, lo que permite una mejor predicción. Por tanto cuenta con un polinomio para la relación entre la salida con sus valores pasados, y un polinomio más por cada entrada exógenas que se relacione con la salida
- **ARMAX:** modelos autorregresivos de media móvil. Al igual que las dos estructuras anteriores la parte regresiva la componen el polinomio para los valores pasados de la propia salida del sistema y polinomios para valores pasados de entradas exógenas que afectan al sistema. Sin embargo el termino del error deja de ser un único termino, para convertirse en un polinomio que estima el error cometido en la predicción en función de los valores pasados del error.
- **Estructura Box-Jenkins (BJ):** es la estructura más completa ya que la función de transferencia que relaciona el ruido con la salida, posee dos polinomios independientes uno para el numerador y otro para el denominador.
- **FIR:** modelos de Respuesta Finita. No tienen en cuenta valores pasados de la salida a estimar. La estructura esta formada por un termino del error y los polinomios de valores pasados de entradas exógenas.

Estos modelos han sido ampliamente usados en los campos de economía, salud y ciencia sociales. Si bien estos modelos son sencillos de obtener y presentan buenos resultados tienen muchas limitaciones cuando los sistemas son complejos, con ecuaciones no lineales acopladas. Además son modelos que necesitan reajustarse si las condiciones del sistema cambian. En este contexto las Redes Neuronales Artificiales adquieren especial atención para generar modelos autoregresivos no lineales.

Estos tipos de modelos, además de ser muy flexibles pudiendo modelar sistemas altamente complejos y no lineales, poseen cierta capacidad de aprendizaje y adaptación lo que se conoce como Inteligencia Artificial (IA). Esto les permite tanto cambiar la función que realiza la RNA como adaptarse a nuevas condiciones del sistema. Las RNA imitan la forma de aprendizaje y de conexión de las neuronas para describir las relaciones entre las variables que influyen en un sistema. De esta forma el desarrollo de las RNA va ligado al estudio del cerebro y su actividad.

Los primeros fundamentos se describen en [30] donde Warren McCulloch, en 1943 proponen una teoría acerca de como funcionan las neuronas y modelan una red neuronal simple con un circuito eléctrico. Más tarde en 1949 se presenta un importante trabajo acerca del aprendizaje en las redes neuronales [31]. Este es el fundamento de la mayoría de funciones de aprendizaje que existen para las RNA.

En 1958 Frank Rosenblatt describe y desarrolla el Perceptrón, que constituye la primera RNA. Aunque tiene sus limitaciones, es usada actualmente y tiene capacidad de generalización.

En 1974 Paul werbos [32] desarrolla el algoritmo de aprendizaje 'backpropagation'. Aunque desde 1960 se usaron las RNA en problemas reales, no fue hasta 1986 cuando David Rumelhart y G. Hinton [33] redescubrieron el algoritmo de aprendizaje 'backpropagation' cuando se extendió y se dio real importancia a las RNA. Desde entonces las RNA han sido ampliamente estudiadas y aplicadas sobre todo en el ámbito de control automático

En cuanto a la aplicación de modelos predictivos en invernaderos, estos modelos se ha usado para predecir el comportamiento dinámico de la temperatura del aire en el interior así como la humedad. Con esto se pretendía ayudar a los productores a mejorar las condiciones del cultivo. Los primeros modelos estaban basados en primeros principios [34], [35] [36] . describiendo los balances de masa y energía. Estos modelos tomaban mucho tiempo en generarse debido a la necesidad de calibrar el modelo con el sistema real mediante la realización de ensayos. La identificación de sistemas facilito esta labor al permitir obtener el modelo ajustado mediante los datos de e-s. Algunos de los trabajos mas importantes realizados con sistemas lineales se encuentran en [37] y [38], [39].

En [37] usaron modelos Arx para predecir la temperatura interior de un invernadero. En su estudio analizaron la posibilidad de usar un mismo modelo para las cuatro estaciones del año. Como conclusión obtuvieron que el modelo debía ser reajustado para obtener una predicción aceptable durante todo el año. En [6] se presenta una tabla con diferentes ejemplos de aplicación y características de modelos lineales aplicados a invernaderos.

Las predicción mediante RNA también se aplica a los invernaderos mejorando los resultados obtenidos con modelos lineales. Aunque un invernadero tiene ciertas componentes cuasi-lineales [7] y puede ser modelado linealmente entorno a un punto de trabajo [40] los modelos no lineales permiten describir mejor las diferentes ecuaciones no lineales que afectan al clima del invernadero. El invernadero es un sistema MIMO, con ecuaciones no lineales acopladas entre sí. Las RNA permiten además la generalización del modelo . En 2016 Taki [41] compara la predicción obtenida mediante una RNA perceptrón con modelos lineales multivariantes para predecir la transferencia de energía y la temperatura interior. Temperatura exterior, velocidad del viento, humedad interior, radiación, temperatura de suelo son tenidas en cuenta como entradas. Las RNA mostraron mejores resultados que los modelos lineales

En 2017 Castañeda-Miranda y Castaño [42] también usaron un RNA multi perceptrón con las entradas temperatura ambiente, radiación, humedad interior y exterior y velocidad del viento para predecir la temperatura interior bajo condiciones de de frío y congelación.

El impacto de la fertirrigación es tenido en cuenta en el clima interior añadiendo la temperatura de suelo como entrada en [43]. En [44] se modela tanto la temperatura como la humedad interior para estudiar el fuerte acoplamiento que existe entre ambas.

Otros autores como [6] o [45] también han usado RNA en predicción de invernaderos. En [6] se presenta una tabla con diferentes ejemplos de aplicación y características de RNA aplicados a invernaderos.

2.5.1. modelos lineales

Como se ha descrito anteriormente la regresión lineal pretende describir las relaciones entre una variable dependiente, y un conjunto de variable independientes. Para poder describir estas relaciones matemáticamente la regresión lineal toma las siguientes hipótesis:

- Linealidad. La relación entre la variable dependiente y cada una de las variables independiente es lineal.
- Homocedasticidad: la dispersión o varianza de los datos debe ser constante. Proceso estacionario

Las siguientes hipótesis son referidas a los residuos, es decir a la diferencia entre la predicción y el valor real de la salida.

- Normalidad: los residuos deben seguir una distribución normal.
- Independencia. Los errores no están correlacionados con sus valores pasados ni con las variables de entrada.
- Homogeneidad. El valor promedio del error es cero

Asumiendo estas hipótesis se pueden aplicar los modelos de regresión lineales. En este apartado se va a describir en detalle las características de los modelos Arx, que es el modelo lineal empleado en este trabajo. Se comienza describiendo los modelos AR, ya que los modelos ARX son una extensión de estos considerando perturbaciones.

La familia de modelos lineales parametricos AR describen el valor de la salida a predecir a partir de sus valores pasados y un ruido que representa una señal no manipulable y de comportamiento impredecible que afecta al sistema [2]. Matemáticamente este ruido se define como un ruido blanco con media nula. Entonces matemáticamente un modelo AR se puede escribir como:

$$y(t) = a_1y(t - 1) + a_2y(t - 2) + \dots + a_ny(t - na) + v(t) \quad (2.4)$$

Donde $y(t)$ es la predicción de la salida en el instante actual t , na es el número de valores pasados de la salida usados el cual define el orden del modelo, a_n son los coeficientes estimados del modelo para cada valor pasado de y , e v la componente impredecible de ruido blanco y que supone un error en la predicción. Por ejemplo un modelo AR de orden dos quedaría definido como:

$$y(t) = a_1y(t - 1) + a_2y(t - 2) + v(t) \quad (2.5)$$

Siendo a_1 y a_2 los coeficientes a estimar a partir de los datos reales del sistema.

Sin embargo estos modelos no tienen en cuenta posibles entradas al sistema, como perturbaciones, que pueden afectar a la salida del sistema y que están presentes en la mayoría de procesos físicos. De estas perturbaciones se puede tener información si las medimos y se puede tener cierto conocimiento de como afecta a la dinámica ayudándonos a modelar el sistema. Por ello surgen los modelos ARX, donde partiendo de los modelos AR se añade una variable exógena, de ahí su nombre, que añade información de como afectan una o varias entradas al sistema. La estructura de un modelo ARX se expresa como:

$$y(t) = a_1y(t-1)+a_2y(t-2)+\dots+a_nay(t-na)+b_1u(t-nk)+b_2u(t-nk-1)+\dots+b_nbu(t-nk-nb+1)+v(t) \quad (2.6)$$

Donde a la estructura del modelo AR, se le ha añadido la perturbación u , con un coeficiente b para cada valor pasado de dicha perturbación. nb indica el orden del regresor, es decir el número de valores pasados que se usan de la señal u . Por último nk indica el retardo de tiempo que sufre la señal de salida en verse afectada ante una variación de la entrada u , y se mide contando el número de muestras o "samples" que ocurren desde que se sufre la perturbación hasta que se afecta la salida. Un ejemplo concreto donde $na=2$, $nb=1$ y $nk=1$, se definiría como un modelo ARX(2,1,1) y tomaría la forma:

$$y(t) = a_1y(t-1) + a_2y(t-2) + b_1u(t-1) + b_2u(t-2) + v(t) \quad (2.7)$$

Esta estructura admite tantas variables exógenas como se quieran incluir, aplicando el principio de superposición. Sin embargo esto incrementa la complejidad del modelo y es importante determinar el orden adecuado de los regresores de cada variable. El coste computacional va incrementando conforme lo hacen el orden de los regresores. Por ello es necesario un método que nos permita establecer que estructura resulta la más óptima, teniendo en cuenta tanto la precisión de la predicción como el coste computacional de la misma. Antes de abordar los detalles del método, resulta adecuado analizar nuestro objetivo. Si estamos ajustando un modelo a una serie de datos de estimación, el error en la estimación se verá reducido cuanto más altos sean los ordenes de los regresores que componen el modelo. Esto se debe a que estamos dotando al polinomio de un mayor orden y por tanto flexibilidad, ajustándose así mejor a los datos. Sin embargo podríamos estar incurriendo en un error de sobreparametrización. Esto ocurre debido a que se estaría modelando esas mediciones concretas, de forma local, con sus particularidades y el ruido superpuesto en las medidas, pero no estaríamos captando la dinámica real del sistema subyacente en los datos [2]. Surge entonces la necesidad de criterios que penalicen la complejidad de los modelos. De forma general esto puede describirse como:

$$criterio(a) = f(error\ del\ modelo) + ag(complejidad) \quad (2.8)$$

El objetivo es minimizar el valor del criterio usado. El compromiso entre complejidad y el error cometido se establece a través del factor de ponderación a . En este trabajo se han empleado dos criterios ambos debido a Akaike [46].

- Criterio del error final (FPE)
- Criterio de información de información de Akaike (AIC)

Matlab tiene implementados el criterio del error final en la toolbox de identificación y por tanto se escogerán aquellos modelos que minimicen tanto el valor de FPE. Estos criterios han sido desarrollados para modelos lineales. También cabe destacar que estos criterios no necesitan un conjunto de datos diferente del conjunto de estimación para determinar los ordenes más adecuados. Si se dispone de un conjunto de datos diferente al usado en la estimación del modelo, se puede recurrir a la validación cruzada.

La validación cruzada consiste en reservar un conjunto de datos diferente al usado en la estimación del modelo, e intentar reproducirlos con los modelos estimados. Los errores cometidos en la estimación nos permitirán comparar los modelos. Puesto que se dispone de una cantidad suficiente de datos, la validación cruzada también se empleara como criterio a la hora de seleccionar los modelos.

Los tres métodos comentados hasta ahora nos proporcionan información acerca de que modelo reproduce mejor el sistema real a predecir. Sin embargo en términos absolutos no nos proporcionan información acerca de la validez de los modelos o de veracidad de los supuestos iniciales. Para esto se recurre al análisis residual que se explica en la sección [2.6](#)

Una vez seleccionada la estructura mas idónea solo queda estimar los coeficientes de los polinomios que constituyen el modelo. En este trabajo se ha usado el método de los mínimos cuadrados disponible en la identification toolbox de Matlab. De forma descriptiva, se trata de un método directo que selecciona aquellos coeficientes que minimizan la suma de los cuadrados de la diferencia entre la predicción y la salida real.

2.5.2. Modelado no lineal

A parte de modelos lineales, en este trabajo se han estudiado modelos no lineales. Concretamente las redes neuronales artificiales (RNA), debido a su gran precisión aproximando y modelando sistemas altamente no lineales así como su capacidad para generalizar los resultados [\[47\]](#).

Las RNA se inspiran y copian el funcionamiento de las neuronas del cerebro humano. Como concepto clave hay que entender la conexión entre neuronas. Cada neurona esta conectada con otras pero hay conexiones más fuertes que otras. Las diferentes formas de conectarse las neuronas entre sí, formando diferentes estructuras, permite procesar señales externas, acumular conocimiento o el aprendizaje. Esta propiedad de las redes neuronales es copiada e implementada computacionalmente. Las neuronas adoptan ahora el nombre de nodos y las diferentes conexiones entre neuronas se ajustan mediante pesos relativos, dotando de más importancia a unas señales que a otras. De esta forma, ajustando los pesos relativos, que supone el entrenamiento de la red, se puede producir la salida deseada del sistema. Para ello, la red neuronal actúa como un modelo de caja negra, relacionando la información de entrada con la salida, lo que le

permite ir ajustando los parámetros de su estructura. La red acumula el conocimiento en el ajuste de los pesos relativos. Cuando existe una red de neuronas con los nodos suficiente, se pueden modelar sistemas muy complejos con muchos parámetros interrelacionados. Pero el principio de funcionamiento es el mismo que una red neuronal pequeña. Al final todas las estructuras se encuentran formadas por unidades muy simples (nodos) muy densamente conectados, generalmente en paralelo. En la figura 2.9 se muestra la estructura de una RNA simple y estándar para ver como funciona.

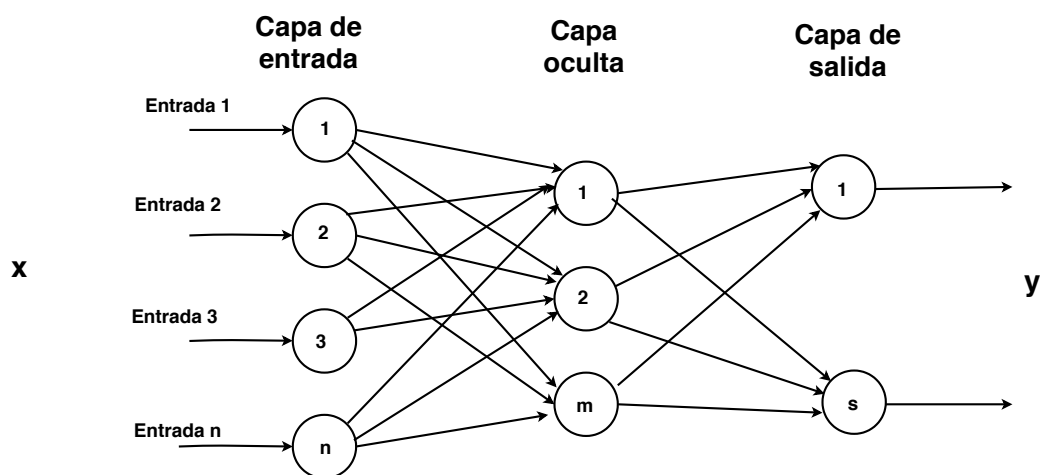


Fig. 2.9. Red neuronal artificial estándar. Fuente: elaboración propia

En la figura 2.9 se diferencian las siguientes partes:

- Capa de entrada: se encarga de distribuir las entradas del sistema entre los nodos de la primera capa oculta.
- Capa de salida: se encarga de obtener cada una de las componentes del vector de salida y . Puede haber una más componentes de salida. Esta capa puede realizar operaciones matemáticas dando más peso a la señal procedente de una neurona que a las de otra. Normalmente suele ser lineal, de forma que solo se suman las señales que llegan a la capa de salida.
- Capa oculta o interna: es llamada así debido a que no es de entrada ni de salida. En este caso solo se muestra una capa con m nodos o neuronas, pero puede haber más de una capa oculta. Estas capas están formadas por las llamadas neuronas, encargadas del aprendizaje.

En la figura 2.9 se ha mostrado una red simple. Esta red puede aumentar su tamaño añadiéndole mas capas internas o más neuronas. Sin embargo el principio fundamental no se ve alterado: están formadas por neuronas, que realizan una operación muy simple y transmiten esa información a otra neurona. En la figura 2.10 se muestra el esquema de una neurona no lineal, que se compone de la siguientes partes:

- Las señales de entrada, las cuales son ponderadas mediante los pesos relativos o coeficientes ajustables, denotados por w_k . Estas señales de entradas proceden de una capa anterior, ya sea la capa de entrada o de otra capa oculta anterior. En redes con conexiones laterales estas señales pueden provenir también de neuronas de la misma capa.
- Un sumatorio donde se suman todas la señales de entrada ponderadas.
- Una función de activación, denotada por σ , la cual se trata de una aplicación no lineal que hace corresponder una salida en función del rango del dominio al que pertenecen las señales de entradas.
- Una única salida, que se conecta con las neuronas de la siguiente capa oculta, o con la capa de salida.

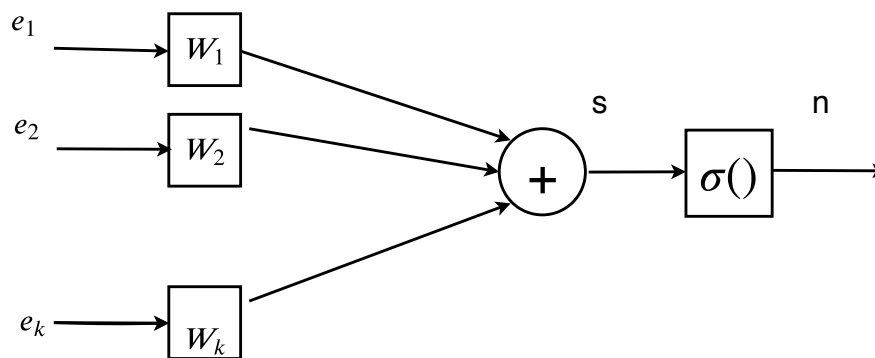


Fig. 2.10. Modelo de una neurona no lineal. Fuente: elaboración propia

De forma matemática la salida de una neurona puede expresarse como:

$$y_n = \sigma\left(\sum(\omega_k * e_k)\right) \quad (2.9)$$

Como se observa, la estructura y operaciones que realiza una neurona no son excesivamente complejas. Sin embargo la adición de estas unidades pueden crear estructuras altamente complejas. Existen RNA ,que cumpliendo ciertos requisitos, son predictivos globales, teniendo la capacidad de modelar cualquier función acotada y derivable [48].

En la red neuronal de la figura 2.9, cuando las neuronas toman una función sigmoideal para la función de activación, se corresponde con una estructura neuronal concreta llamada perceptrón [49]. Se trata de una red neuronal, con una o múltiples capas ocultas, sin conexiones laterales y con una función lineal en la capa de salida. Esta estructura tiene la capacidad de ser un predictor global [50]

A parte de esta estructura, existen múltiples configuraciones que permiten crear diferentes tipos de redes, cada una con unas características propias. En las redes aquí consideradas no se permite la conexión lateral entre nodos. O lo que es lo mismo los nodos de una misma capa no pueden conectarse, y las señales fluyen siempre de izquierda a derecha. Estas RNA reciben el

nombre de directas o *feedforward* en inglés. Por contra, las redes con conexiones laterales son llamadas competitivas y aquí no se entrara en detalle de su funcionamiento.

Puesto que el perceptrón con una sola capa, función de activación hiperbólica y capa de salida lineal es un aproximador global se va a usar este modelo

Esta estructura ha sido usada en la predicción en invernaderos por otros autores [2], [7], [51], [45] y predicción de demanda energética [52] [42].

En otros trabajos para predicción climática en invernaderos se han usado redes neuronales recurrentes con dynamic back propagation(algoritmo [53], en [54] modelo radial

Una parte muy importante de la red, la constituye la función de activación. Esta se encarga de transformar las señales de entrada en señales de salida, más atenuadas o pronunciadas, dependiendo del rango en el que se encuentren las entradas. Existen diferentes funciones, cuya elección determina en gran medida las características de la red, como su rapidez de aprendizaje, capacidad de generalización, cantidad de datos a usar para entrenamiento o tamaño de la red. Tal es la importancia, que dicha función determina la familia y el nombre de la RNA Existen muchos tipos de funciones de activación, de tipo lineal, escalón, pero son dos las funciones más usadas: la función sigmoideal y la función de funciones base.

- Función sigmoideal. Necesitan un elevado número de datos de entrenamiento y resulta más lento. Como bondad presentan un tamaño reducido y la capacidad de generalización. La salida se forma usando funciones de activación, cada cual afecta a todo el dominio de entrada. Por tanto una modificación en una neurona afecta a todo el modelo.
- Función de funciones base: Esta función de activación necesita menos datos para el entrenamiento de la red. Y son más rápidas, haciéndolas adecuadas para la predicción en linea. Sin embargo, cada función modela una parte del comportamiento de la salida, siendo no nula la salida solo en un pequeño rango del dominio de las señales de entrada. Esto le da al modelado carácter local, que impide la generalización. Estas redes son mucho más rápidas ya que pueden trabajar en paralelo, modelando por separado los datos de entrada. Por contra el tamaño de la red puede ser muy elevado, como consecuencia de lo anterior dicho. Las redes que usan esta función adquieren el nombre de Redes de base radial (RBR) debido a la simetría de la función de activación respecto a su centro. Son redes directas y con una sola capa oculta. Estas redes también son aproximadores universales

La red neuronal, entendida como estructura, queda determinada cuando se definen el número de capas y de nodos, sus conexiones, así como la función de activación. Tras esto es necesario ajustar los parámetros de la estructura para reproducir la dinámica del sistema. Para ello se usa un conjunto de datos de entrenamiento, que son datos reales del sistema. A partir de aquí existen dos tipos de procedimientos de entrenamiento:

- Supervisado: existe un conjunto de datos de entrenamiento, compuesto por las entradas reales del sistema y la/as salida/as reales que este proporciona. Esto se conoce como par

e-s. La red neuronal se alimenta con las entradas reales del sistema, y predice una salida del sistema. Esta salida se compara con la salida real que ha sido medida del sistema, y se calcula el error. A partir de aquí existen diferentes reglas para ajustar los parámetros de la red con el objetivo de disminuir el error con cada par e-s usado. El proceso acaba cuando el error es aceptable. Puede requerir realizar el proceso más de una vez. La regla para ajustar los pesos puede ser cualquiera que disminuya el error. El más usado es el conocido como *Backpropagation*: se calcula el error cometido mediante una función. Se calcula el gradiente y se reajustan los parámetros en la dirección negativa del gradiente de la función de error.

- No supervisado. Son los llamados algoritmos de agrupamiento. En este caso solo se cuenta con entradas reales del sistema, pero no las salidas que este produce. El ajuste de la red neuronal, se hace en base a la capacidad de esta de agrupar los datos de entrada entorno a centros diferenciados. La red se va ajustando para diferenciar patrones en los datos de entrada, agrupándolos en tantos centros como nodos tiene la red.

Puesto que en este trabajo se disponen de pares de datos e-s, el entrenamiento realizado es supervisado. -

2.6. Validación de modelos

A parte de los criterios de la información que ya nos ayudaron a seleccionar la estructura más óptima y estimar el error cometido en la predicción, se debe usar un método para aceptar el modelo como válido para el propósito especificado, así como aceptar los supuestos asumidos para la construcción de los modelos. En este trabajo se han empleado dos métodos: la validación cruzada y el análisis de los residuos.

Análisis residual

El análisis de los residuos permite conocer si hay diferencias sistemáticas o estructurales entre el modelo y el sistema real [2]. Los residuos son la diferencia entre la salida real del sistema y la predicción del modelo. Por tanto mide la parte de los datos no explicados por el modelo. Idealmente los residuos deberían aproximarse a un ruido blanco independiente de las entradas del modelo, de forma que se cumpla los supuestos tomados para la regresión lineal. Para verificar esto se realizan dos pruebas, que permiten validar el modelo con un intervalo de confianza. [18]. Por un lado la función de autocorrelación debe mostrar que los residuos no están correlacionados con la salida del sistema. Por otra parte la correlación cruzada debe mostrar que los residuos no dependen de valores pasados de cada una de las entradas usadas en el modelo [17].

De forma concreta el análisis permite: Detectar:

- Si efectivamente la relación entre x e y es lineal.

- Si hay normalidad de los errores
- Si hay valores anormales en la distribución de errores (outliers)
- Si hay varianza constante (propiedad de homocedasticidad)
- Si hay independencia de los errores

Validación cruzada

Como se ha comentado anteriormente la validación cruzada consiste en usar un conjunto de datos diferentes al de estimación del modelo y calcular el error cometido en la predicción. Esto supone una generalización del modelo, que es capaz de obtener buenas estimaciones para conjuntos de datos más allá de los datos usados en la estimación. Este método presenta la ventaja de que es rápido y de fácil aplicación. Sin embargo no nos da información acerca de las carencias concretas del modelo o como podemos mejorarlo.

3. RESULTADOS

A continuación, se van a mostrar los resultados de aplicar los métodos descritos en el capítulo anterior al registro de datos objeto de estudio en este trabajo. Como se comentaba en el apartado 2.1 los datos corresponden a la campaña de 2015. Sin embargo durante los meses de Junio y Julio no hay actividad en el invernadero, puesto que las elevadas temperaturas no permiten el desarrollo de las plantas. De esta forma, el cultivo se planta en Agosto y se recolecta en Mayo. Por ello los datos que se van a mostrar cubren el año 2015 exceptuando los meses Junio y Julio y pertenecen a dos campañas de cultivo diferentes.

3.1. Validación de los datos

En este primer apartado se van a mostrar los datos brutos obtenidos mediante el sistema de adquisición de datos sin ningún tipo de tratamiento. Así mismo se mostrara el resultado de aplicar las técnicas necesarias para depurar y recuperar la información necesaria de los datos brutos para poder trabajar con ellos. Los equipos de medida del invernadero registran una gran cantidad de variables cuyos datos están disponibles con una frecuencia de 30 segundos. Sin embargo, de entre todas ellas se ha preseleccionado un conjunto formado por aquellas variables que ejercen una mayor influencia sobre el consumo del invernadero [45]. Estas se muestran en la tabla 3.1

TABLA 3.1. Conjunto de variables seleccionadas para el estudio

Variable	Tipodevariable	Rango de medida	Unidades
Consumo	Variable dependiente	[0-12]	<i>kW</i>
Radiación	Perturbación	[0-1367]	<i>W/m²</i>
Temperatura exterior	Perturbación	[0,50]	<i>°C</i>
Temperatura interior	Perturbación	[0,60]	<i>°C</i>
Temperatura suelo	Perturbación	[0,60]	<i>°C</i>
Humedad interior	Perturbación	[0-100]	%
Humedad exterior	Perturbación	[0-100]	%
Velocidad del viento	Perturbación	[0,30]	<i>m/s</i>
Estado bomba calefacción	Controlada	{0, 1}	lógica
Estado máquina deshumificadora	Controlada	{0, 1}	lógica
Estado aerotermo	Controlada	{0, 1}	lógica
Apertura ventilación cenital	Controlada	[0-100]	%
Apertura ventilación lateral	Controlada	[0-100]	%

A continuación se analiza en detalle la serie de datos registrada por cada una de estas variables.

Radiación

En la figura 3.1 se muestra la señal original obtenida mediante el piranómetro así como el resultado final obtenido tras la aplicación de las técnicas que se consideraron oportunas.

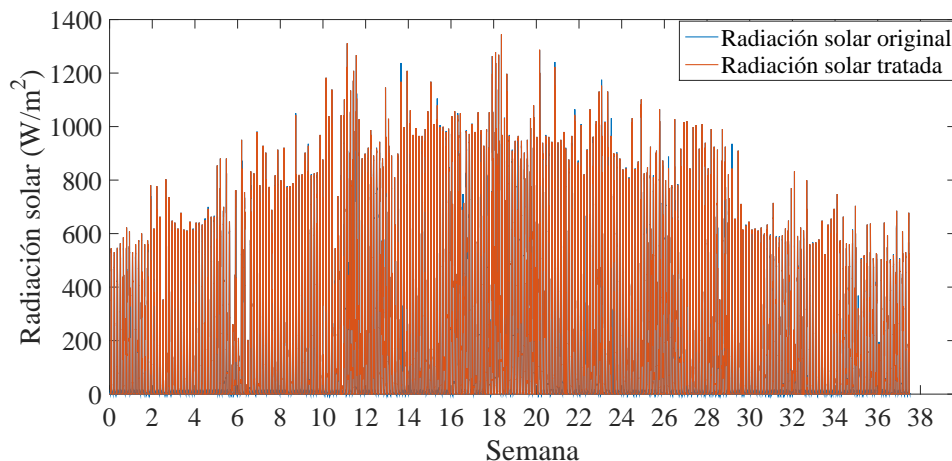


Fig. 3.1. Señal de radiación global horizontal (W/m^2) original y tratada durante el periodo de registro

Se observa en la figura como no existen valores físicamente imposibles en la señal original, es decir menores que 0 o mayores que la constante solar (1367 W). Sin embargo si existen discontinuidades, que han sido detectadas y suplidas con la función *fillmissing* de Matlab, usando los valores disponibles inmediatamente anteriores.

Aunque no existen valores físicamente imposibles si existen falsas medidas provocadas por un mal funcionamiento del sensor.

Estas han sido detectadas mediante la función *filloutliers* de Matlab. Aplicando una ventana de tiempo de 5 puntos (2.5 min). Se trata de una ventana muy pequeña pero es necesaria, ya que la radiación varía rápidamente y no queremos perder esta información.

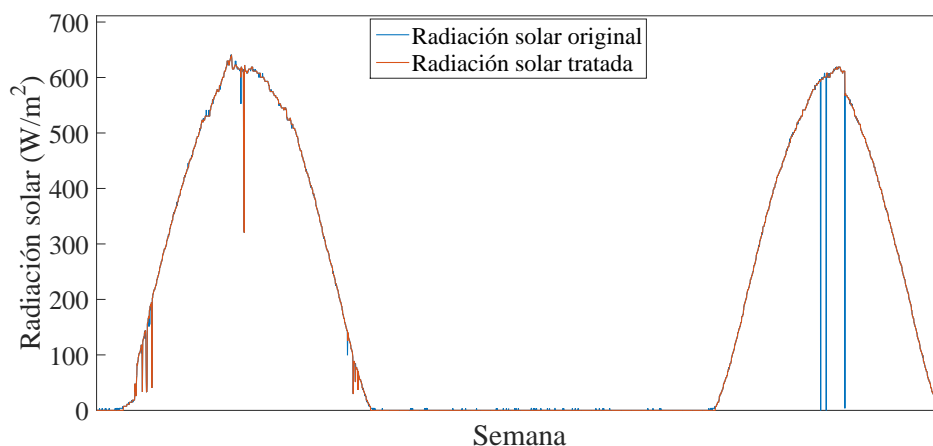


Fig. 3.2. Detalle de la señal de radiación global horizontal (W/m^2) original y tratada

En las figuras 3.2 y 3.3 se muestran ejemplos de los resultados obtenidos

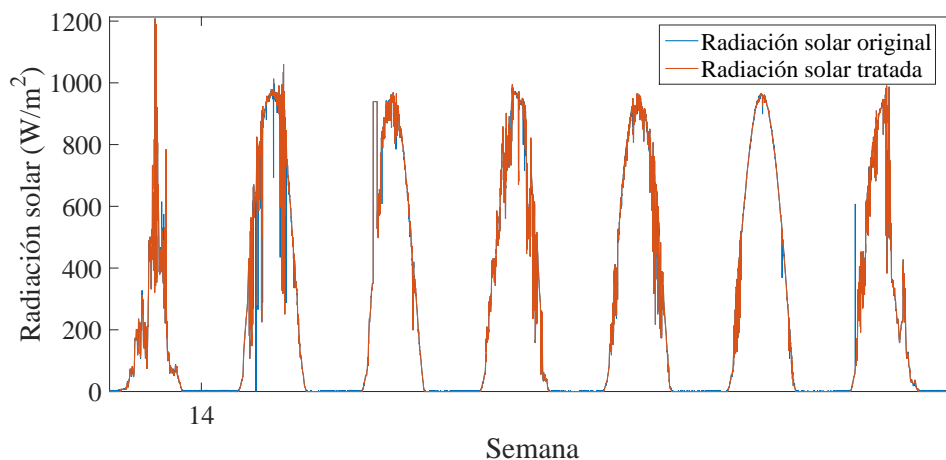


Fig. 3.3. Detalle 2 de la señal de radiación global horizontal (W/m^2) original y tratada

Se observa como se consigue limpiar gran parte de falsas medidas respetando la variabilidad de la radiación.

Viento

La medida de viento presenta por un lado discontinuidades y falsas medidas. Además de bastante ruido. Este ruido se ha eliminado mediante un filtro Savitzky-Golay [21], [20]. Como se explicó en el capítulo anterior, en resumidas cuentas este filtro realiza una interpolación local. Por ello sirve también para reconstruir las discontinuidades y suavizar falsas medidas.

En las figuras 3.4 y 3.5 se muestran los resultados obtenidos así como la señal original.

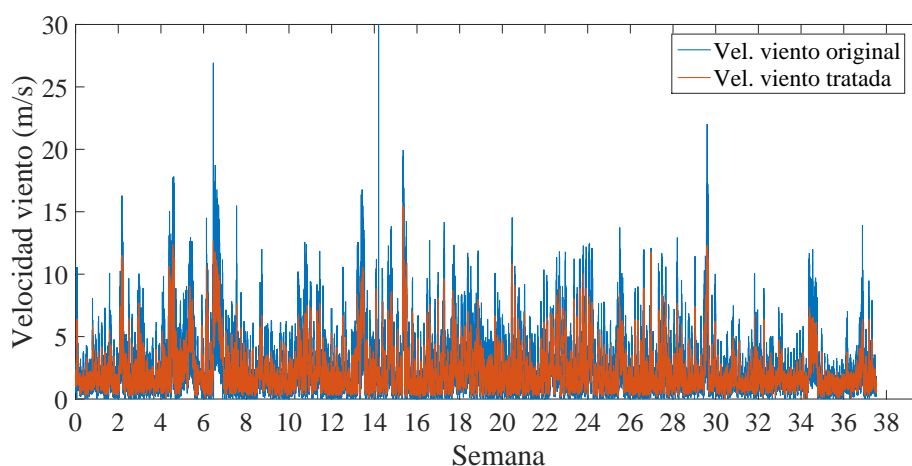


Fig. 3.4. Señal original y tratada de la velocidad del viento (m/s) durante el periodo de registro

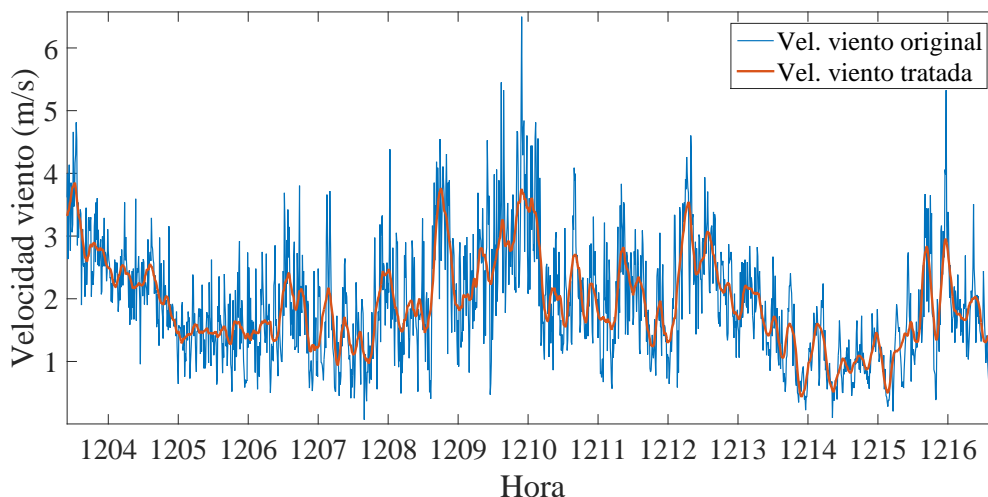


Fig. 3.5. Detalle de la señal original y tratada de la velocidad del viento (m/s)

Se observa como el filtrado suaviza la señal original, captando su tendencia y variabilidad pero eliminando el ruido y las falsas medidas.

Temperatura exterior

Tras el análisis de los datos brutos, se considera que no es necesario un filtrado de los datos. Sin embargo si se detectan falsas medidas, que producen escalones más o menos pronunciados en la serie histórica. También se han detectado discontinuidades en las medidas. Se usa la función *fillmissing* para las discontinuidades y la función *filloutliers* para las falsas medidas, con una ventana móvil de 40 puntos lo que equivale a 20 minutos.

En la figura 3.6 se muestra la señal original y tratada para todo el periodo de registro. Se observan valores físicamente imposibles en la señal original y como son corregidos en la señal tratada

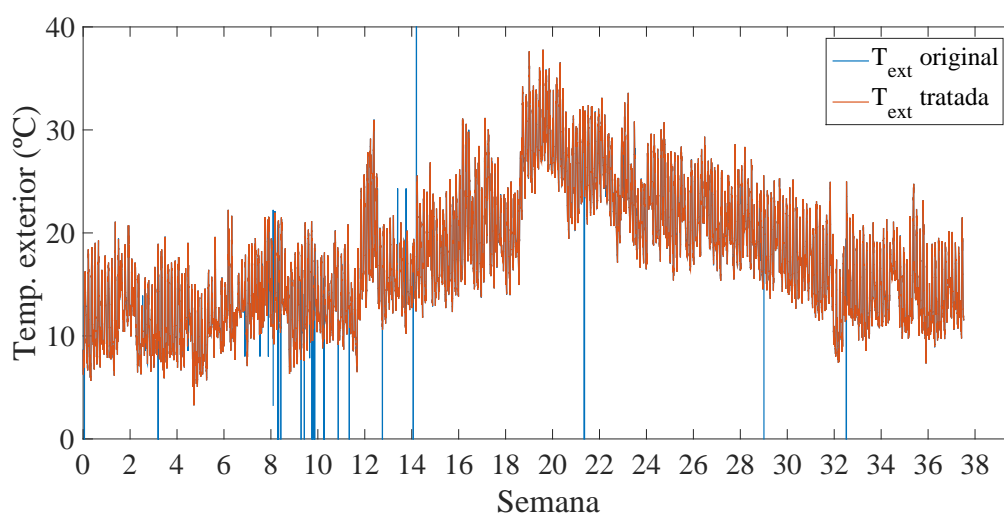


Fig. 3.6. Señal original y tratada de la temperatura exterior (°C) durante el periodo de registro

En la figura 3.7 se muestra un ejemplo de como se corrigen medidas que sin ser físicamente imposibles son falsas medidas.

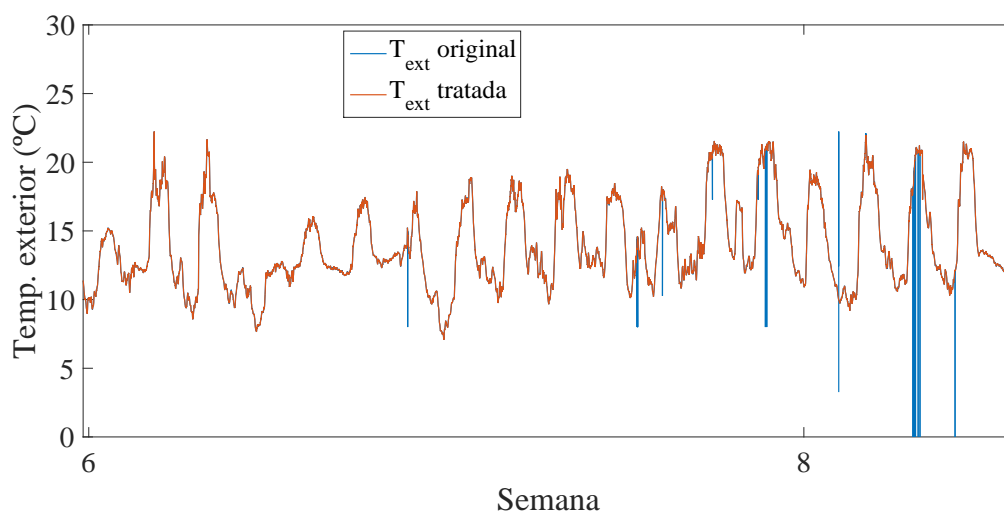


Fig. 3.7. Detalle de la señal original y tratada de la temperatura exterior (°C)

Se ha obtenido pues una señal más limpia de valores imposibles y atípicos.

Temperatura interior

Al igual que con la temperatura exterior se usan las funciones *fillmissing* y *filloutliers*. Sin embargo se ha usado una ventana móvil de 10 puntos lo que equivale a 5 minutos, ya que los resultados obtenidos son mejores.

En la figuras 3.8 se muestra la señal original y tratada para todo el periodo. En la figura 3.9 se muestra un ejemplo de como se limpia la señal original.

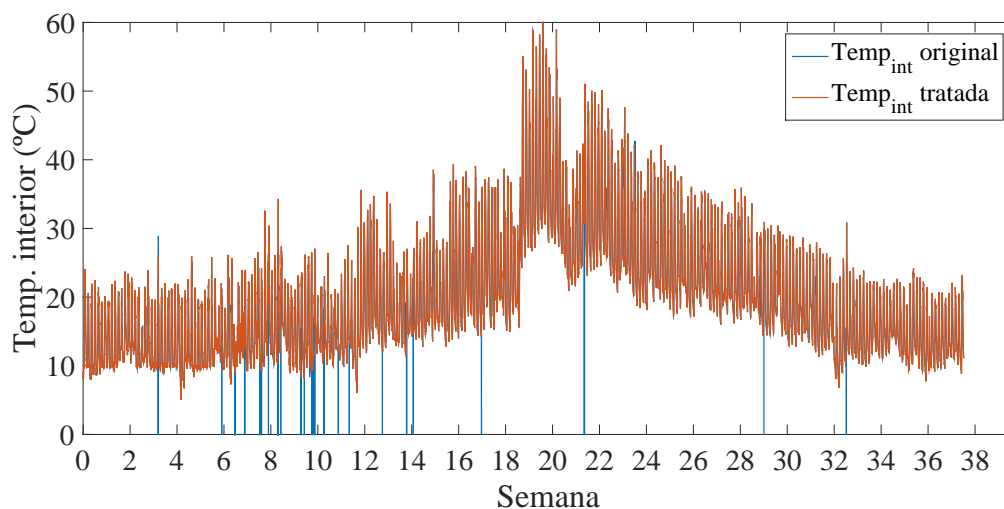


Fig. 3.8. Señal original y tratada de la temperatura interior (°C) durante el periodo de registro

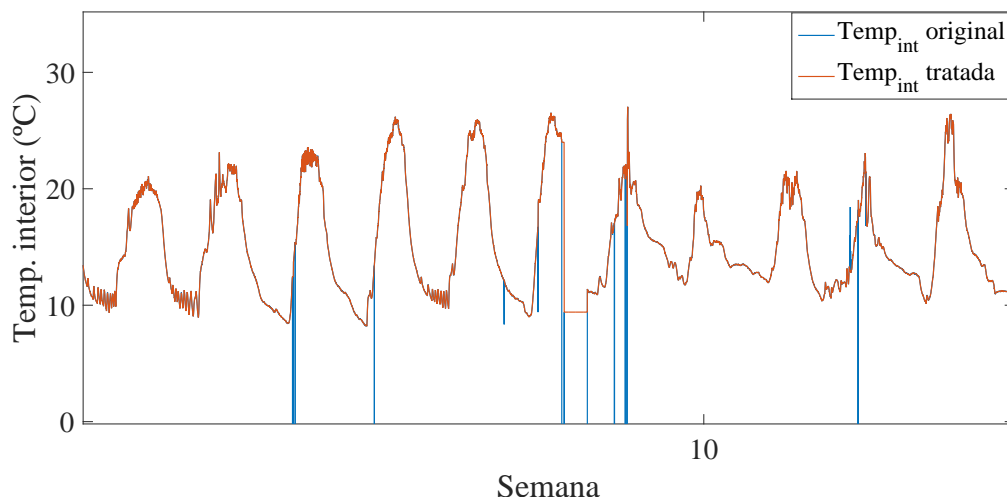


Fig. 3.9. Detalle de la señal original y tratada de la temperatura exterior (°C)

Temperatura suelo

Al igual que las anteriores mediciones de temperatura, se aplican las funciones fillmissing y filloutliers, en este caso con una ventana móvil de 40 puntos lo que equivale a 20 minutos y da buenos resultados.

En las figuras 3.10 se muestra la señal original y tratada para todo el periodo y en la figura 3.11 un ejemplo del tratamiento aplicado a la señal original.

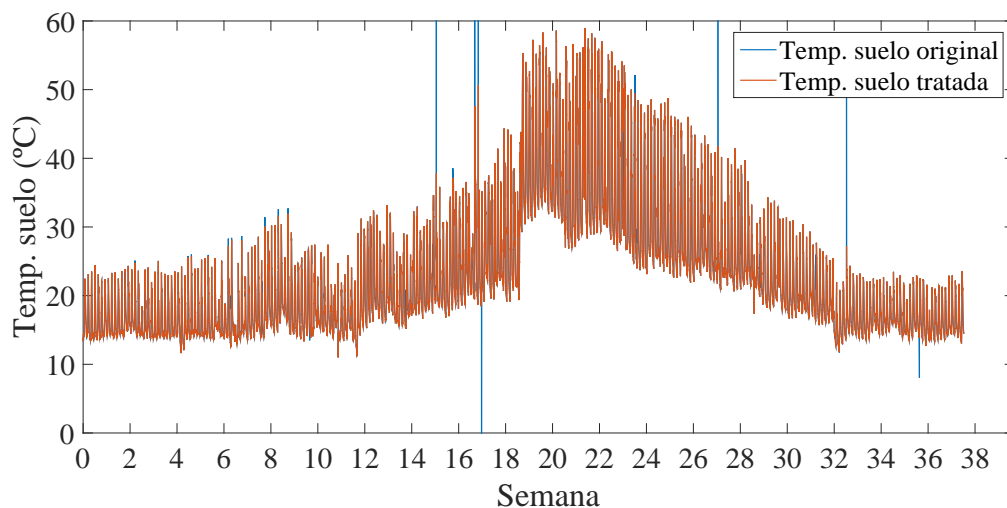


Fig. 3.10. Señal original y tratada de la temperatura del suelo (°C) durante el periodo de registro

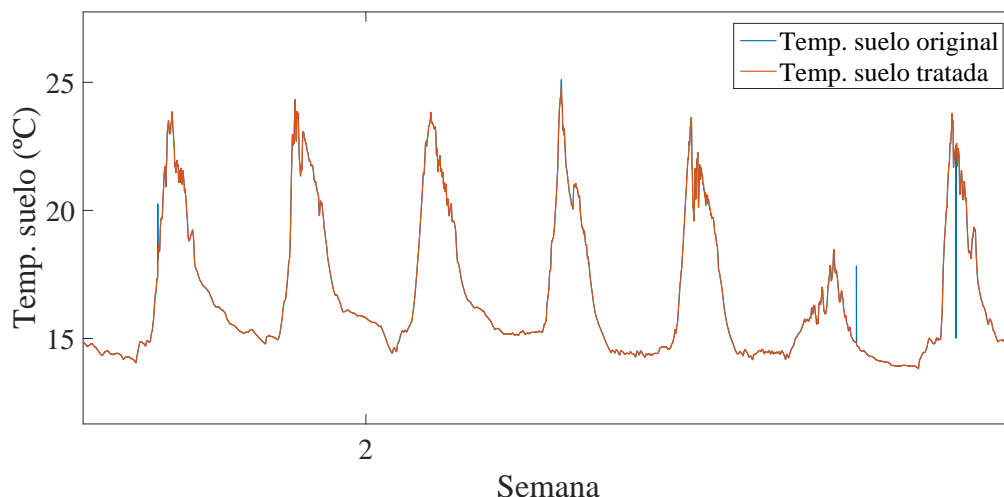


Fig. 3.11. Detalle de l señal original y tratada de la temperatura del suelo (°C)

Humedad interior

En primer lugar se sustituyen los valores menores que cero y mayores que 100, por los valores más cercanos posibles.

Luego, se aplica la función *fillmissings* y *filloutliers* con el método de media móvil con una ventana de 10 min (20 muestras), para eliminar las discontinuidades así como falsas medidas que no siguen la distribución normal de los datos. En las figuras 3.12 y 3.13 se muestran los resultados obtenidos.

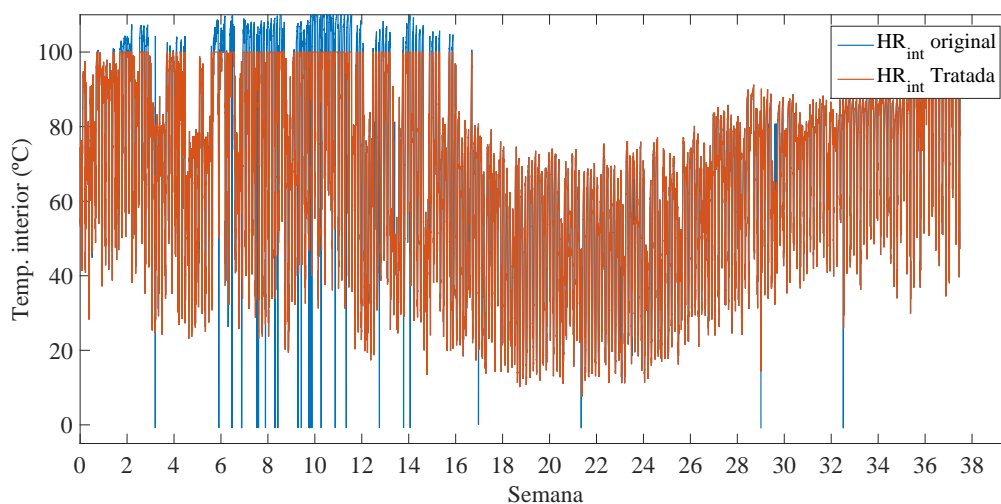


Fig. 3.12. Señal original y tratada de la humedad interior (%) durante el periodo de registro

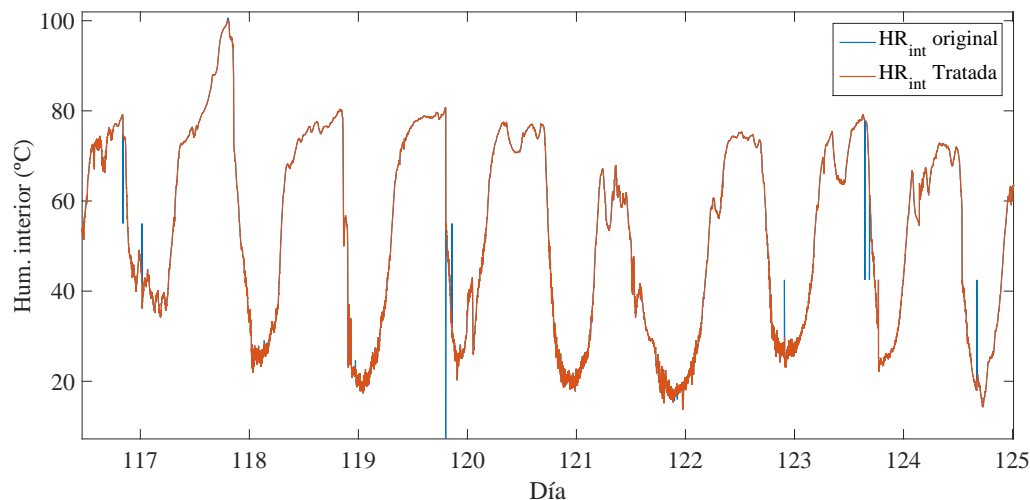


Fig. 3.13. Detalle de la señal original y tratada de la humedad interior (%)

Se observa como existían en la señal original bastantes valores por encima del 100% de humedad relativa, debido a las altas humedades que se alcanzan dentro del invernadero y que el sensor no esta debidamente calibrado. También se encuentran falsas medidas. Ambas han sido corregidos con éxito.

Humedad exterior

Se aplica la función *fillmissings* y *filloutliers* con el método de media móvil con una ventana de 10 min (20 puntos). En las figuras 3.14 y 3.15 se muestran los resultados obtenidos.

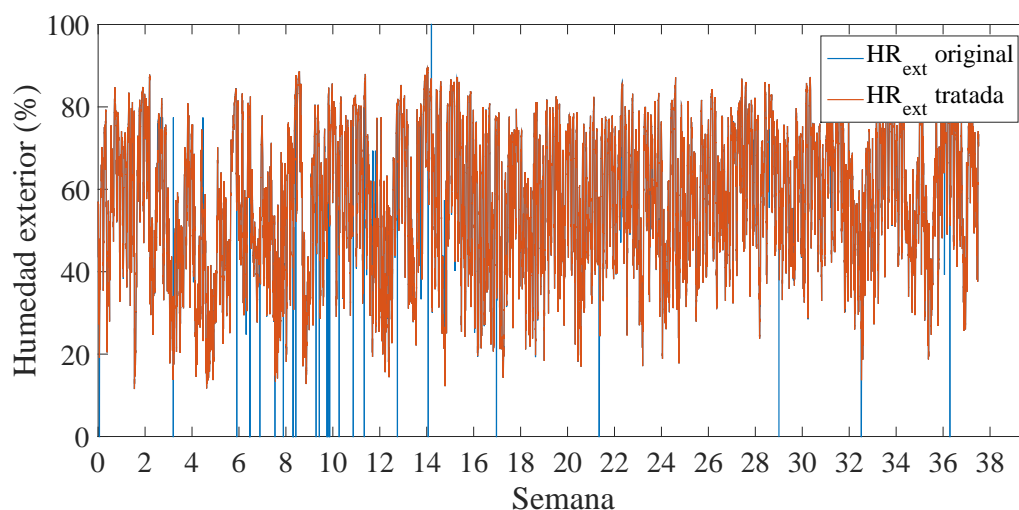


Fig. 3.14. Señal original y tratada de la humedad exterior (%) durante el periodo de registro

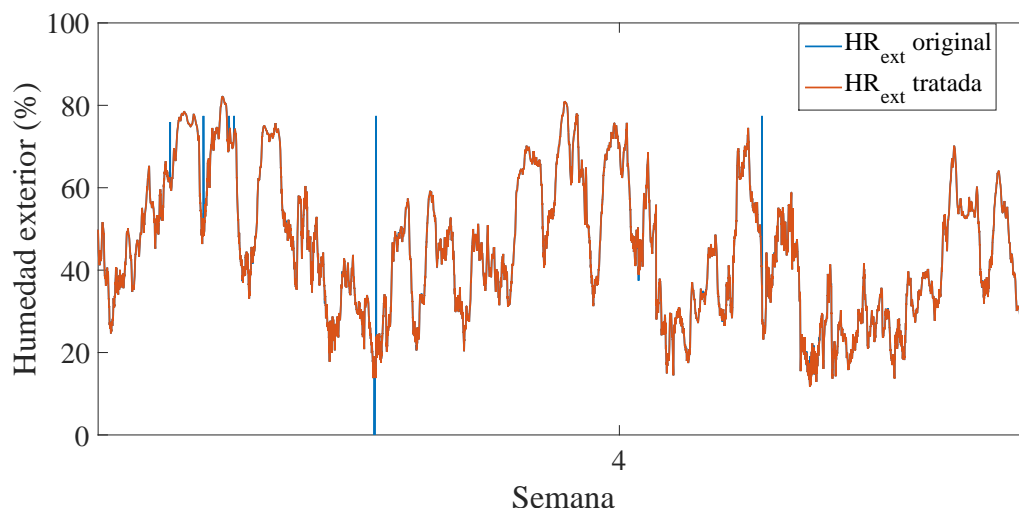


Fig. 3.15. Detalle de la señal original y tratada de la humedad exterior(%)

Una vez analizadas y tratadas las señales correspondiente a las medidas físicas de las perturbaciones, se pasa a representar las señales procedente de los actuadores y equipos. Las señales de activación son valores lógicos, donde -1 significa que el funcionamiento no es automático, mientras que los valores 0 y 1 provoca la parada y actuación del equipo correspondiente de forma programada.

Consigna de calefacción y señal de la bomba

Se muestra en la figura 3.16 la representación de la consigna establecida para la temperatura de calefacción durante todo el periodo de registro.

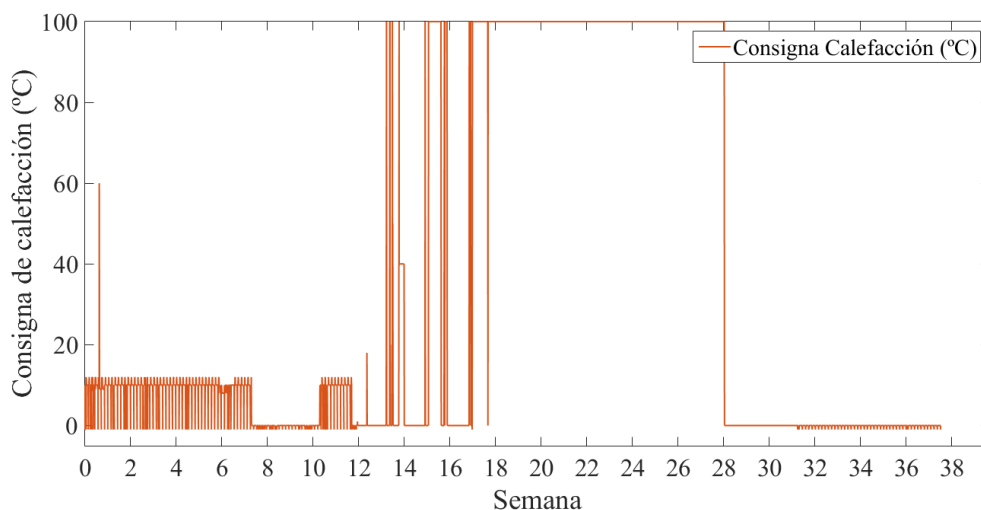


Fig. 3.16. Consigna de calefacción (°C) durante el periodo de registro

Se observa como existen valores extraños, menores que 0 y superiores a 20 °C lo cual no tiene mucho sentido. En la figura 3.17 se acotan los ejes para poder obtener más información

de esta señal.

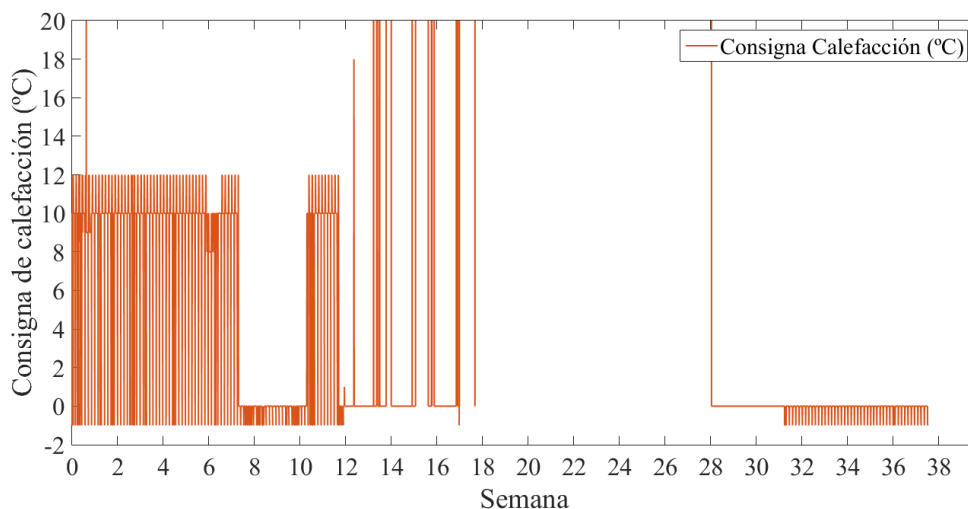


Fig. 3.17. Zoom de la consigna de calefacción (°C) durante el periodo de registro

Para que la calefacción funcione de forma automática, la consigna debe de tener un valor definido y lógico desde el punto de vista físico y funcional de la calefacción. Por tanto el funcionamiento automático de la calefacción solo se realizo hasta la semana 18, reduciéndose a la semana 12 si tenemos en cuenta que deben ser valores lógicos y con sentido físico. En este periodo la consigna varia entre 8 y 12 °C, lo cual tiene sentido. También se observa instantes en los que la consigna establecida cae al valor -1. Para entender porque ocurre esto, en la figura 3.18 se muestra en detalle la consigna establecida para la calefacción así como la señal de activación de la bomba.

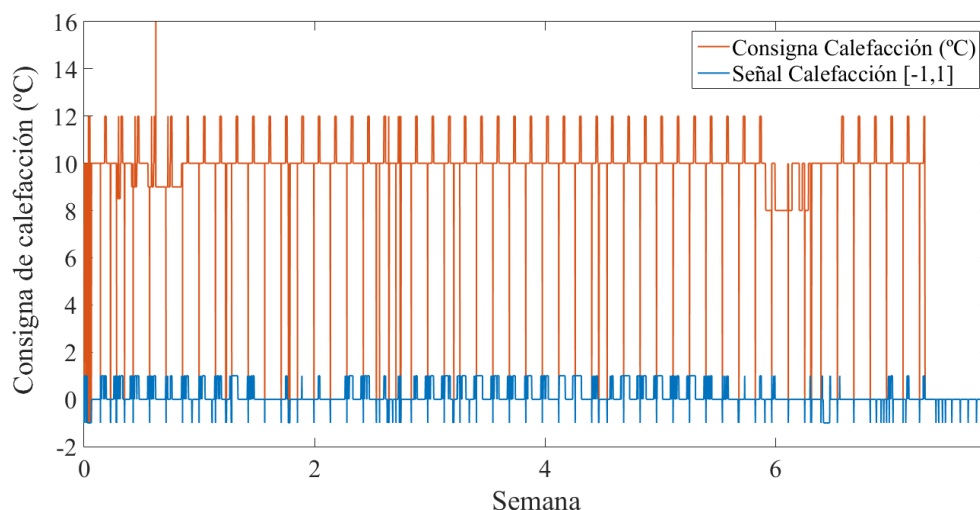


Fig. 3.18. Detalle de la consigna de calefacción (°C) y señal lógica de actuación de la bomba calefacción

Se observa como con cierta frecuencia periódica la señal de activación cae a -1 y es en esos momentos cuando la consigna de calefacción también se va a -1. Sin embargo esto no parece

afectar el comportamiento de la bomba de calefacción como se muestra en la figura 3.19 donde se aprecia la señal de activación de la bomba junto con el consumo, percibiéndose la actuación de la bomba.

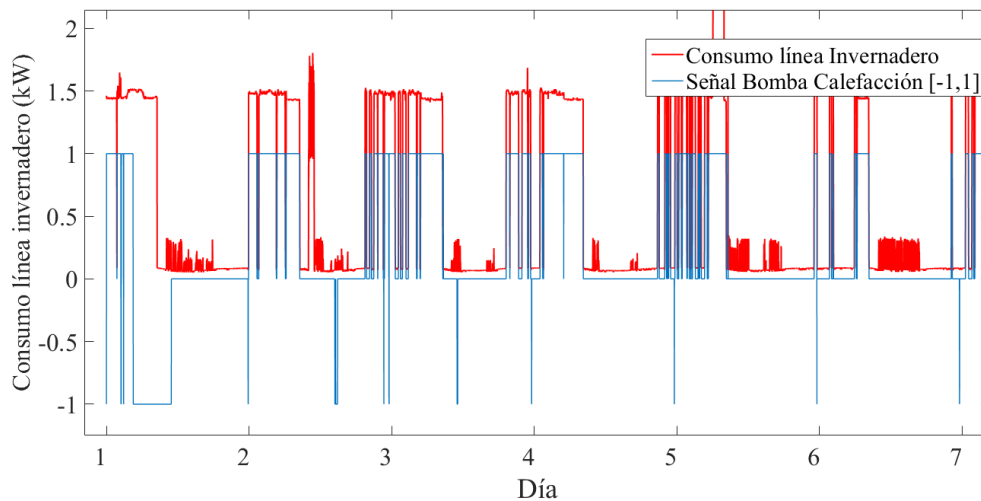


Fig. 3.19. Señal de actuación de la bomba calefacción y consumo (kW) de la línea invernadero

Podemos concluir que si queremos datos fiables de la actuación programada de la calefacción mediante consignas, debemos restringirlos al periodo comprendido de la semana 1 a la 12.

Actuación Soplante

A continuación se va a analizar la actuación de la soplante. En primer lugar, se muestra en la figura 3.20 la señal de activación de la soplante.

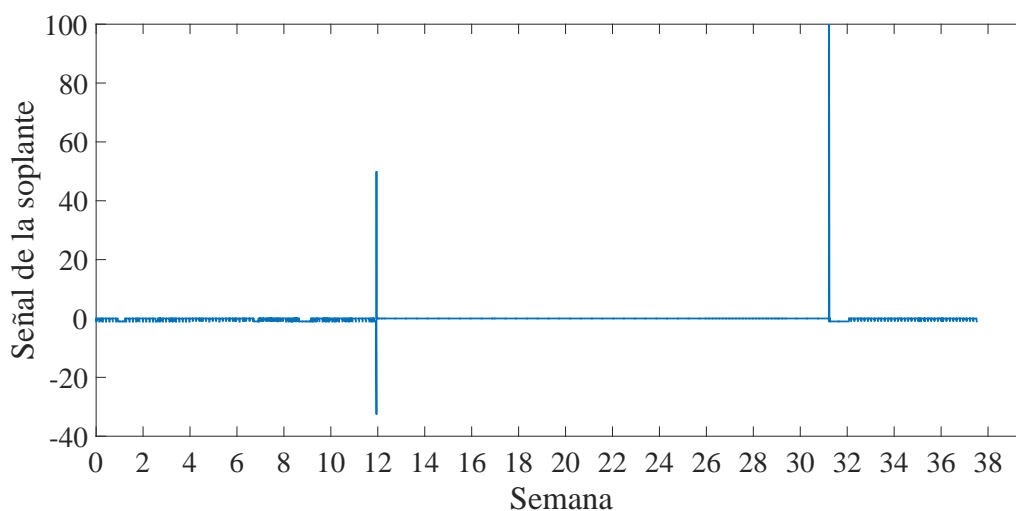


Fig. 3.20. Señal de actuación de la soplante durante el periodo de registro

Se observan tanto valores positivos como negativos. La señal de activación es una variable

lógica y debería tomar valores entre -1 y 1. Sin embargo existen valores inusuales en algunos instantes que indican un mal registro o funcionamiento de la señal. En la figura 3.21, obviando estos valores, se realiza un zoom sobre la figura para obtener una mejor percepción de los datos.

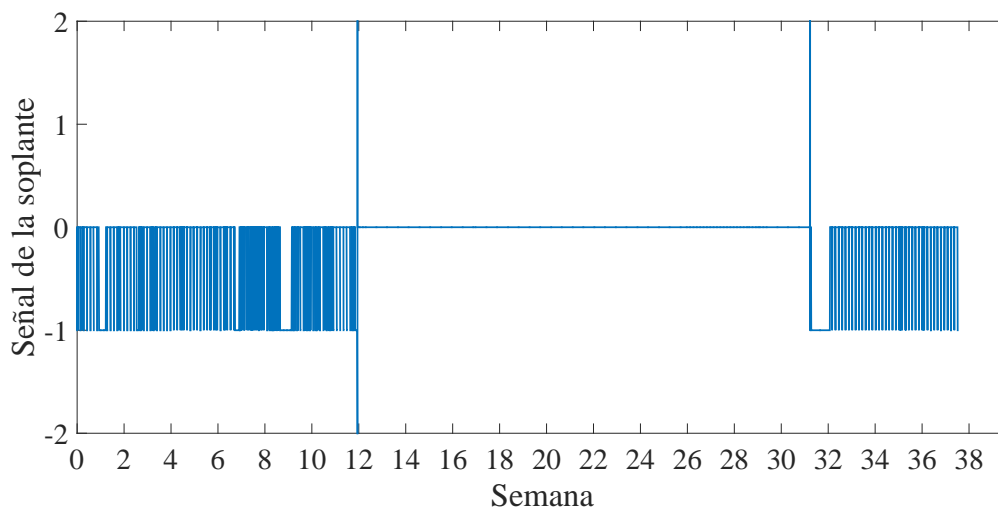


Fig. 3.21. Zoom de la señal de actuación de la soplante durante el periodo de registro

Se observa como la señal de activación de la soplante no toma el valor 1 en ningún instante y por tanto no funciona de forma programada en ningún momento.

Actuación Aerotermo

A continuación se analiza la actuación del aerotermo. Se muestra en la figura 3.22 la señal lógica enviada al aerotermo.

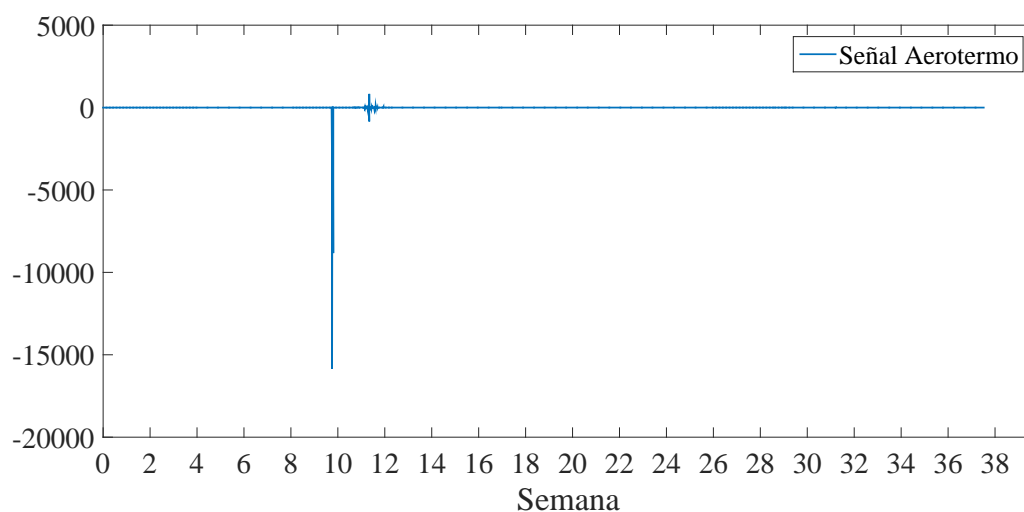


Fig. 3.22. Señal de actuación del Aerotermo durante el periodo de registro

Esta señal debería tener valores entre -1 y 1. Sin embargo se observa una serie de valores

muy grandes que nos impiden obtener ninguna conclusión. Desecharemos estos valores como erróneos para acotar la representación, la cual se muestra en la figura 3.23

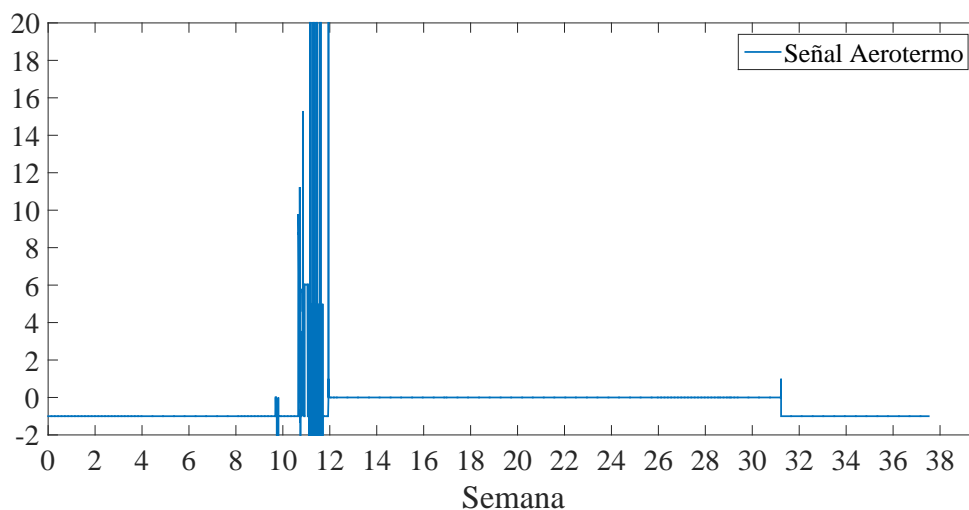


Fig. 3.23. Zoom de la señal de actuación del Aerotermino durante el periodo de registro

Se observan valores extraños. Además, el valor de la señal de activación no se establece en ningún momento en 1, por tanto parece que el aerotermino no ha funcionado mediante consigna. Para confirmar esto en la figura 3.24 se muestra la señal junto al consumo en la línea invernadero.

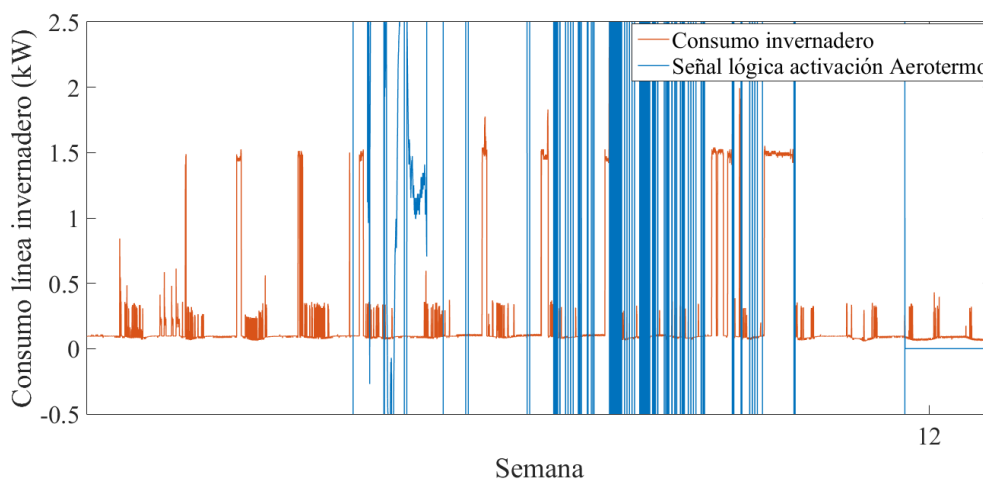


Fig. 3.24. Señal de actuación del Aerotermino y consumo de la línea invernadero (kW)

Se observa como la señal aerotermino no produce un aumento en el consumo debido al funcionamiento del aerotermino. Además esta hipótesis ha sido contrastada con personal de la planta experimental quienes confirmaron el no funcionamiento del aerotermino en el periodo de registro de los datos.

Actuación Ventilación

A continuación se va a analizar la actuación del sistema de ventilación natural. En la figura 3.25 se representa la consigna de ventilación, para todo el periodo de registro.

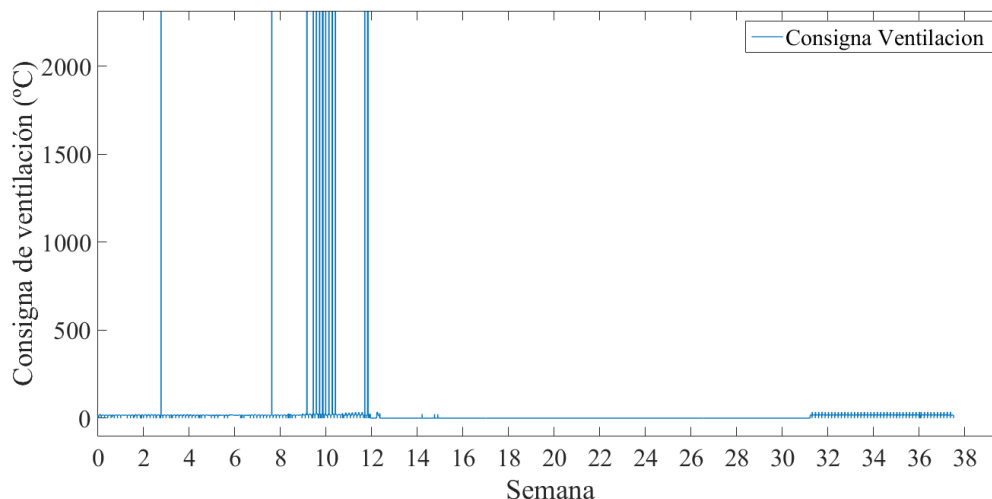


Fig. 3.25. Consigna (°C) de la ventilación, establecida para la actuación automática de las ventanas durante el periodo de registro

Al igual que ocurría con la calefacción existen valores sin sentido que impiden obtener ninguna conclusión. Por ello se acotan los ejes en la representación tal y como se muestra en la figura 3.26

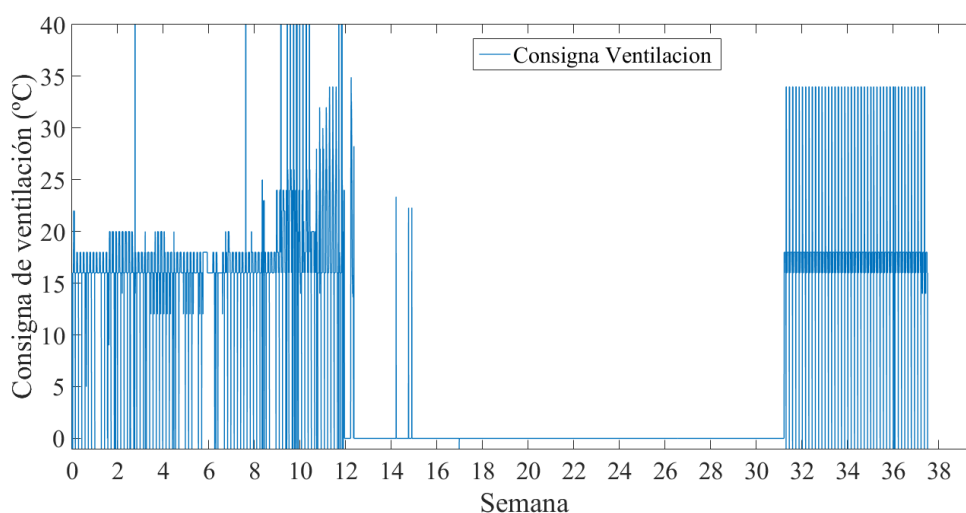


Fig. 3.26. Zoom de la consigna de la ventilación (°C) , establecida para la actuación automática de las ventanas durante el periodo de registro

En esta gráfica 3.26 se observa como la consigna de ventilación varía durante la temporada entre los valores de 25 y 16 °C. Al igual que ocurría con la calefacción, se observa como con

cierta periodicidad la consigna toma el valor -1, ocasionado por un cambio en la consigna o un cambio en la señal de apertura.

También se observa como no existe consigna de la semana 12 a la 32, por lo que el funcionamiento automático de las ventanas se restringe entre las semanas 0-12 y la 32-38. Esto queda de manifiesto en la figura 3.27, donde se representa la señal de apertura enviada a cada ventana. Decir en este punto que la señales a los actuadores no son valores entre -1 y 1, sino la apertura en tanto por cien que debe tener la ventana.

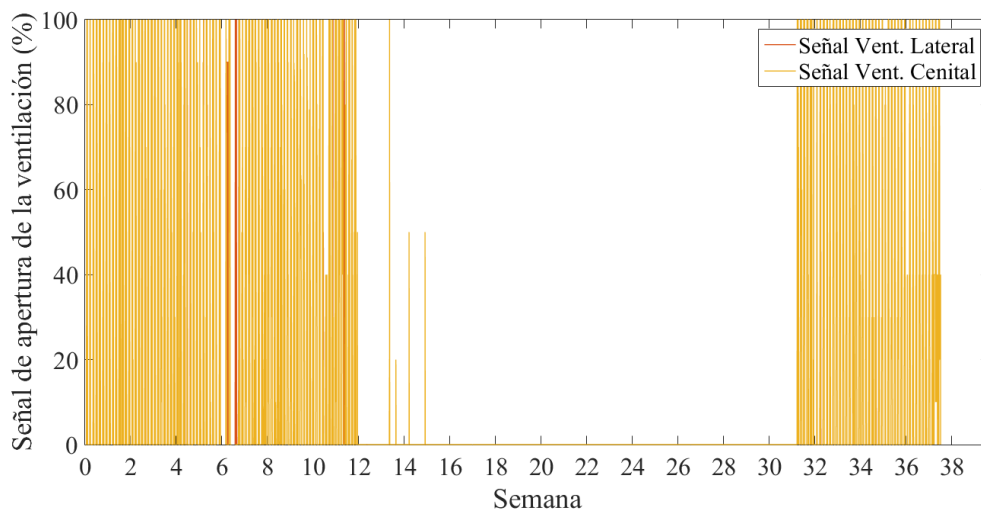


Fig. 3.27. Señal de apertura (%) enviada a la ventilación cenital y lateral durante el periodo de registro

Por último en la figura 3.28 se muestra el consumo de la línea invernadero debido a la actuación de la ventilación. Conjuntamente se representa la señal de apertura enviada (se ha puesto en tanto por uno para facilitar la representación).

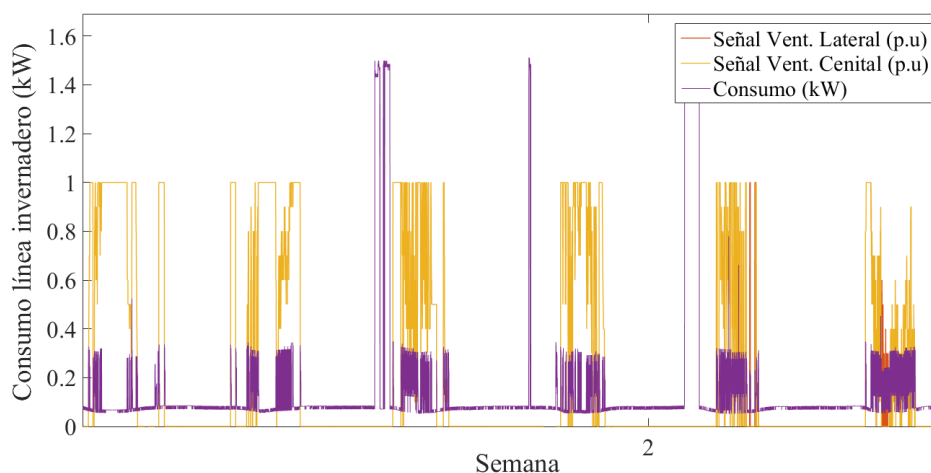


Fig. 3.28. señales de actuación de las ventanas (p.u) y consumo de la línea invernadero (kW)

Se observa como la señal de apertura toma valores entre 0 y 1 que corresponden a valores entre 0 y 100 en tanto por ciento. Cuando se produce un cambio en el valor de la señal de

apertura, se manifiesta en la línea invernadero el consumo debido a la ventilación. También se ha visto en otras gráficas que el cambio de la consigna establecida para la ventilación, provoca una activación momentánea de los actuadores de las ventanas.

Se concluye que el periodo fiable de actuación programada de la ventilación mediante consignas, se restringe a las semanas 0-12 y la 32-38.

Actuación Deshumificadora

A continuación se va a analizar la actuación de la deshumificadora. En la figura 3.29 se muestra la señal deshumificadora en el periodo de registro.

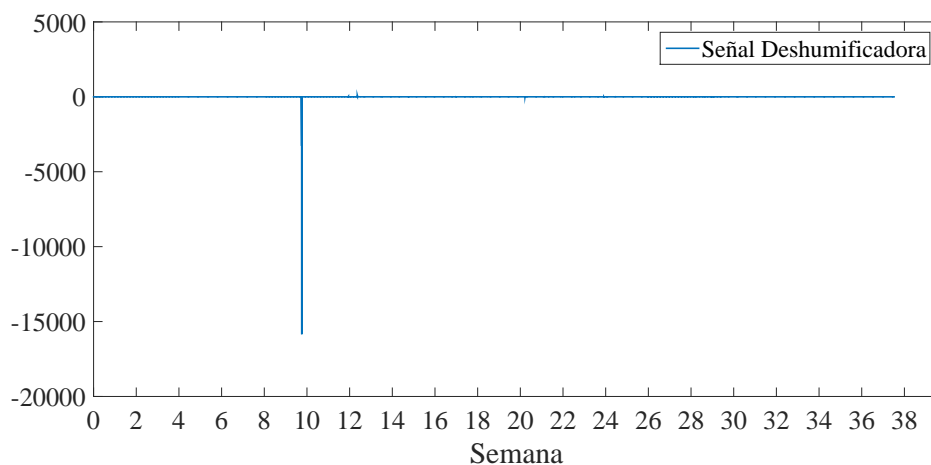


Fig. 3.29. Señal de actuación de la máquina deshumificadora durante el periodo de registro

Existen valores extraños muy altos que nos impiden obtener conclusiones. Por ello se presenta la figura 3.30, donde se han acotado los ejes para apreciar mejor los datos.

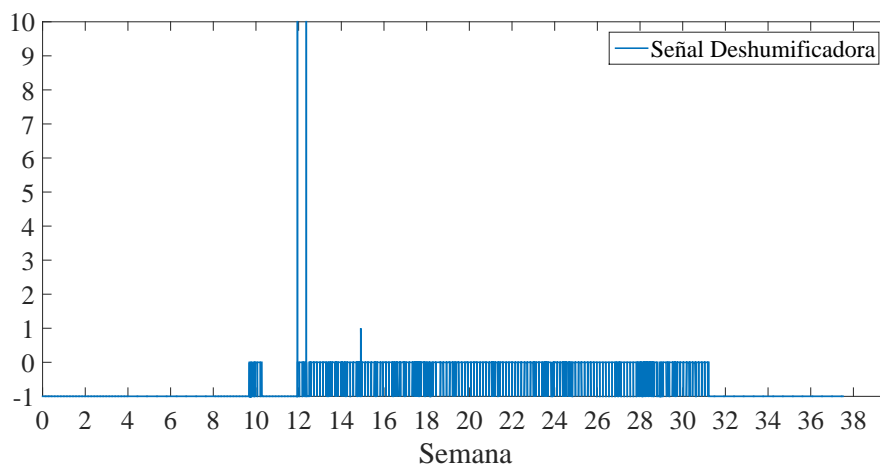


Fig. 3.30. Zoom de la señal de actuación de la máquina deshumificadora durante el periodo de registro

Se observa como el funcionamiento de la máquina no ha sido en modo automático. Perso-

nal de la planta nos confirma que la máquina no funciona de forma automática aunque si se realizaron ensayos manuales.

En la figura 3.31 se muestran dos periodos donde el consumo se eleva debido al funcionamiento de la máquina deshumificadora. Sin embargo en esos periodos la señal de activación no estaba a 1, lo que indica el funcionamiento manual.

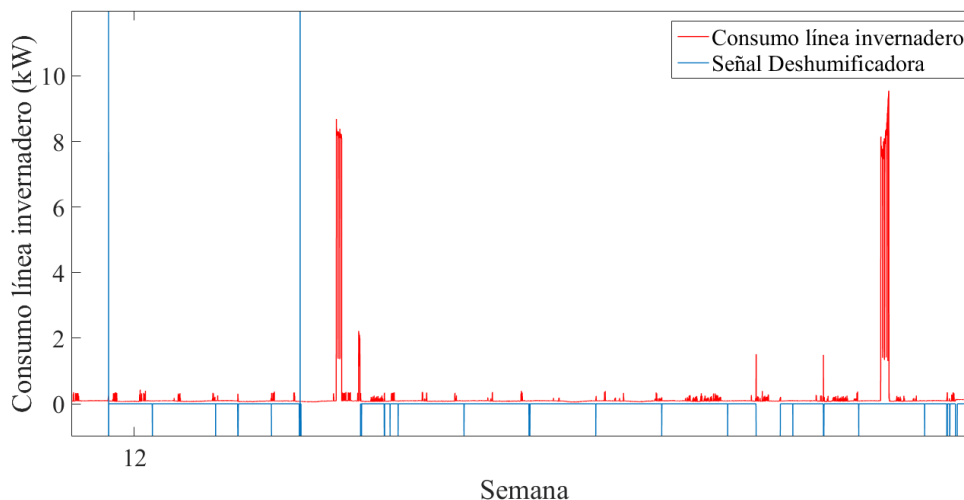


Fig. 3.31. Señal de actuación de la máquina deshumificadora y consumo de la línea invernadero (kW)

Se concluye que con los datos y con la información aportada por el personal de la estación que el funcionamiento de la máquina deshumificadora no ha sido de forma automática mediante las consignas programadas. Sin embargo, debido a que este supone uno de los mayores consumos y que aparecen en la línea de invernadero durante el periodo de registro, se va a suponer su señal de activación como se explica más adelante.

Se va a pasar ahora al análisis de los consumos registrados tanto en la línea invernadero como en la línea biomasa. Se trata de las señales más importantes ya que son aquellas que queremos modelar.

Consumo en la línea invernadero

Se va a analizar el consumo registrado en la línea invernadero. Se intentará relacionar este con los diferentes equipos conectados a la línea, así como establecer los patrones de consumo tipo de cada uno de los equipos. En primer lugar se muestran en la figura 3.32 los datos brutos registrados en la línea invernadero.

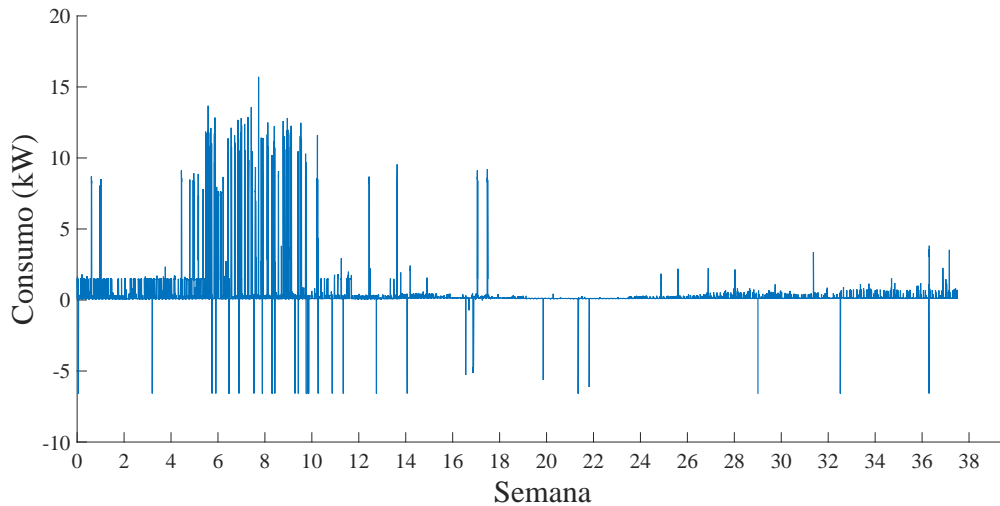


Fig. 3.32. Señal original del consumo de la línea invernadero (kW) durante el periodo de registro

Se observa en la figura 3.32 como existen valores negativos que son físicamente imposibles, ya que el consumo no puede ser negativo. También existen discontinuidades en la medida como se muestra en la figura 3.33.

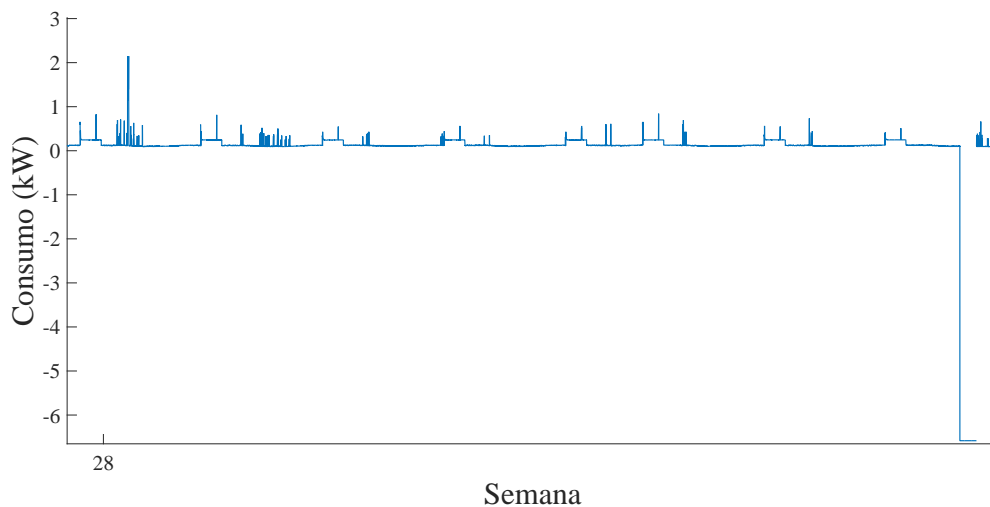


Fig. 3.33. Detalle de la señal original del consumo de la línea invernadero (kW) durante el periodo de registro

Se van a eliminar los valores negativos y reemplazar los huecos mediante la función fillmissing por el valor disponible inmediatamente anterior. Además podemos descartar el filtrado de esta variable pues no existe un ruido que distorsione la señal. En la figura 3.34 se observa el resultado.

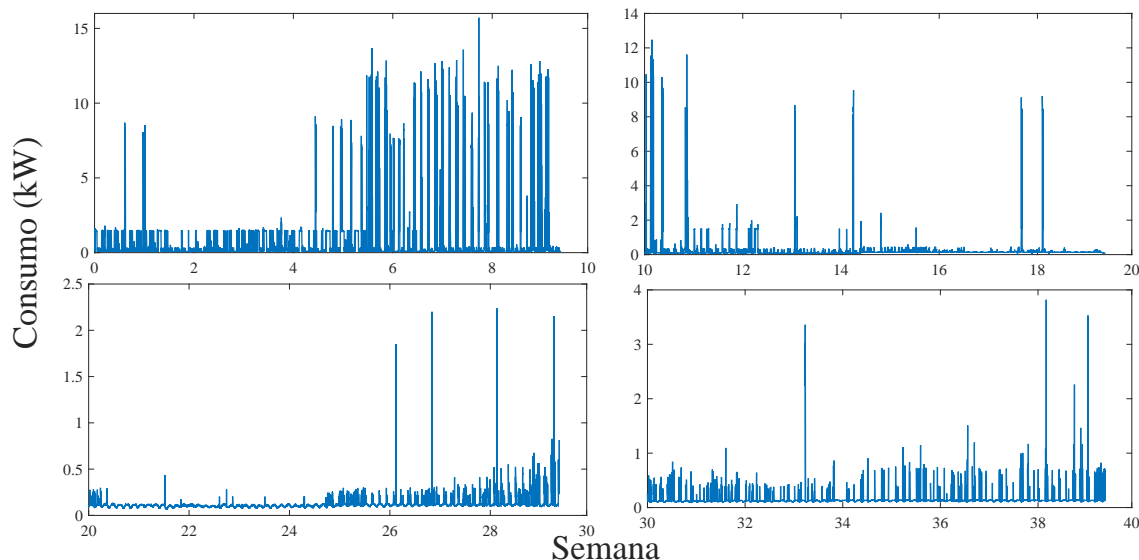


Fig. 3.34. Consumo línea invernadero (kW) una vez eliminados los valores negativos y las discontinuidades

Se ha descompuesto en 4 gráficas para facilitar su análisis. En primer lugar se observa como se han corregido los valores negativos. Atendiendo a los máximos observamos que el pico de consumo está en torno a los 16 kW, superior a la potencia máxima que se puede alcanzar en la línea invernadero según la tabla 2.1. Se buscará una explicación a este hecho aunque el pico de arranque en equipos se presenta como principal opción. Los consumo más altos se dan en las primeras semanas, lo cual tiene sentido ya que se trata de Enero y Febrero los meses mas fríos, por lo que tanto la calefacción como la deshumificadora pudieron ser usadas. El consumo nunca es 0, existe un consumo base debido a los sensores y equipos de adquisición. Este consumo se aprecia en la figura 3.35. Sobre este consumo aparecen los incrementos que se deben al uso de los equipos de control de clima.

Se va a intentar identificar el perfil de consumo que produce cada equipo existente en la línea invernadero y que aparecen superpuestos en la figura 3.34

En primer lugar, en la figura 3.35 se muestra el consumo debido al sistema SCADA y los sensores asociados. Esta figura es un detalle de la figura 3.34 en un periodo donde no hay funcionamiento de otros equipos, permitiendo aislar el consumo base constante que supone el equipo de medida.

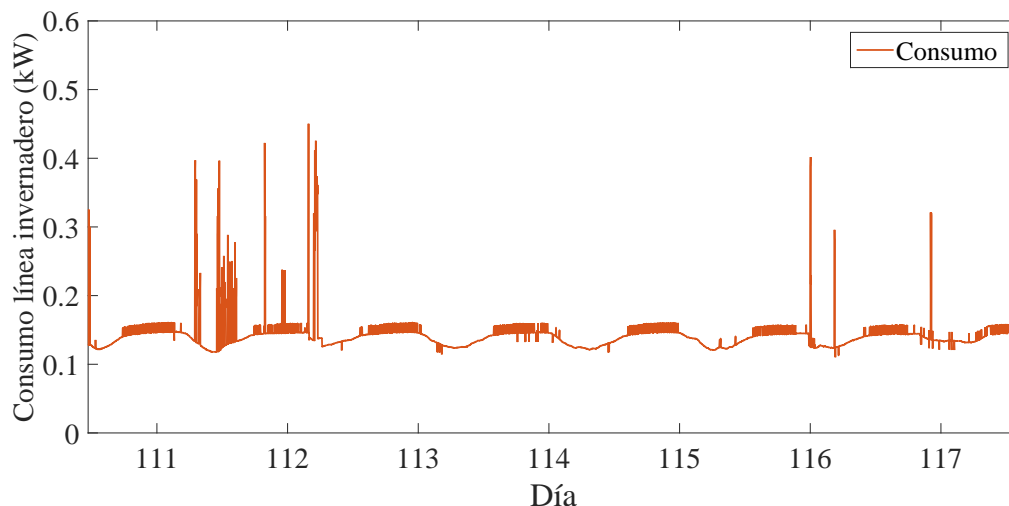


Fig. 3.35. Perfil tipo del consumo del equipo de medida de la línea invernadero (SCADA+sensores)

Se observa como este consumo tiene una forma sinusoidal (posiblemente debido al funcionamiento de los sensores con el día y la noche). El consumo está acotado entre 100 y 200 W, valores compatibles con las potencias asignadas a los sensores y SCADA en la tabla 2.1. Los picos que elevan el consumo por encima de 200 W se deben a la actuación de las ventanas.

En la figura 3.36 se muestra el consumo debido a la actuación de la bomba de calefacción.

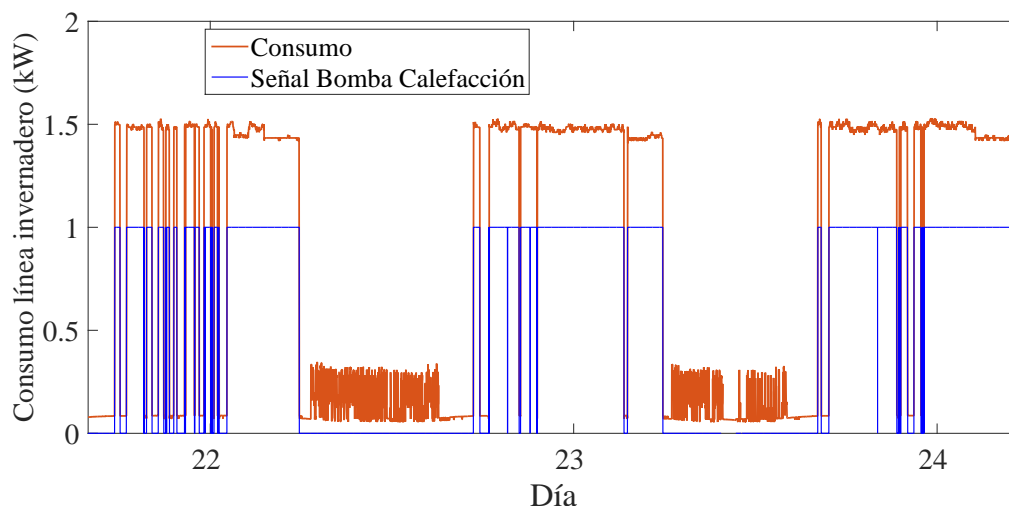


Fig. 3.36. Perfil tipo del consumo de la bomba de calefacción (kW)

Se observa como el consumo de la bomba de calefacción se mantiene bastante constante mientras está en funcionamiento. Este consumo se mantiene entorno a 1500 W, valor compatible con la potencia nominal indicada en la tabla 2.1

En la figura 3.37 se muestra el perfil de consumo debido a la máquina deshumificadora.

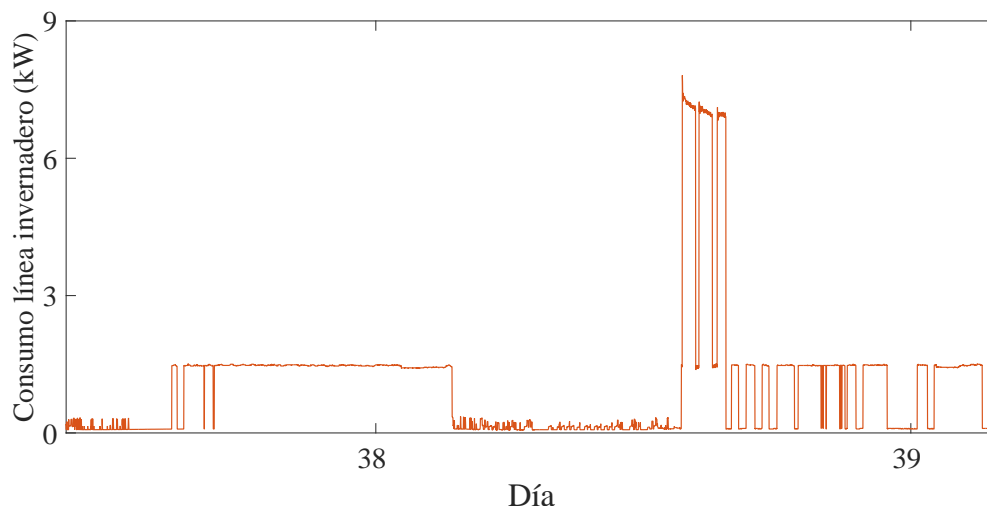


Fig. 3.37. Perfil tipo del consumo (kW) de la máquina deshumificadora

Se trata de las tres barras que elevan el consumo hasta los 8 kW. Este consumo varía debido al ciclo frigorífico de la máquina. Además puede verse un comportamiento típico de los invernaderos tecnificados. En los primeros instantes de la figura se observa el consumo debido a la actuación de la ventilación durante un día. Después llega la noche y entra en funcionamiento la calefacción (consumo constante a 1.5 kW). Amanece de nuevo y vuelve a funcionar la ventilación hasta que vuelve a anochecer y en este caso arranca la máquina deshumificadora y más tarde también la calefacción.

Aunque la máquina no ha funcionado en modo programado, si que ha funcionado de forma manual y por ello puede verse su perfil en el consumo.

Por último, en la figura [3.28](#) se mostró el perfil de consumo debido a la actuación de las ventanas. Los motores de las ventanas producen incrementos del consumo de entre 160 y 350 W cuando se cambia el valor de la señal de apertura de las ventanas. La actuación de las ventanas también puede percibirse en las figura [3.35](#) y [3.36](#)

En el registro de consumo de la línea invernadero existen una serie de consumos que no se han podido asignar al perfil de ninguno de los equipos existentes en la línea. Estos valores extraños aparecen de la semana 4 a la 12 y por lo general superan la potencia máxima asignada a esa línea (12 kW). En la figura [3.38](#) se muestra un ejemplo de estos valores.

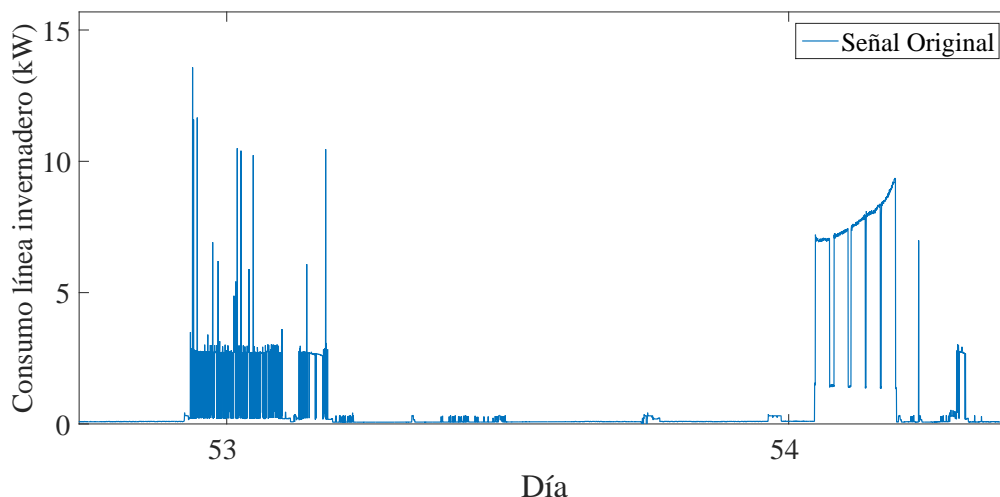


Fig. 3.38. Consumo Desconocido

Se trata de los valores de consumo a la izquierda de la gráfica. Aunque estos valores son similares a los alcanzados por la máquina deshumificadora, los cuales pueden verse a la derecha de la gráfica, se observa como el perfil es diferente del perfil de consumo que implica la operación normal de la máquina deshumificadora, el cual está perfectamente definido. Además se trata de valores puntuales, que no son sostenidos en el tiempo, por lo que puede deberse a arranques y paradas de la máquina deshumificadora dentro de un ensayo. Tampoco puede deberse al aerotermo ya que este no funciona además que su consumo es menor y sería constante durante cierto tiempo.

Por tanto, debido a que estos valores no son ocasionados por la actuación automática de ningún equipo, que es lo que aquí se intenta predecir, no son de interés y se considerarán falsas medidas.

La detección de dichas falsas medidas sin eliminar otras informaciones resulta bastante complicada. Se ha echo usando la función `movmedian` con una ventana móvil de 4 puntos lo que equivale a 2 minutos, Es una ventana muy pequeña, pero es necesaria si no se quiere desechar información que puede resultar importante.

En la figura [3.39](#) se muestra el resultado de aplicar dicha función a la señal original (sin valores negativos ni discontinuidades).

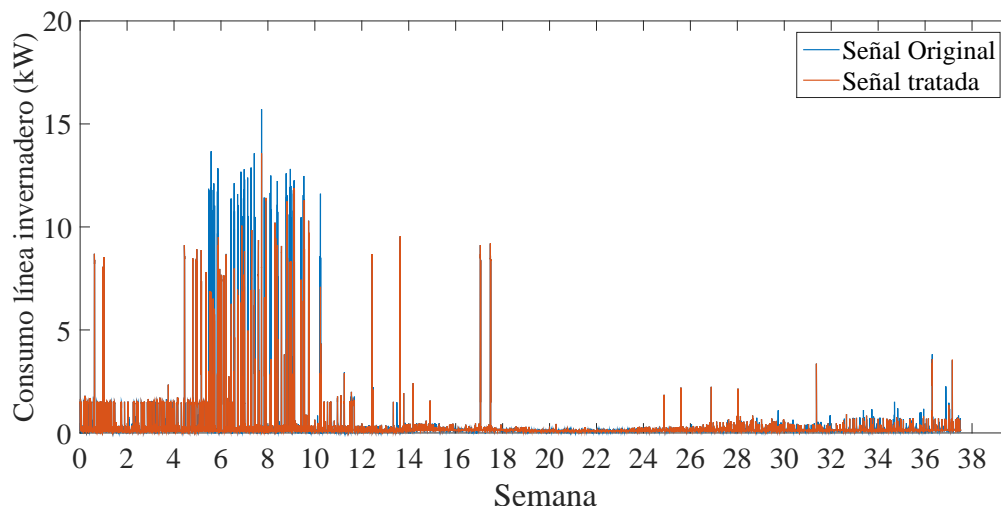


Fig. 3.39. Señal del consumo de la línea invernadero original y depurada de valores extraños

Por último en la figura 3.40 se muestra como se logran eliminar estos valores preservando el resto de información.

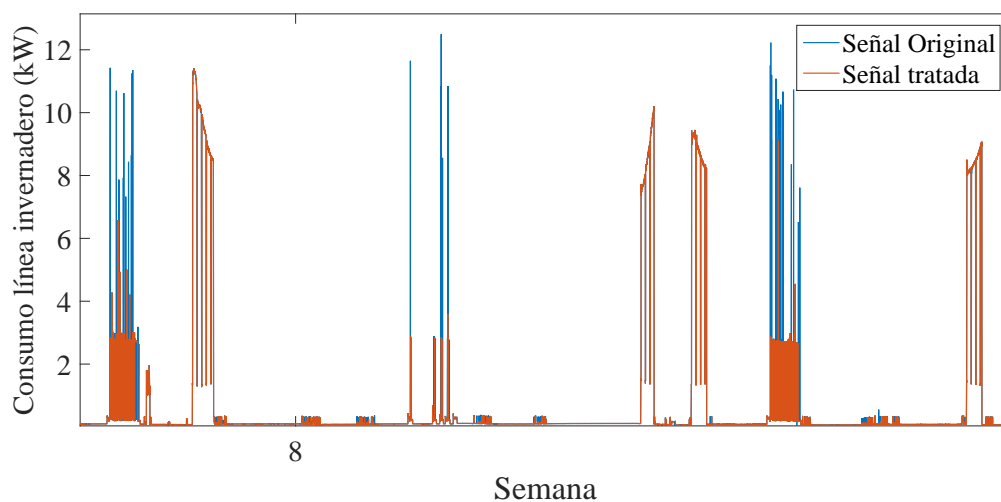


Fig. 3.40. Detalle de la señal del consumo de la línea invernadero original y depurada de valores extraños

Como existen ciertos valores que resultan muy difícil de eliminar incluso usando la función se tendrá cuidado a la hora de seleccionar los datos para el entrenamiento y validación de los modelos. Además también se han eliminado parte de estos valores extraños de forma manual, usando la función 'brush' de Matlab directamente sobre la gráfica de consumo.

Consumo de la línea biomasa

Se va a analizar el consumo registrado en la línea biomasa. Se intentara relacionar este con los diferentes equipos conectados a la línea, así como establecer los patrones de consumo tipo

de cada uno de los equipos. En primer lugar se muestran en la figura 3.41 los datos brutos registrados en la línea biomasa.

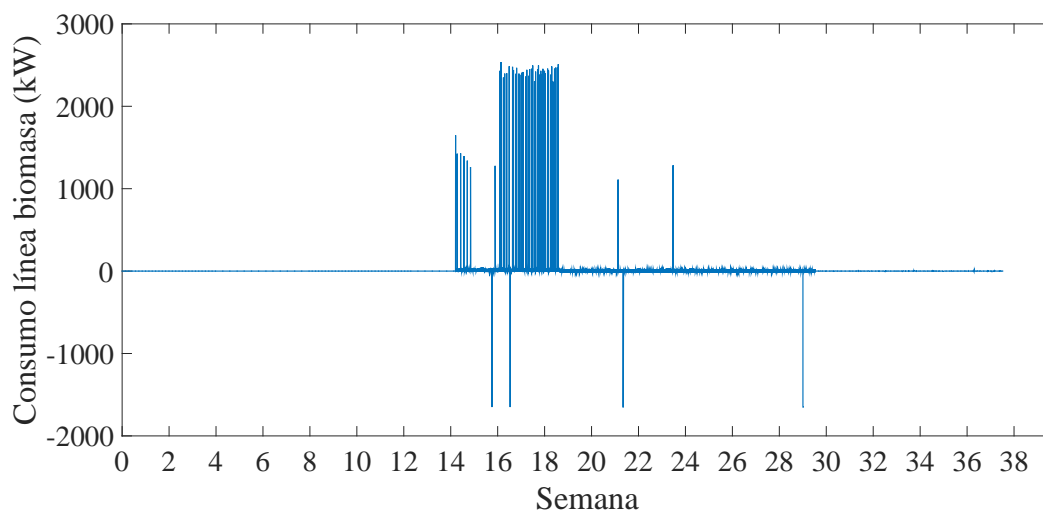


Fig. 3.41. Señal original del consumo de la línea biomasa (kW) durante el periodo de registro

Como puede observarse, existen valores atípicos con valores de consumo excesivamente altos, que son imposible y que impiden obtener conclusiones de la gráfica. En la figura 3.42 solo se representan valores positivos y se acota la representación para poder obtener información del registro.

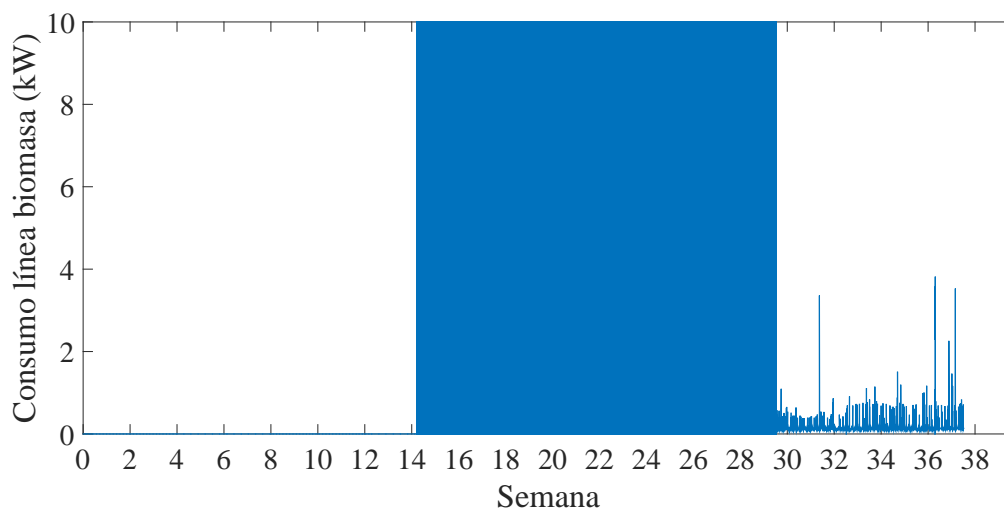


Fig. 3.42. Zoom de la señal original del consumo de la línea biomasa (kW) durante el periodo de registro

Analizando la figura podemos observar tres periodos diferenciados. De la semana 0 a la 14 no existe registro de datos ya que el valor del consumo se mantiene constante a 0. Por otra parte de la semana 14 a la 30 se registra una serie de valores atípicos, oscilantes de valores negativos a positivos y con una magnitud demasiado grande para poderse asignar a algún equipo. Este

registro se debe a un mal funcionamiento del sensor o a algún ruido en la señal que impide obtener información de este periodo. En la figura 3.43 se muestra en detalle este periodo.

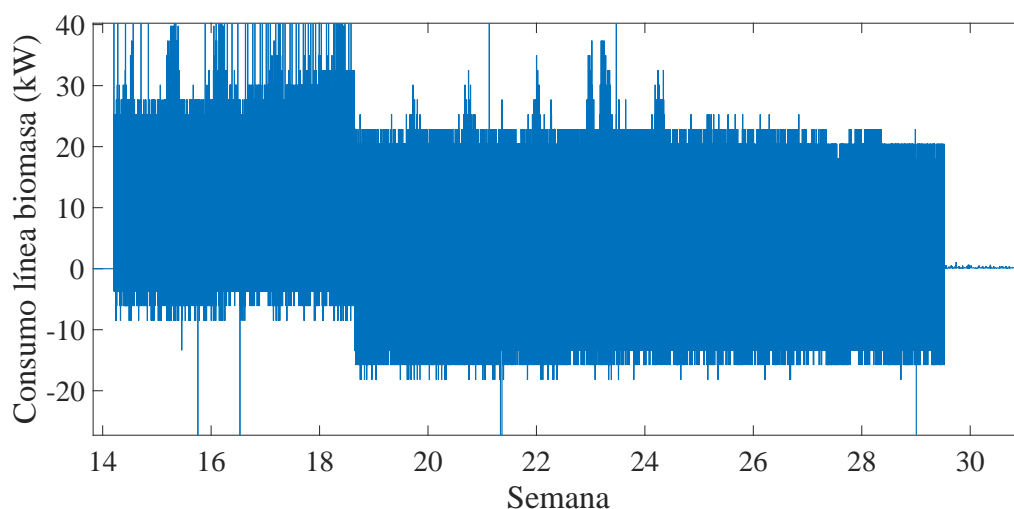


Fig. 3.43. Valores atípicos registrados en la señal consumo de la línea biomasa de la semana 14 a 30

Por último en la figura 3.44 se muestra el último periodo, de la semana 30 al final del periodo de registro.

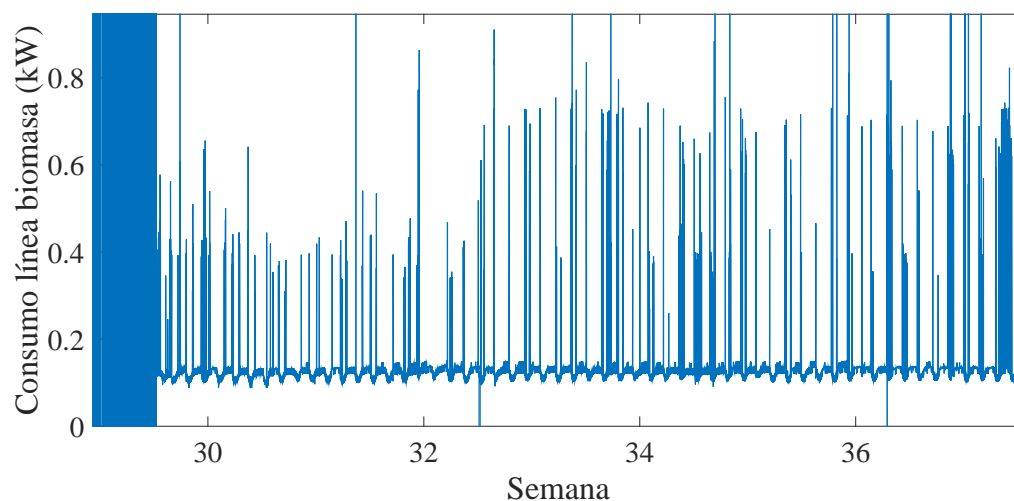


Fig. 3.44. Consumo línea biomasa Detalle

En este periodo, si se observa información legible. Se aprecia el consumo debido al equipo de medida (SCADA+Sensores) con su perfil sinusoidal de consumo entre 100 y 200 W, al igual que en la línea invernadero, mostrado en la figura 3.35. Sobre este consumo base se producen picos de consumo. Sin embargo estos son muy pequeños para deberse a la soplante, con una potencia nominal de 1800 W, y muy grandes para deberse a los equipos auxiliares de la caldera de biomasa (210 W). Además de que la calefacción no funcionó en ese periodo como se demostró anteriormente.

Por tanto, no se puede obtener información alguna de la línea de biomasa del uso de la soplante o la caldera de biomasa.

Como conclusión de este apartado, se han tratado y obtenido las variables candidatas a ser entradas de los modelos. Debido a la falta de datos para el resto de equipos y periodos, los equipos cuyo funcionamiento puede ser predicho mediante consigna se limita a la ventilación y a la calefacción de la semana 0 a la 12,

Como último paso del tratamiento de los datos, se ha usado la función *zscore* que permite a la serie histórica, obteniendo varianza 1 y media nula.

3.2. Selección de variables de entradas

En este apartado se ha estudiado que variables se deben incorporar al modelo y cuales se deben desechar por existir colinealidad. Las variables objeto de estudio son aquellas de tipo perturbación. Al tratarse de variables físicas es importante ver cuanta información añaden y cuales son desechables al estar su origen en un mismo fenómeno físico. Se parte por tanto de 7 variables independientemente medidas. Estas son:

TABLA 3.2. Variables estudiadas en la selección de variables

Variable	Unidades
Radiación	W/m^2
Temperatura exterior	$^{\circ}C$
Temperatura interior	$^{\circ}C$
Temperatura suelo	$^{\circ}C$
Humedad interior	%
Humedad exterior	%
Velocidad del viento	m/s
Temporal	KJ

No se incluye la variable de salida, el consumo, que por supuesto debe estar en la construcción del modelo. Por otra parte, se ha añadido una variable temporal. Esta pretende dar información acerca del estado y edad del cultivo, por ello se ha calculado como la integral de la radiación. La fotosíntesis, proceso en el cual las plantas se nutren incorporando carbono procedente del CO_2 del aire, solo se realiza cuando hay disponible radiación solar. Por tanto el crecimiento de la planta está directamente relacionado con la cantidad de radiación que estas reciben.

3.2.1. Análisis de correlación simple

En primer lugar se realiza el análisis de coeficientes de correlación simple bivariada, para tener una idea de la relación entre variables y la información compartida con la variable a

predecir. El resultado se muestra en la tabla 3.3

TABLA 3.3. Coeficientes de correlación simple bivariada

	y	Rad	Text	Tint	Tsuelo	Viento	HRext	HRint	tiempo
y	1	-0.091	-0.198	-0.153	-0.149	-0.027	-0.024	0.112	-0.213
Rad	-	1	0.546	0.657	0.622	0.355	-0.524	-0.704	-0.025
Text	-	-	1	0.944	0.918	0.187	-0.285	-0.783	0.338
Tint	-	-	-	1	0.970	0.233	-0.299	-0.827	0.225
Tsuelo	-	-	-	-	1	0.227	-0.253	-0.801	0.225
Viento	-	-	-	-	-	1	-0.331	-0.397	-0.095
HRext	-	-	-	-	-	-	1	0.588	0.152
HRint	-	-	-	-	-	-	-	1	-0.138
tiempo	-	-	-	-	-	-	-	-	1

Puesto que la relación es simétrica entre dos variables, la tabla también es simétrica por lo que solo se muestra el triangulo superior . Por otra parte en la tabla 3.4, se muestran los coeficientes de significancia de cada coeficiente de correlación mostrado en la tabla anterior.

TABLA 3.4. Matriz de significancia

	y	Rad	Text	Tint	Tsuelo	Viento	HRext	HRint	tiempo
y	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Rad	-	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Text	-	-	1	0.000	0.000	0.000	0.000	0.000	0.000
Tint	-	-	-	1	0.000	0.000	0.000	0.000	0.000
Tsuelo	-	-	-	-	1	0.000	0.000	0.000	0.000
Viento	-	-	-	-	-	1	0.000	0.000	0.000
HRext	-	-	-	-	-	-	1	0.000	0.000
HRint	-	-	-	-	-	-	-	1	0.000
tiempo	-	-	-	-	-	-	-	-	1

La matriz de significancia muestra valores menores que 0.05 por lo que aceptamos los coeficientes de correlación de la tabla 3.3, con un riesgo de equivocarnos y que no exista correlación entre las variables menor del 5 %.

Analizando la tabla 3.3 se obtienen las siguientes conclusiones.

- Para la variable a predecir, consumo del invernadero, observamos lo siguiente: Casi nula correlación lineal con la radiación. Negativamente relacionado con temperatura interior, temperatura exterior y temperatura de suelo. Tiene sentido físico si pensamos que durante la noche los consumos son mayores debido a la calefacción, mientras que los valores de las temperaturas son menores. Con la temperatura interior, exterior y de suelo lo hace con

coeficientes parecidos, si bien parece estar más correlacionado con la temperatura exterior. Esto sugiere que podríamos dejar solo la temperatura exterior. Esta poco relacionado con el viento y la humedad exterior. Se estudiara si es adecuado despreciarlas.

- Radiación. Según los coeficientes: escasa relación lineal con el consumo. Positivamente relacionada con la temperatura interior, temperatura exterior, y temperatura de suelo. Negativamente relacionada con la humedad interior. El aire tiene más capacidad de humectarse cuanto más alta es su temperatura, y por tanto más difícil alcanzar altos valores de humedad relativa, por lo que físicamente tiene sentido asumiendo las correlaciones positivas entre temperatura y radiación.
- Temperatura exterior. Según los coeficientes esta: positivamente fuertemente relacionada con la radiación, temperatura interior, y temperatura de suelo. Esto tiene sentido físico y algunas de estas variables son por tanto candidatas a ser eliminadas. Como se comento anteriormente también guarda una fuerte relación lineal con la humedad interior.
- Temperatura interior y temperatura de suelo. Ambas están fuertemente relacionadas con la temperatura exterior. Además los coeficientes de las tres temperaturas con el resto de variables, son del mismo signo y valor muy parecido, lo que sugiere que están midiendo el mismo fenómeno físico. Esto tiene bastante sentido, si pensamos que la temperatura esta fuertemente relacionada con la radiación, y que debido a la permeabilidad y poca consistencia del invernadero su constante de tiempo es rápida teniendo poco desfase con las condiciones exteriores. Por otra parte la temperatura del suelo esta medida a 3cm por lo que también puede estar bastante relacionada con la temperatura exterior.
- Humedad exterior: tiene una escasa relación lineal con el consumo. Esta fuertemente relacionada con la humedad interior y la radiación.
- Humedad interior: posee una mayor correlación con el consumo que la humedad exterior. Esta muy fuertemente relacionada con las temperaturas y la radiación. Tiene sentido que la humedad en el interior se vea más afectada por estos parámetros que la humedad exterior.

Existen por tanto tres variables con un coeficiente de correlación entre ellas superior al 0.9 que son candidatas a ser reducidas. Son las variables de temperatura. Además, también se estudiará si la humedad interior puede estar suficientemente correlacionada con las anteriores variables para ser eliminada. También si puede despreciarse la humedad exterior o el viento.

A partir de la matriz de coeficientes de correlación R^2 , pueden obtenerse los coeficientes de determinación. El coeficiente R al cuadrado, coeficiente de determinación, nos da la información compartida por ambas variables. Por tanto $1 - R^2$ nos da la información no común entre ambas. En la tabla [3.5](#) se presenta la matriz de coeficientes de determinación.

TABLA 3.5. Matriz de coeficientes de determinación

	y	Rad	Text	Tint	Tsuelo	Viento	HRext	HRint	tiempo
y	1	0.008	0.039	0.023	0.022	0.001	0.001	0.013	0.045
Rad	-	1	0.298	0.431	0.387	0.126	0.2747	0.496	0.001
Text	-	-	1	0.891	0.842	0.035	0.081	0.613	0.114
Tint	-	-	-	1	0.941	0.054	0.089	0.684	0.050
Tsuelo	-	-	-	-	1	0.051	0.064	0.641	0.051
Viento	-	-	-	-	-	1	0.110	0.158	0.009
HRext	-	-	-	-	-	-	1	0.346	0.023
HRint	-	-	-	-	-	-	-	1	0.019
tiempo	-	-	-	-	-	-	-	-	1

Atendiendo a los coeficientes de determinación podemos añadir al análisis anterior:

- Temp.int y Temp.ext: comparten el 89,1 % de la información
- Temp.int y Temp.suelo: comparten el 94.1 % de la información
- Temp.suelo y Temp.ext: comparten el 84.2 % de la información
- Consumo: comparte un 0.1 % de la información con el viento y la humedad exterior.

3.2.2. Análisis de componentes principales

Se va a emplear este análisis para determinar la viabilidad de eliminar algunas de las variables anteriormente planteadas.

En primer lugar se realiza el análisis con las 8 variables a ser estudiadas. En la tabla 3.6 se muestra la variabilidad que aporta cada nueva componente ortogonal al total de la varianza de los datos y el tanto por ciento explicado por cada componente respecto al total.

TABLA 3.6. Variabilidad aportada por cada componente

	λ	% explicado
1º Componente	4.520	56.50
2º Componente	1.477	18.46
3º Componente	0.757	9.46
4º Componente	0.651	8.14
5º Componente	0.373	4.66
6º Componente	0.135	1.70
7º Componente	0.065	0.81
8º Componente	0.022	0.28

Con las cuatro primera componentes se explica el 90 % de la variación de los datos. Por lo que se prestara especial atención a estas componentes que explican la mayor parte de la varianza en los datos. Por último, el bajo peso de las dos últimas componentes principales indica que dos de las variables originales puede ser eliminadas. Gráficamente puede verse este resultado en la figura 3.45.

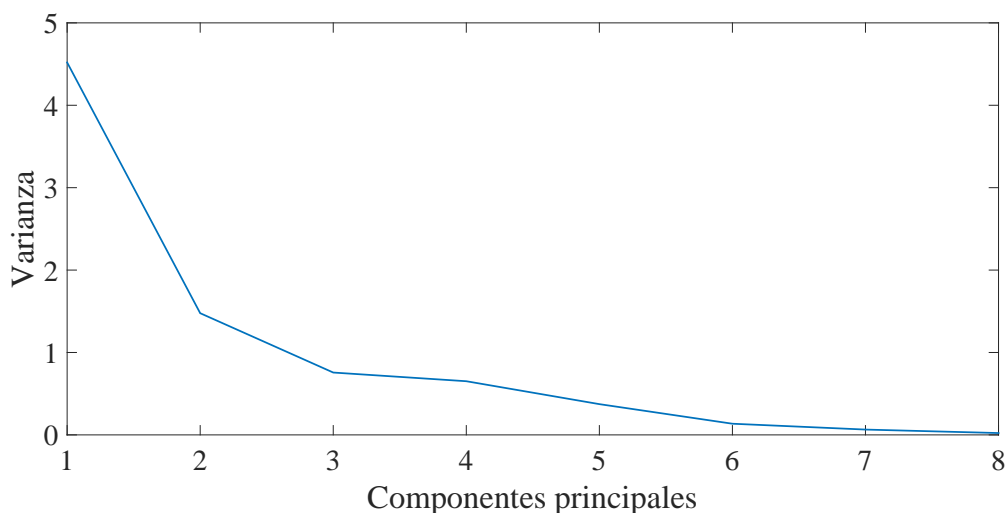


Fig. 3.45. Varianza explicada por cada componente

En la tabla 3.7 se muestran los coeficientes de correlación entre las variables originales y los nuevos ejes ortogonales.

TABLA 3.7. Matriz de coeficientes de correlación entre las variables originales y las nuevas componentes ortogonales

	1° Com.	2° Com.	3° Com.	4° Com.	5° Com.	6° Com.	7° Com.	8° Com.
Rad	0.368	-0.241	-0.124	-0.074	0.873	0.072	0.116	-0.068
Text	0.426	0.260	-0.064	0.077	-0.254	0.341	0.690	-0.290
Tint	0.445	0.179	-0.139	0.185	-0.073	0.141	-0.126	0.822
Tsuelo	0.434	0.200	-0.134	0.238	-0.085	0.104	-0.668	-0.483
Viento	0.187	-0.438	0.769	0.408	-0.064	0.110	0.001	0.008
HRext	-0.245	0.499	0.049	0.687	0.310	-0.310	0.151	-0.019
HRint	-0.439	0.093	-0.011	0.118	0.168	0.860	-0.126	0.036
tiempo	0.096	0.594	0.591	-0.495	0.180	0.009	-0.098	0.029

La representación gráfica de estos valores para cada componente permite una mejor visión y análisis de las relaciones existentes. En la siguientes gráficas se muestran en los ejes x e y las componentes principales, cuyos valores representan los valores de correlación lineal con las variables originales representadas en el plano xy. En la primera figura se observa la correlación que guarda cada variable original con la primera y segunda componente.

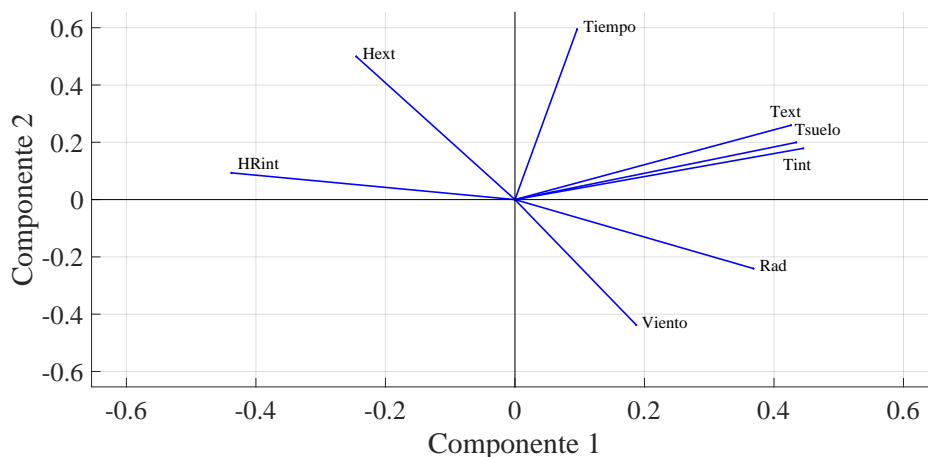


Fig. 3.46. Correlación de cada variable original con las componentes principales 1 y 2

Se observa como como la componente principal esta fuertemente y positivamente relacionada con las Temperaturas y la radiación. Lo hace negativamente con la humedad exterior e interior. Puede ser una variable imaginaria relacionada con las temperaturas.

La segunda componente esta influenciada en orden decreciente por: tiempo, humedad exterior , viento y radiación y temperaturas. Puede ser una variable relacionada con el momento del día.

Se observa como las tres variables relacionadas con las temperaturas, T_{int} , T_{ext} y T_{suelo} ocupan un lugar muy parecido en el plano, por lo que se cumplen las condiciones para eliminar alguna de las tres variables. La humedad exterior y el viento tienen ambas coeficientes altos con ambas componentes por lo que en principio se complica despreciarlas. Puede ser factible pensar que la humedad interior puede ser suplida por las temperaturas, debido a que ocupan lugares simétricos debido a su correlación negativa con la componente 1. Estas dos componentes representan el 74.96 % de la información. Se procede a analizar el resto de componentes pero teniendo en cuenta su menor peso.

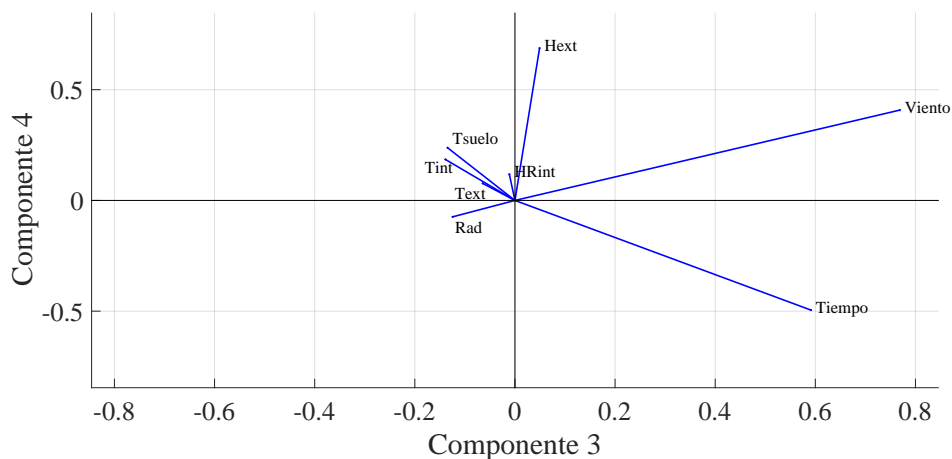


Fig. 3.47. Correlación de cada variable original con las componentes principales 3 y 4

La tercera componente está muy relacionada con el viento. Y muy poco con las temperaturas que tienen coeficientes muy parecidos. Puede ser que esta variable intente explicar el viento. La cuarta está muy relacionada con la humedad exterior y el tiempo. Por tanto puede estar describiendo la humedad. En esta componente los coeficientes de temperatura exterior, interior y suelo difieren más. Sin embargo estamos hablando de la cuarta componente por lo que no nos impiden despreciar alguna de ellas.

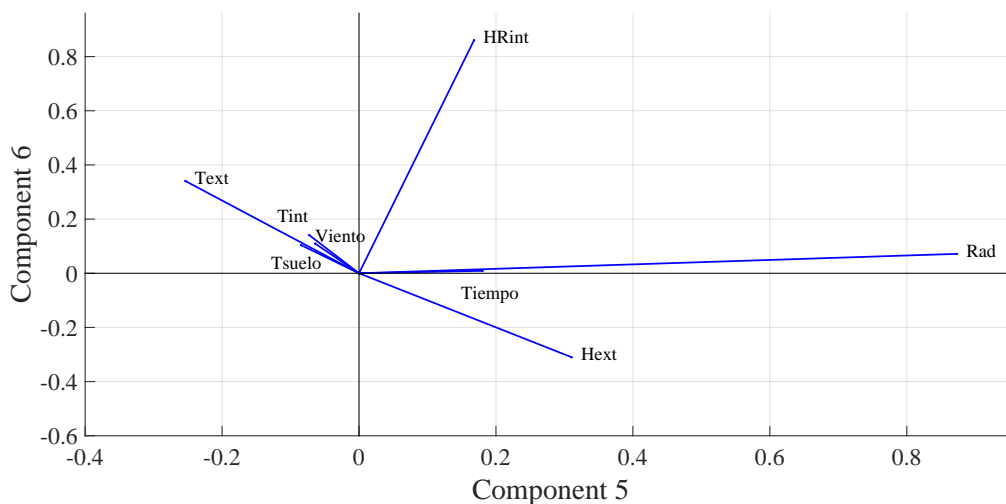


Fig. 3.48. Correlación de cada variable original con las componentes principales 5 y 6

En la quinta y sexta componente la temperatura interior, temperatura exterior y temperatura suelo vuelven a ocupar un sitio muy parecido en el plano. Por tanto sin analizar el resto de componentes, puesto que ya estamos por encima del 90 % de la información explicada, se decide eliminar dos de las tres variables: temperatura interior, temperatura exterior y temperatura suelo. Y se volverá a realizar el análisis de las componentes principales por si fuera conveniente algún reajuste más.

En cuanto a que variable independiente de las tres se debe retener, la respuesta no es inmediata. En el sentido práctico la temperatura exterior es una variable medida con mucha frecuencia y que puede estar disponible, lo cual es conveniente si queremos disponer de los datos que alimentan al predictor. Desde el punto de vista del modelo, lo mejor sería retener aquella variable que aporta más información. Para conocer esto, se recurre a la correlación parcial. En la sección 3.2.3 se realiza dicho análisis, el cual arroja que la variable a retener debe ser la temperatura exterior. Se elimina por tanto y se vuelve a realizar el análisis de componentes principales sin ella.

2º iteración Análisis de componentes principales

En la tabla 3.8 se muestra como queda ahora la variabilidad que aporta cada nueva componente ortogonal al total de la varianza de los datos y el tanto por ciento explicado por cada componente respecto al total

TABLA 3.8. Variabilidad aportada por cada componente

	λ	% explicado
1º Componente	2.959	49.315
2º Componente	1.307	21.782
3º Componente	0.732	12.202
4º Componente	0.515	8.582
5º Componente	0.358	5.969
6º Componente	0.129	2.150

El reparto es ahora más proporcional, y se han eliminado la 7º y 8º componentes que explicaban menos del 1 % de la varianza total.

Gráficamente puede verse este resultado en la figura [3.49](#)

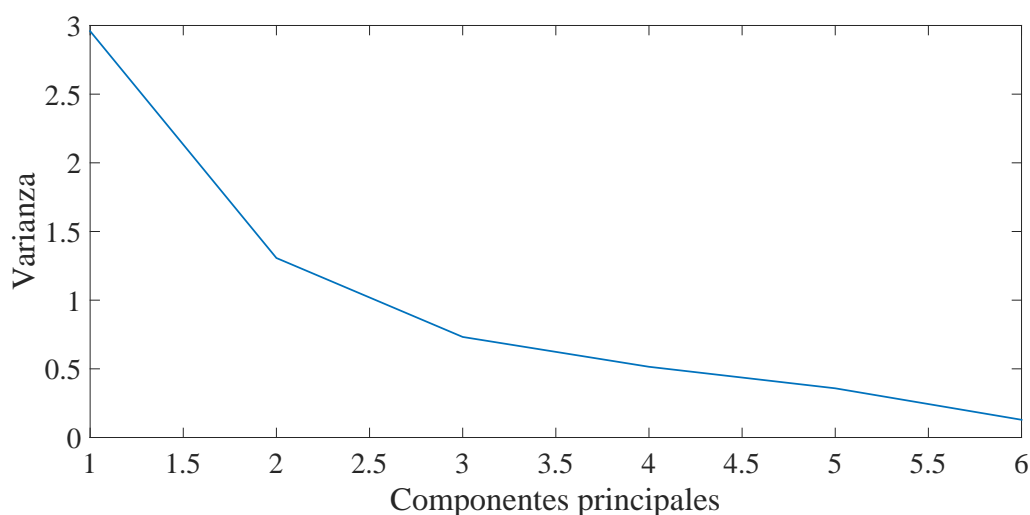


Fig. 3.49. Varianza explicada por cada componente

En la tabla [3.9](#) se muestran los coeficientes de correlación entre las variables originales y los nuevos ejes ortogonales.

TABLA 3.9. Matriz de coeficientes de correlación entre las variables originales y las nuevas componentes ortogonales de la 2º iteración del análisis de componentes

	1º Com.	2º Com.	3º Com.	4º Com.	5º Com.	6º Com.
Rad	-0.485	-0.0899	-0.182	0.265	0.802	0.095
Text	-0.452	0.407	-0.101	0.352	-0.433	0.556
Viento	-0.307	-0.336	0.883	0.031	-0.045	0.102
HRext	0.406	0.338	0.278	0.753	0.127	-0.244
HRint	0.544	-0.094	0.079	-0.073	0.268	0.782
tiempo	-0.064	0.768	0.306	-0.482	0.280	-0.035

La representación gráfica de estos valores para cada componente permite una mejor visión y análisis de las relaciones existentes. En la figura 3.50 se observa la correlación que guarda cada variable original con la primera y segunda componente

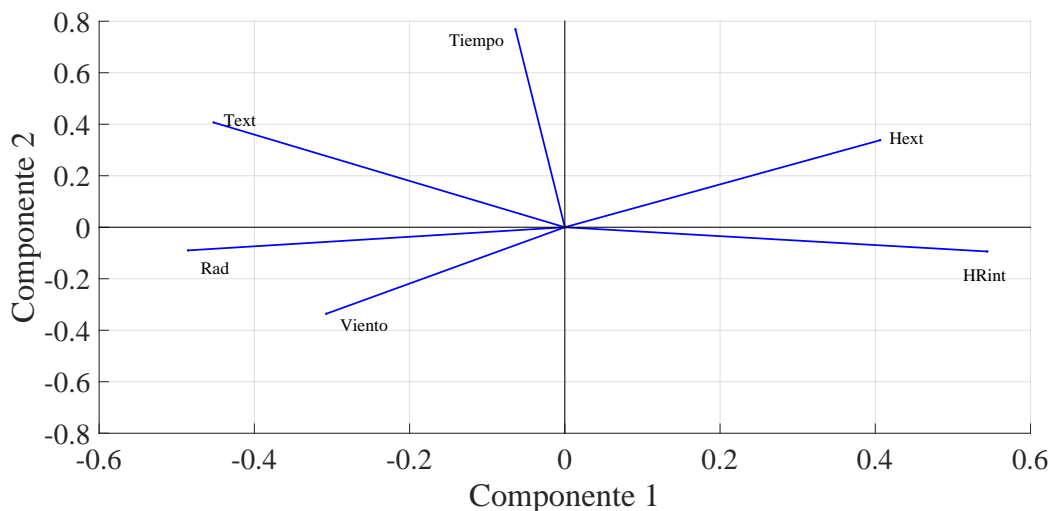


Fig. 3.50. Correlación de cada variable original con las componentes principales 1 y 2

Se observa que tanto la humedad exterior como el viento están relacionados con la 1ª y 2ª componente. Por lo que no se va a eliminar ninguna de estas variables. También se va a conservar la humedad interior que está fuertemente relacionada con la primera componente. Puesto que cada variable ocupa un lugar diferente en el plano, sin que exista relación entre ellas, y que la primera y segunda componente explican el 70% de la varianza de los datos, se da por finalizado el análisis.

3.2.3. Correlación parcial

La correlación parcial se define como la correlación entre dos variables si las demás variables no varían, es decir, el valor de las demás variables son fijos. Esto en la práctica permite analizar la correlación entre la variable a predecir, consumo, con una de las variables independientes, controlando el resto de las variables. Con esto nos aseguramos que la correlación calculada se debe únicamente a variaciones en la variable independiente y no una tercera. Con esto podemos estimar con qué variable de T interior, T exterior o T suelo aporta más información y por tanto con cuál quedarnos.

Puesto que en nuestro modelo vamos a seguir usando el resto de variables y se pretende eliminar dos, vamos a ver cuánta variabilidad aporta cada una de las variables candidatas con el resto controladas. Usando la función de Matlab $[\text{rho}, \text{pval}] = \text{partialcorri}(y, x)$ nos permite evaluar la correlación entre y , con cada variable de x controlando el resto de las variables contenidas en x . De esta forma en la tabla 3.2.3 se muestra la varianza que aporta T_{ext} cuando el resto de variables se mantiene constante y sin tener en cuenta T_{int} ni T_{suelo} . Esto mismo también puede observarse para T_{int} y T_{suelo} .

TABLA 3.10. Coeficientes lineales de correlación parcial entre y y el resto de variables

	T_{ext}	T_{int}	T_{suelo}
Rad	-0.060	-0.104	-0.061
Text	0.115	-	-
Tint	-	-0.005	-
Tsuelo	-	-	-0.050
Viento	0.038	-0.040	-0.047
HRext	-0.034	-0.035	-0.024
HRint	-0.064	-0.050	-0.065
Tiempo	-0.017	-0.045	-0.035

Atendiendo a los coeficientes de la tabla 3.2.3, nos quedamos con la variable T_{ext} que aporta una varianza de 0.115 mientras T_{int} y T_{suelo} un 0.005 y 0.05 respectivamente.

3.3. Modelado empírico lineal

Puesto que no se tienen datos de funcionamiento mediante consignas de todos los equipos, el modelado se centrará en los equipos que si los tienen y en aquellos periodos que estén disponibles. En la figura 3.51 se muestra la selección de datos de para la estimación y validación de los modelos, pertenecientes al periodo de la semana 1 a la 12. Se ha intentado que estos datos sean de la mayor calidad posible, evitando incluir falsas medidas o ruido en la señales.

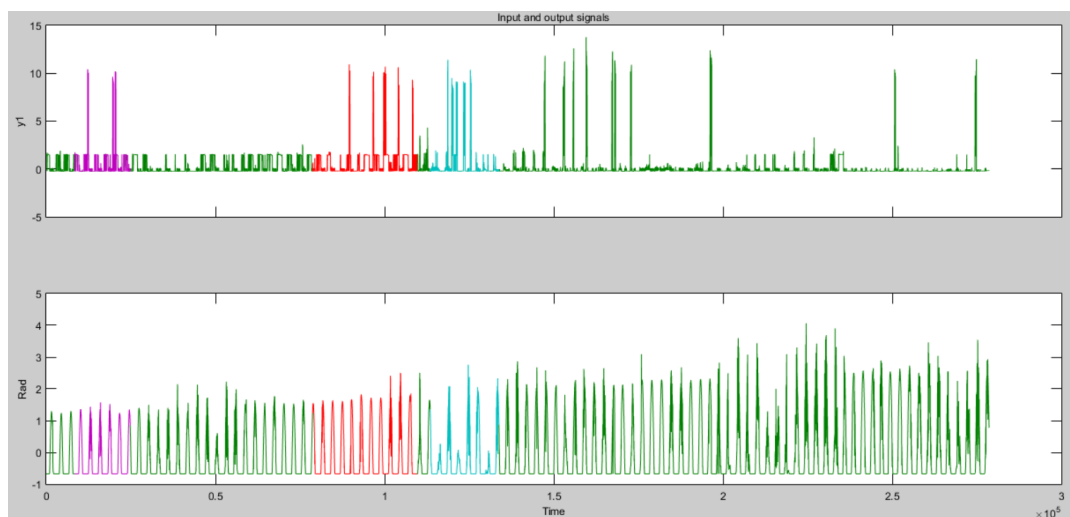


Fig. 3.51. Conjunto de datos seleccionados para la estimación (naranja) y para la validación cruzada 1 (amarillo) y validación cruzada 2 (rojo)

En este periodo funcionan de manera automática la ventilación y la calefacción. Como se comento anteriormente no lo hace la máquina deshumificadora, pero se ha supuesto su señal de activación cuando el consumo de la línea invernadero esta por encima de 7 kW. En la siguientes

figuras puede verse algunos pares de variables e-s para los conjunto de datos usados para la estimación y validación. Estas gráficas han sido obtenidas de la herramienta de identificación de Matlab que permite representar la salida conjuntamente con cada una de las entradas.

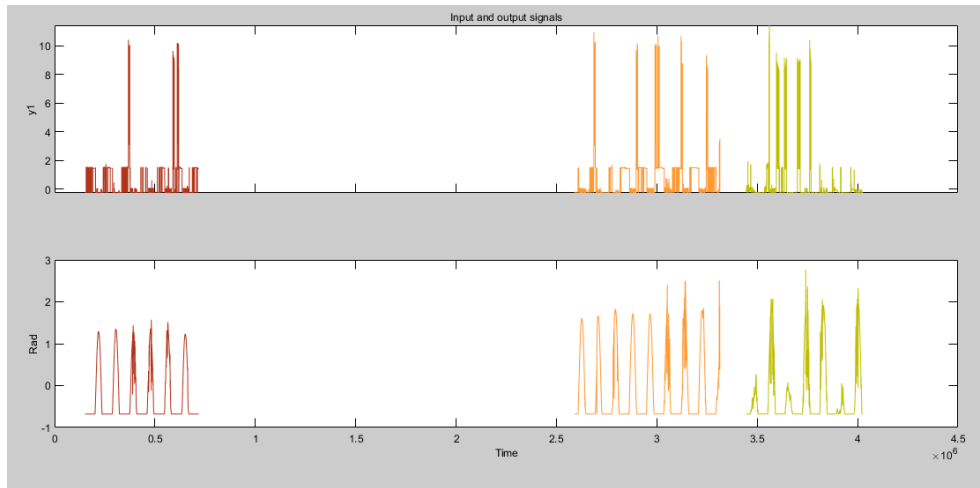


Fig. 3.52. Conjunto de datos para la estimación (rojo), validación cruzada 1 (azul) y validación cruzada 2 (morado). Par e-s, radiación-consumo

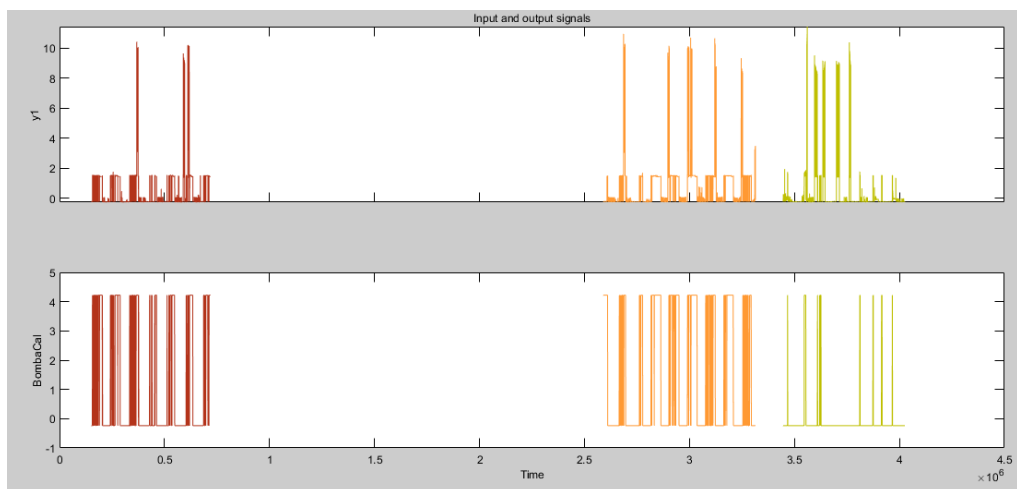


Fig. 3.53. Conjunto de datos para la estimación (naranja), validación cruzada 1 (amarillo) y validación cruzada 2 (rojo). Par e-s, bomba calefacción-consumo

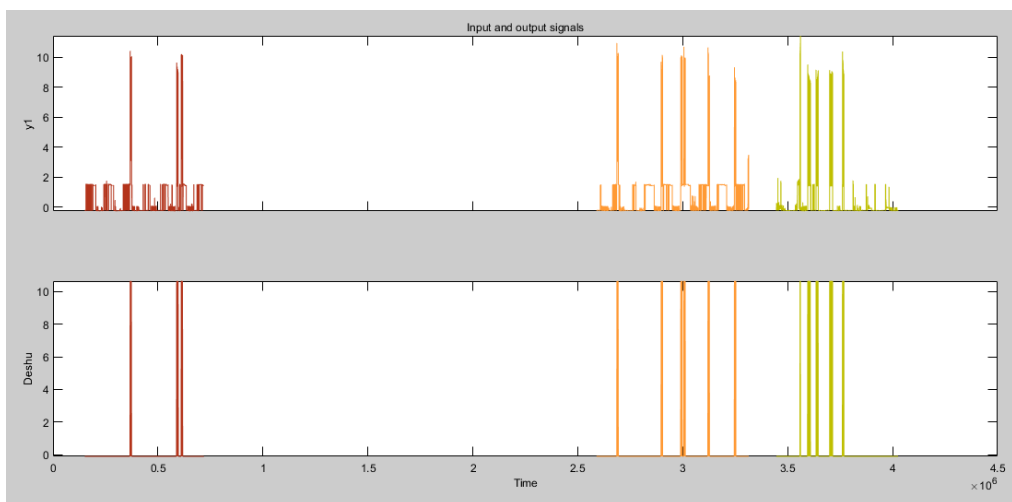


Fig. 3.54. Conjunto de datos para la estimación (naranja), validación cruzada 1 (amarillo) y validación cruzada 2 (rojo). Par e-s, Deshumificadora -consumo

Modelo ARX basado en estado de actuadores.

Se comienza con la obtención de modelos lineales Arx, usando como entradas las variables climáticas preseleccionadas en el apartado 2.4 junto con los estados de los actuadores. En la tabla 3.11, se muestra las variables de entrada tenidas en cuenta en la construcción del modelo.

TABLA 3.11. Variables usadas en la construcción de los modelos Arx basados en estados de actuadores

Variable	Entrada
Radiación	1
Temperatura exterior	2
Humedad interior	3
Humedad exterior	4
Velocidad del viento	5
Tiempo	6
Señal bomba calefacción	7
Apertura Ven.Cen	8
Apertura Ven.Lat	9
Señal deshumificadora	10

Como se observa el número de entradas del modelo es alto. En consecuencia la estructura óptima, en cuanto al número de valores pasados de cada entrada, ha de ser hallada entre una multitud de combinaciones posibles. Gracias a su gran capacidad de cálculo, Matlab permite obtener la estructura más óptima evaluando diferentes configuraciones. Sin embargo, esta capacidad no es infinita, por lo que se realizan varias iteraciones con el siguiente procedimiento:

- En el primer paso se establece un rango con ordenes bajos para los regresores y se selec-

cionan las estructuras más óptimas.

- En la segunda iteración se deja mayor grado de libertad a aquellos regresores que han alcanzado el límite establecido para su orden.

En total se han probado 8.388.608 modelos en la primera iteración y 2.752.512 en la segunda. En la tabla 3.12 se muestran las estructuras probadas y aquellas seleccionadas como más óptimas. Por otra parte en la tabla 3.13 se muestra el ajuste en la predicción y el error en función de los criterios para las estructuras seleccionadas.

TABLA 3.12. Combinaciones de estructuras probadas para el modelo Arx basado en actuadores y selección de las más óptimas

	na	nb	nk
1° Iter.	1:8	1:2	0:1
Arx721	7	2 2 1 1 2 1 2 2 2 2	1 0 0 0 1 1 0 0 0 0
Arx111	1	1 2 1 1 2 1 2 2 1 2	1 0 1 0 1 1 0 0 0 0
2° Iter.	1:7	1:4 1:4 1 1 1:4 1 1:4 1:4 1:4 1:4	0:2 0 0:1 0 0:2 0:2 0 0 0 0
Arx712	7	1 2 1 1 1 1 4 1 2 3	2 0 0 0 2 2 0 0 0 0
Arx111	1	1 2 1 1 1 1 2 1 2 2	1 0 0 0 2 2 0 0 0 0
3° Iter.	1:7	1 2 1 1 1 1 1:8 1 2 1:4	0:4 0 0 0 0:4 0:4 0 0 0 0
Arx 712	7	1 2 1 1 1 1 4 1 2 4	2 0 0 0 0 3 0 0 0 0
ARx112	1	1 2 1 1 1 1 2 1 2 2	2 0 0 0 4 4 0 0 0 0

TABLA 3.13. Ajustes y error de las estructuras seleccionadas para los modelos Arx basados en actuadores

	Ajuste estimacion(%)	FPE	Val. cruzada(%)	Val. cruzada2(%)
1° Iter.				
Arx721	80.98	0.02439	84.53	86.69
Arx111	81.42	0.02472	85.14	87.14
2° Iter.				
Arx721	81.07	0.02181	84.34	86.31
Arx111	81.4	0.02472	85.14	87.15
3° Iter.				
Arx712	80.35	0.01954	82.28	84.58
ARX112	81.39	0.02472	85.23	87.14

En la 1° iteración se seleccionan dos estructuras óptimas. La estructura Arx721, posee orden

mayor, sin embargo dado a sus buenas prestaciones como se muestra en la tabla 3.13 se decide conservar. La segunda iteración se realiza entorno a estos dos estructuras escogidas.

La segunda iteración arroja como óptimas estructuras similares a las anteriores. Aunque el criterio FPE de akaike para la estructura Arx721 se ha reducido de forma considerable.

Se realiza una última iteración por tener mas información. Se vuelve a obtener como óptimas estructuras muy parecidas, aunque de mayor orden. Sin embargo para la estructura Arx712 tiene un peor ajuste en la validación cruzada, lo que puede estar indicando sobreparametrización.

De entre las 6 estructuras generadas como óptimas, finalmente se van a escoger 2, una con $na=1$ y la otra $na=7$. Se escogerán aquellas estructura con mejor ajuste y menor orden, que serán validadas con el análisis residual.

- Estructura con $na = 1$. Se preselecciona la estructura de la 2º iteración presenta el menor orden, 15 coeficientes respecto a los 16 del primero. El ajuste es muy similar, siendo menor para los datos de estimación y ligeramente mayor en la validación cruzada 2, por lo que podemos estar evitando sobreparametrización.
- Estructura con $na = 7$. Se preselecciona de la 1º iteración ya que presenta los mejores ajustes en la validación cruzada y menor en los datos de estimación, por lo que se estaría evitando sobreparametrización. En cuanto al orden, ambos presentan 24 coeficientes.

A continuación se muestran las gráficas de ajuste de las estructuras seleccionadas y el análisis marginal.

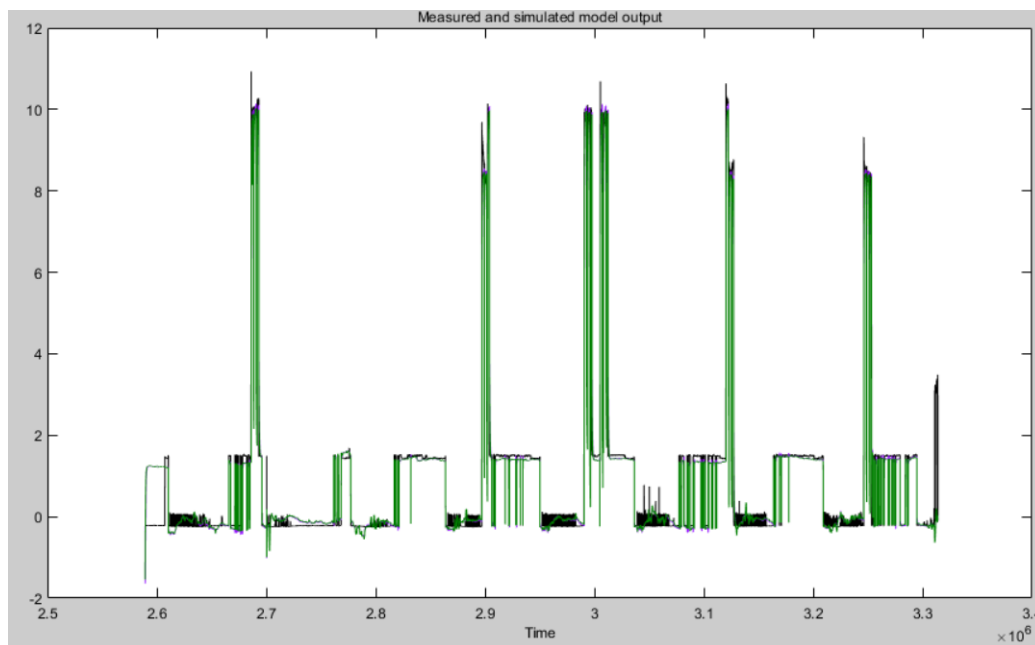


Fig. 3.55. Consumo real del invernadero (curva negra) y predicho por los modelos (Arx111 en verde y Arx721 en morado) basados en el estado de actuadores, para el conjunto de datos de estimación

En la gráfica 3.55 se observa como el ajuste a los datos de estimación es bastante bueno. Como puede verse existen valores negativos del consumo. En este punto hay que aclarar que esto es debido a que todos los datos han sido normalizados. Esto consiste en que su varianza es la unidad y su media es cero. Para obtener una media igual a cero, los valores son trasladados entorno a cero, de forma que la suma de los valores positivos y negativos sea cero.

Se presenta en la figura 3.56 el ajuste de los modelos al conjunto de datos de validación.

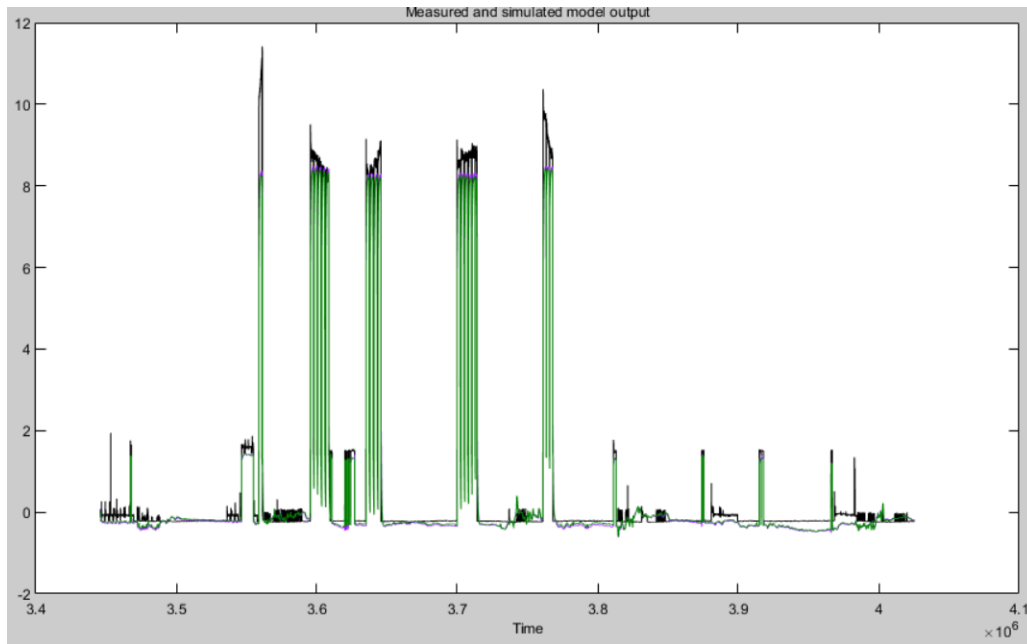


Fig. 3.56. Consumo real del invernadero (curva negra) y predicha por los modelos (Arx111 en verde y Arx721 en morado) basados en el estado de actuadores, para el conjunto de datos de validación

Se observa como la predicción es bastante buena. El modelo consigue reproducir tanto el consumo de la calefacción como de la deshumificadora con bastante éxito. El consumo de las ventanas también se estima bien aunque con mayor dificultades, debido a que su funcionamiento es de forma discontinua.

En la figura 3.57 se presenta gráficamente el ajuste de los modelos al conjunto de datos de la validación 2.

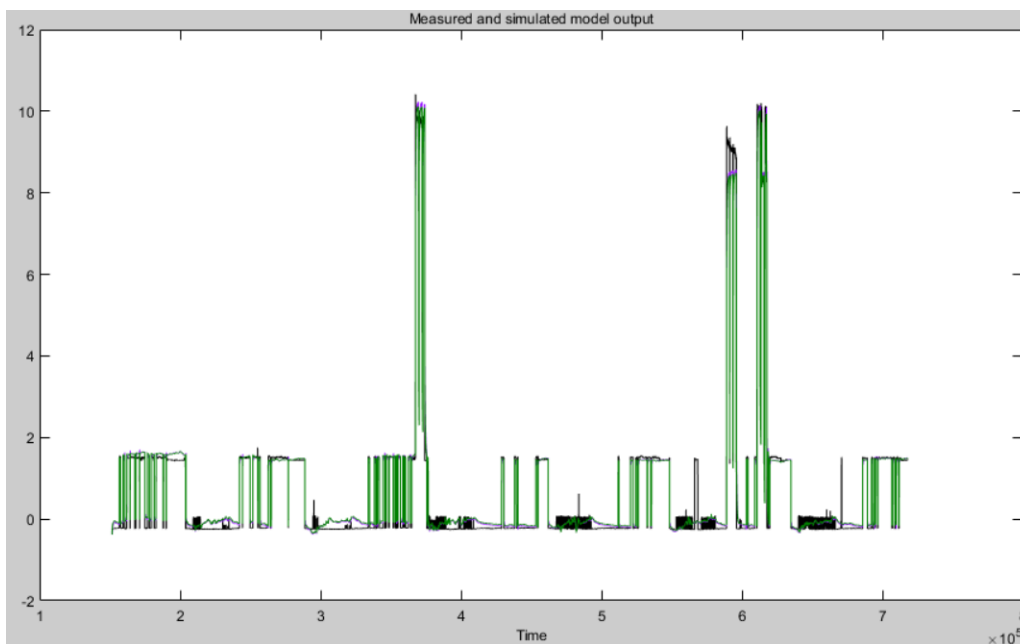


Fig. 3.57. Consumo real del invernadero (curva negra) y predicho por los modelos (Arx111 en verde y Arx721 en morado) basados en el estado de actuadores, para el conjunto de datos de validación 2

La predicción de los datos de validación 2, también son bastante buenos. Esto ocurre incluso cuando estos datos están bastante alejados en el tiempo de los datos que han sido usados en la estimación. Esta característica del modelo es muy positiva, y se debe fundamentalmente a que se están usando como entradas los estados de los actuadores, lo cual determina directamente el consumo.

A continuación se muestra el análisis de residuos para dar validez al modelo.

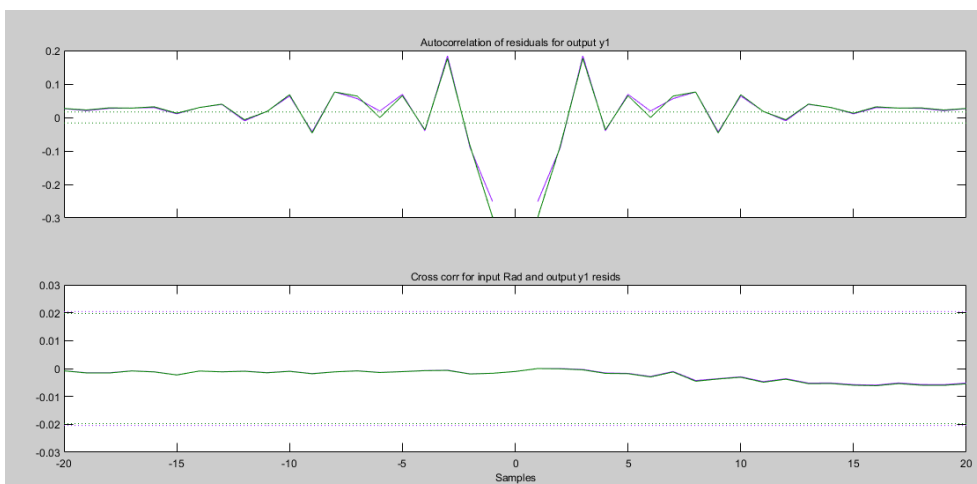


Fig. 3.58. Análisis residual. En la parte superior la autocorrelación. En la parte inferior la correlación cruzada entre radiación y el consumo

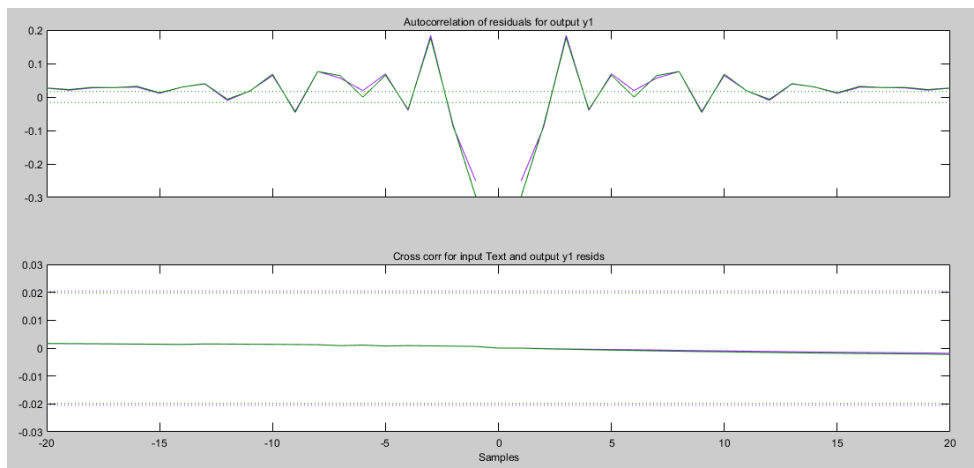


Fig. 3.59. En la parte superior la autocorrelación. En la parte inferior la correlación cruzada entre temperatura exterior y el consumo

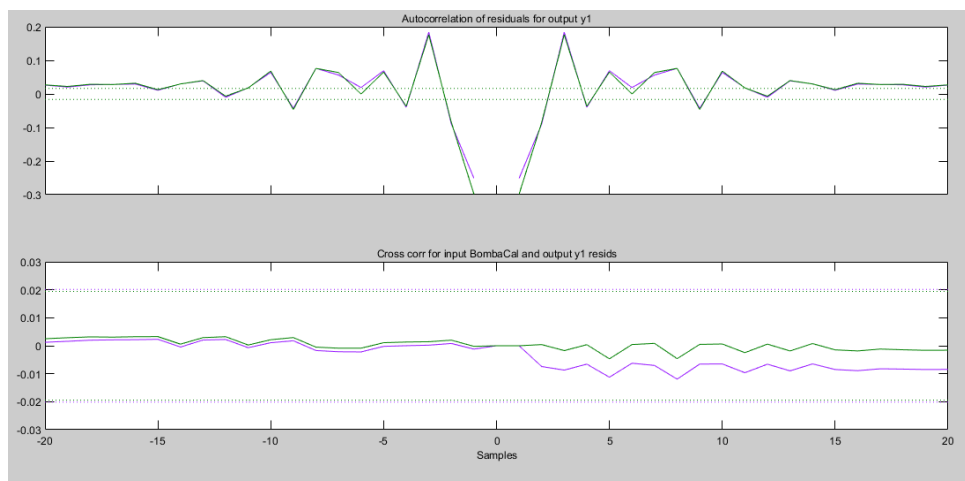


Fig. 3.60. En la parte superior la autocorrelación. En la parte inferior la correlación cruzada entre la señal de la bomba de calefacción y el consumo

Se observa claramente que la autocorrelación no se encuentran en los rangos aceptables, existiendo una relación entre el error y el valor pasado de la salida. Sin embargo, en casos donde el modelo está dado mayormente en función de las perturbaciones y donde se presenta un ajuste bueno, este tipo de análisis no es pertinente.

El análisis importante en este tipo e casos es el de la correlación cruzada. Se observa como la correlación entre el error y las entradas es prácticamente cero, con intervalo de confianza que se estableció en el 98 %. Para el resto de entradas no representadas el análisis residual es muy similar.

Modelo ARX basado en consignas

En este apartado se intentara sustituir las entradas de los estados de los actuadores por sus correspondientes consignas. De esta forma se obtendrá una herramienta más útil, ya que no se

necesita saber el estado de los actuadores, si no las consignas programadas en estos, así como las variables meteorológicas externas. Esto es posible hacerlo para la calefacción y la ventilación. Sin embargo, puesto que la deshumificadora no ha funcionado de forma programada, no existe una consigna y se ha mantenido el estado del actuador como entrada en el modelo de predicción. En la tabla 3.14 se muestran las variables de entrada usadas en la construcción de los modelos Arx basados en consignas. Se observa como se ha reducido en uno el número de variables al ser la consigna de ventilación común para las ventanas cenitales y laterales.

TABLA 3.14. Variables usadas en la construcción de los modelos Arx basados en consignas

Variable	Entrada
Radiación	1
Temperatura exterior	2
Humedad interior	3
Humedad exterior	4
Velocidad del viento	5
Tiempo	6
Consigna bomba calefacción	7
Consigna Ventilación	8
Señal deshumificadora	9

Para estimar la mejor estructura se sigue el mismo procedimiento que el desarrollado en el apartado anterior. En la primera iteración se prueban 2.097.152 estructuras y en la segunda 82.944. En la tabla 3.3 se muestran las estructuras seleccionadas como las mejores y sus ajustes y errores cometidos en la predicción. Puesto que estos modelos están basados en consignas se le ha añadido la letra 'C' a su nombre para diferenciarlos de los modelos basados en actuadores.

TABLA 3.15. Mejores estructuras generadas para los modelos Arx basados en consignas, sus ajustes y errores de predicción

	Nº de coeficientes	na	nb	nk
ArxC111	12	1	1 2 1 1 1 1 1 1 2	1 1 0 1 0 1 0 0 0
ArxC510	17	5	1 1 1 1 1 1 1 1 4	0 0 1 0 1 0 0 0 0
	Ajuste estimacion	FPE	Val. cruzada	Val. cruzada2
ArxC111	63.96	0.03781	74.36	67.23
ArxC510	62.7	0.03336	72.08	65.10

Como era de esperar los resultados obtenidos son peores que con los modelos basados en los estados de los actuadores. En las siguientes gráficas se muestra el ajuste a los datos de estimación y validación

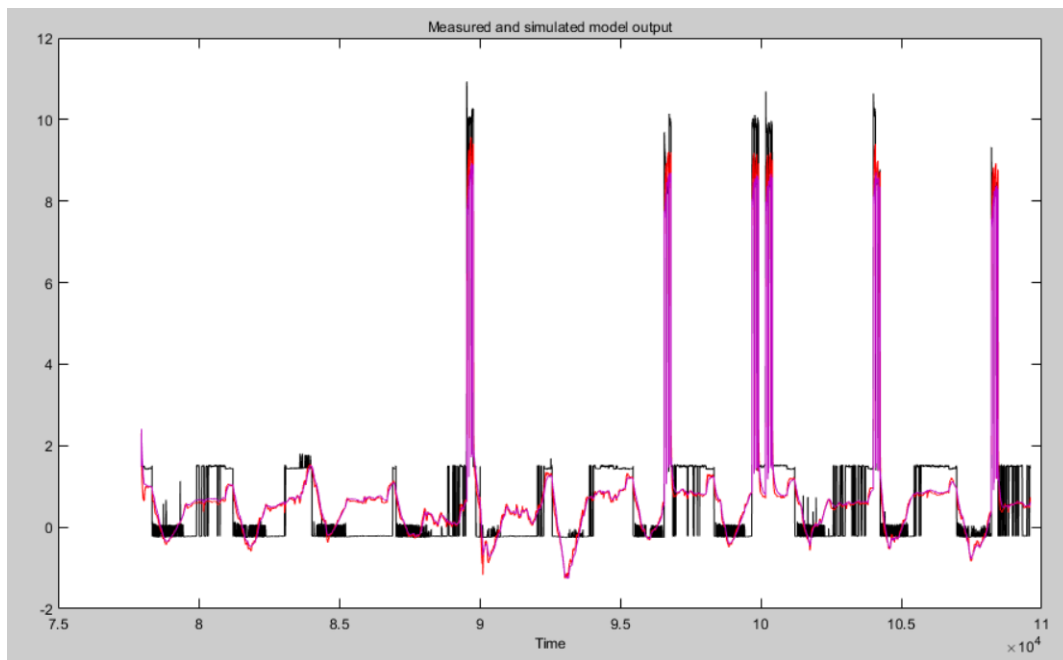


Fig. 3.61. Consumo real del invernadero (curva negra) y predicho por los modelos (ArxC111 en rojo y ArxC510 en morado) basados en consignas, para el conjunto de datos de estimación

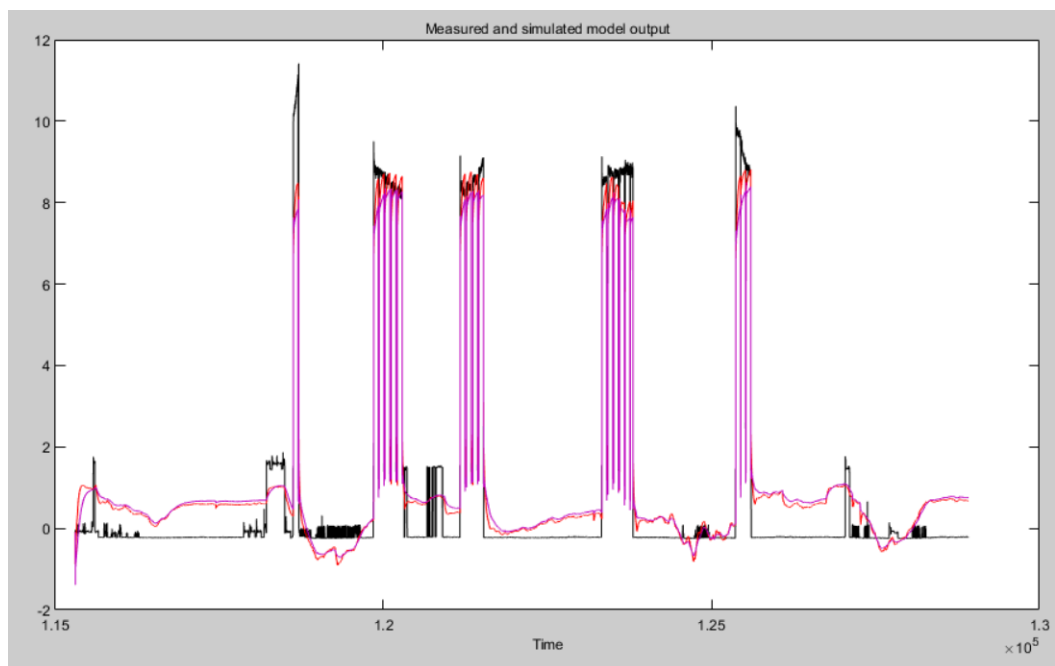


Fig. 3.62. Consumo real del invernadero (curva negra) y predicho por los modelos (ArxC111 en rojo y ArxC510 en morado) basados en consignas, para el conjunto de datos de validación

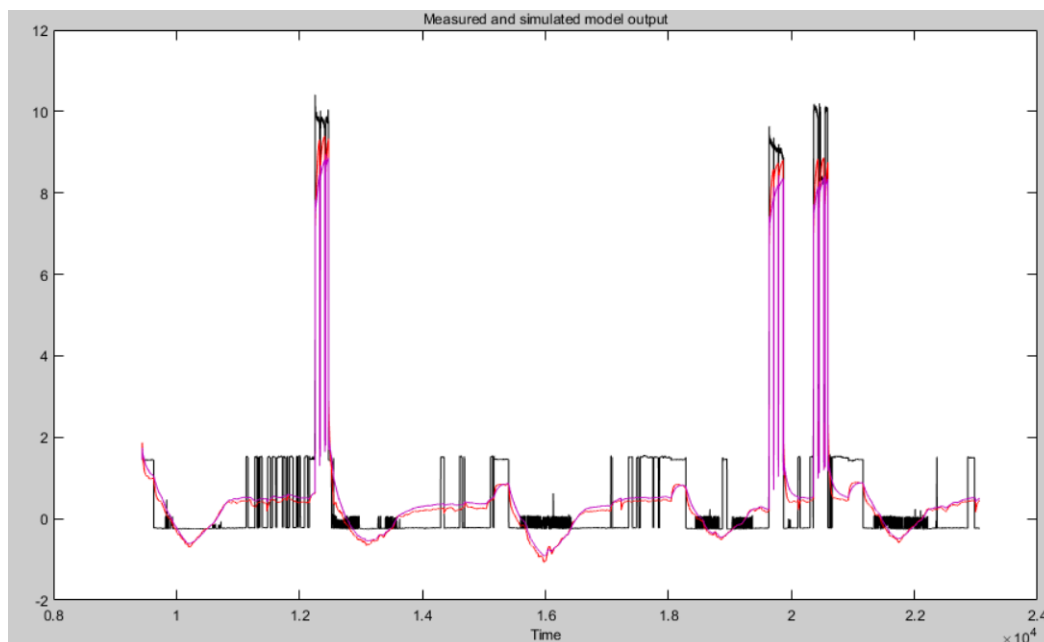


Fig. 3.63. Consumo real del invernadero (curva negra) y predicho por los modelos (ArxC111 en rojo y ArxC510 en morado) basados en consignas, para el conjunto de datos de validación 2

Se observa en las gráficas anteriores como los modelos no consiguen reproducir el consumo debido a la calefacción y ventilación. Además el modelo presenta en su predicción valores que caen por debajo del mínimo de consumo real, lo que podría producir estimaciones de consumo negativo. En cuanto a la comparación entre las dos estructuras generadas, ambos modelos tienen un comportamiento muy parecido, siendo las diferencias mínimas.

3.4. Modelado empírico no lineal

Se va a intentar mejorar los resultados obtenidos en el apartado 3.3, usando modelos no lineales que permitan usar las consignas como entradas obteniendo buenas predicciones.

Como anteriormente se dijo, para la estructura de la red neuronal se utilizará el perceptrón con una capa, y se comenzará con un número de nodos igual a diez. La cantidad de nodos adecuados se determinará mediante la validación cruzada. Esta decisión tiene sus motivos en trabajos previos realizados como [7] y [2].

Se parte de los dos modelos lineales basados en consignas ArxC111 y ArxC510 mostrados en la tabla 3.15 y se usarán el número de valores pasados de los regresores de estos modelos para las estructuras neuronales que se van a generar. La herramienta Identificación de sistemas de Matlab permite establecer el número de valores pasados de los regresores, así como el retardo puro. En la figura 3.64 se muestra una captura de la interfaz gráfica de la selección del número de nodos y valores pasados mediante la herramienta de Matlab.

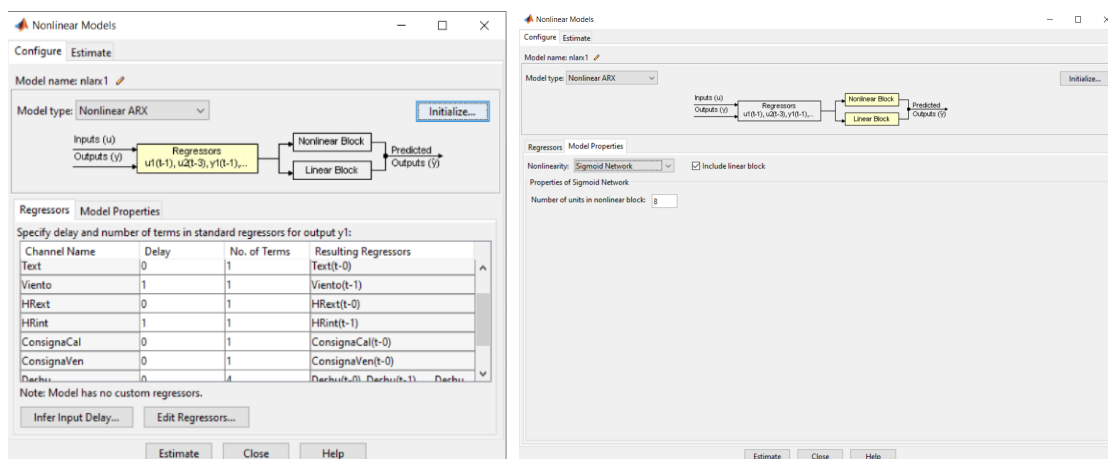


Fig. 3.64. Captura de la selección del número de nodos a la derecha y valores pasados para cada regresor a la izquierda.

Mediante la validación cruzada se obtienen las siguientes estructuras como las mejores:

- Modelo Narx111 con 10 nodos
- Modelo Narx510 con 12 nodos

El ajuste a los datos de predicción, validación uno y validación dos se muestra en las figuras 3.65, 3.66 y 3.67.

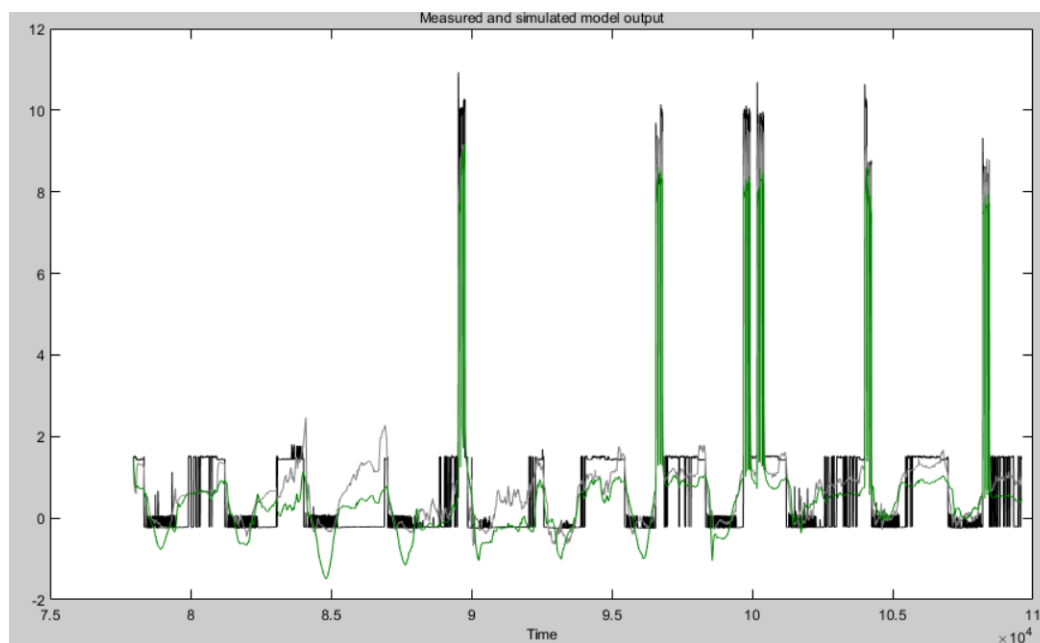


Fig. 3.65. Consumo real del invernadero (curva negra) y predicha por los modelos no lineales basados en consignas (narx111 en gris y narx 510 en verde), para el conjunto de datos de estimación

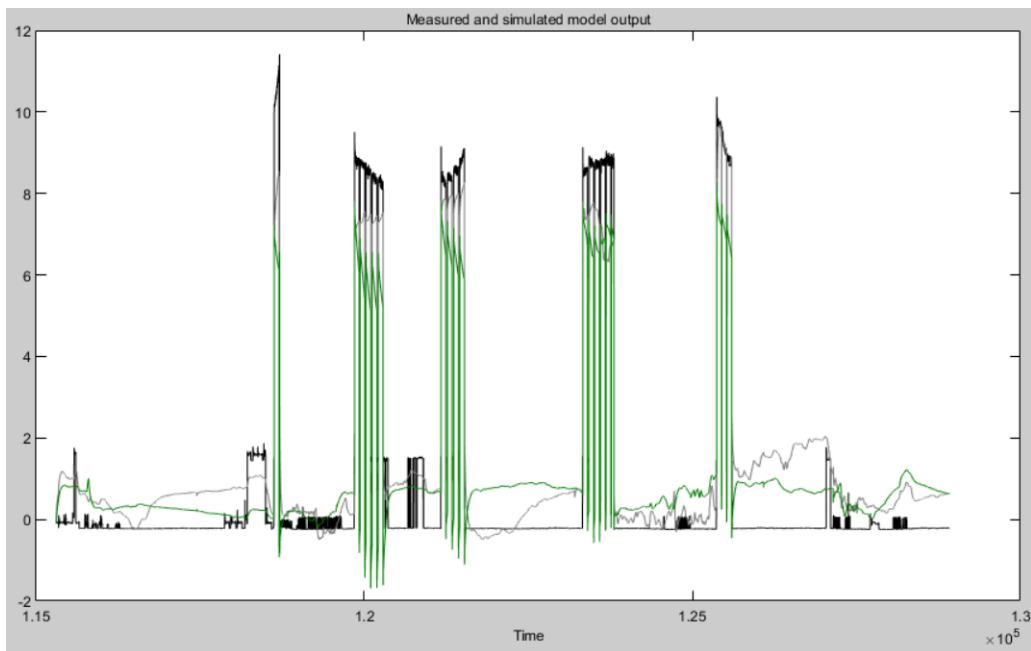


Fig. 3.66. Consumo real del invernadero (curva negra) y predicha por los modelos no lineales basados en consignas (narx111 en gris y narx 510 en verde), para el conjunto de datos de validación

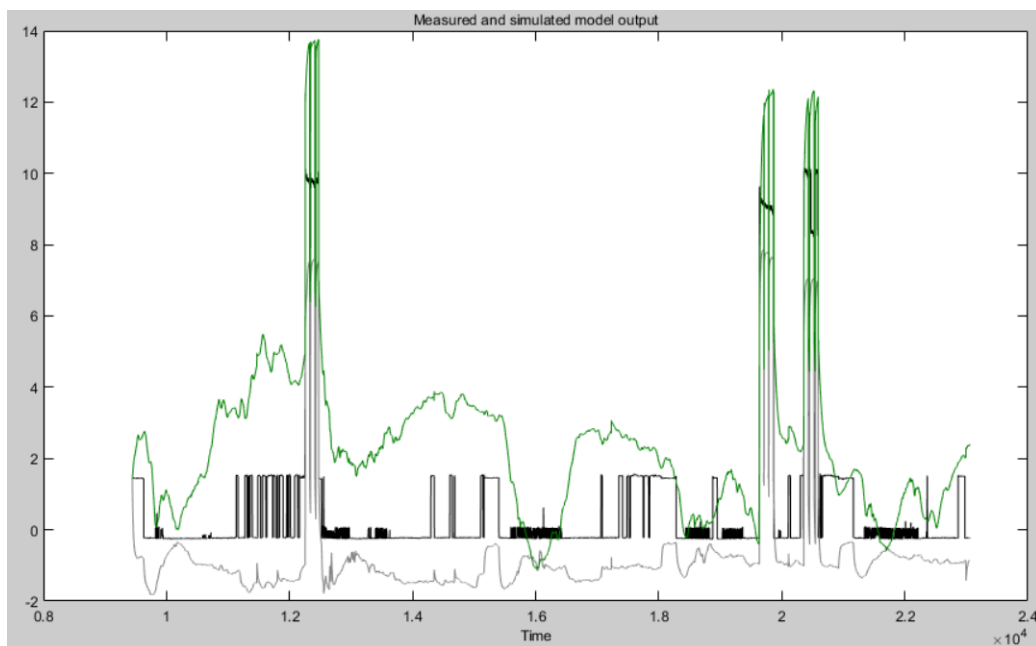


Fig. 3.67. Consumo real del invernadero (curva negra) y predicha por los modelos no lineales basados en consignas (narx111 en gris y narx 510 en verde), para el conjunto de datos de validación 2

Los resultados obtenidos no son demasiados satisfactorios. Los modelos no son capaces de estimar el consumo debido a la calefacción y ventilación. En la figura [3.65](#) el modelo narx111, presenta un mejor ajuste, evitando la estimación de picos negativos de consumo. Por otro lado el mal ajuste con la validación 2 hace pensar que una variable esta afectando de forma negativa a la predicción, ya que en los modelos previos, a pesar de ser un conjunto de datos alejados en

el tiempo del conjunto de estimación, la predicción era aceptable.

Por ello se elimina la variable tiempo ya que se piensa que es la causante de impedir la correcta predicción y se vuelven a generar las RNA a partir de los modelos lineales ArxC111 y ArxC510 pero eliminando la entrada Tiempo. Además como estos modelos ArxC habían sido obtenidos teniendo en cuenta la entrada Tiempo, se vuelve a calcular la estructura Arx basada en consignas mas óptima sin tener en cuenta la entrada tiempo. Se obtiene la nueva estructura $na = 3$, $nb = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 3]$ y $nk = [0\ 1\ 0\ 1\ 1\ 0\ 0\ 0]$. Se generan las estructuras neuronales a partir de los modelos lineales mencionados. Tras probar diferentes números de nodos y mediante la validación cruzada se obtienen las siguientes RNA como las mejores. Se les ha añadido la notación 'ST' para indicar que estan modeladas sin la variable tiempo.

- Modelo Narx111ST con 8 nodos.
- Modelo Narx510ST con 11 nodos.
- Modelo Narx310ST con 7 nodos.

En la tabla [3.16](#) se presentan los ajustes y los errores cometidos en la predicción.

TABLA 3.16. Ajustes y errores de predicción de los modelos Narx basados en consignas y sin tiempo generados

	Ajuste estimacion	FPE	Val. cruzada	Val. cruzada2
NarxC111ST	64.18	0.03748	76.89	67.23
Narx310ST	68.07	0.0331	76.08	64.21
NarxC510ST	47.58	0.03474	73.60	64.75

En las siguientes figuras se presenta el ajuste de cada modelo para cada conjunto de datos, estimación, validación y validación 2.

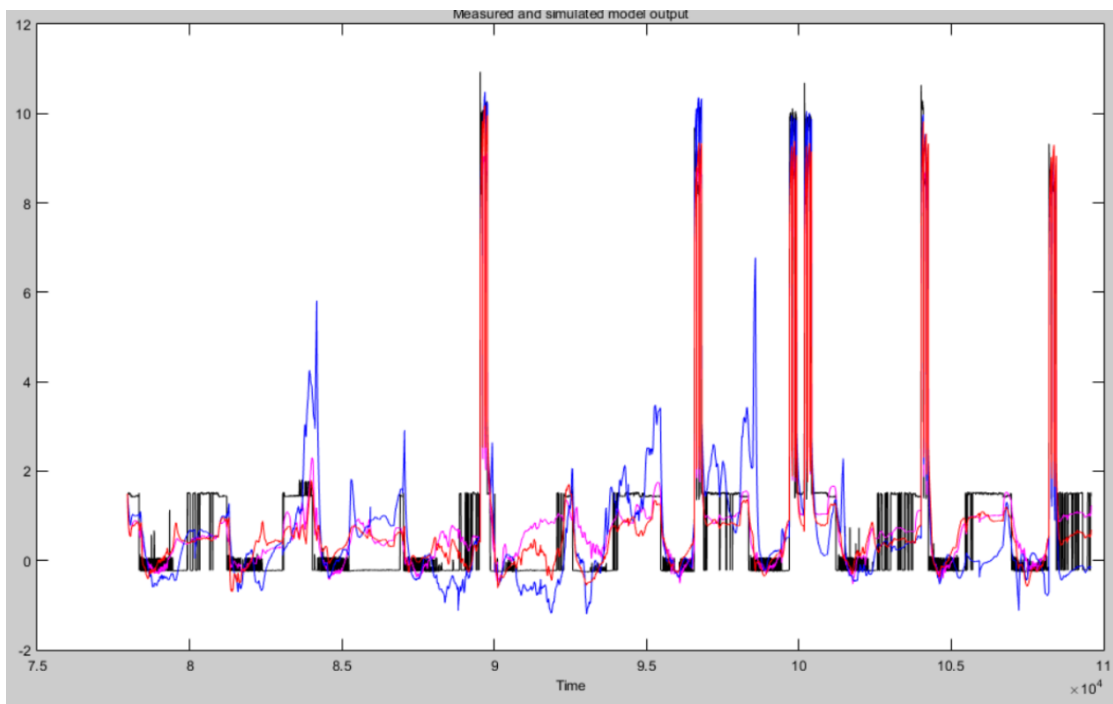


Fig. 3.68. Consumo real del invernadero (curva negra) y predicha por los modelos no lineales sin la variable tiempo (narx310ST en rojo, narx111ST en magenta y narx510ST en azul), para el conjunto de datos de estimación

Se observa un comportamiento bastante inestable del modelo Nlarx510. Los otros dos modelos predicen el consumo de forma pobre pero aceptable.

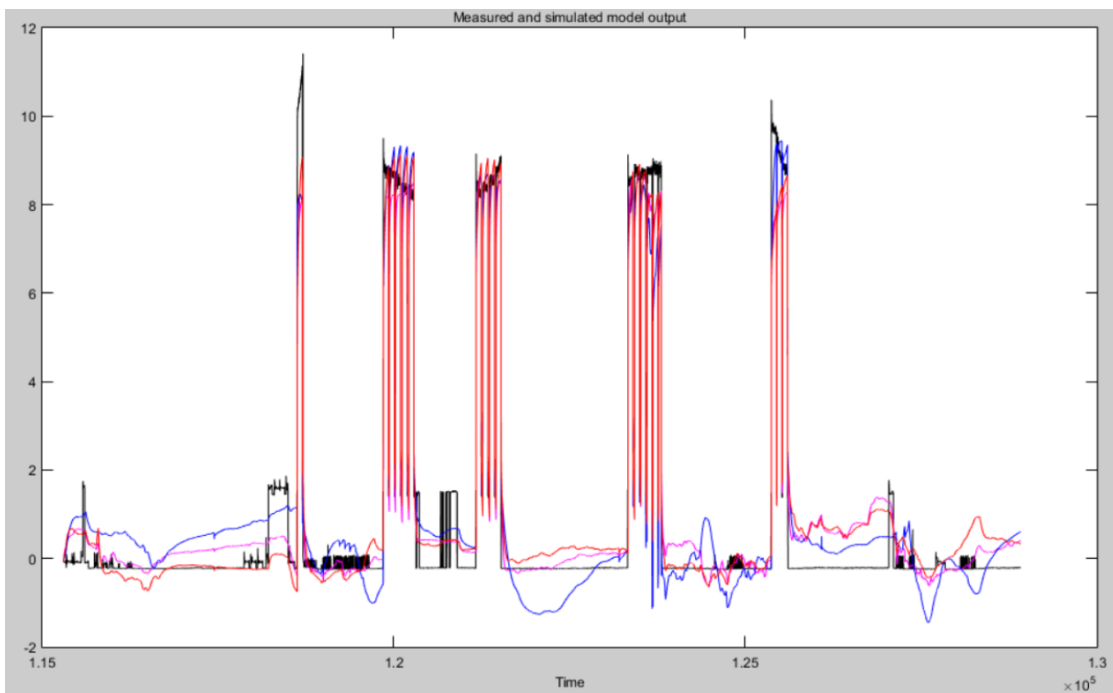


Fig. 3.69. Consumo real del invernadero (curva negra) y predicha por los modelos no lineales sin la variable tiempo (narx310ST en rojo, narx111ST en magenta y narx510ST en azul), para el conjunto de datos de Validación

Los modelos Nlarx111 y Nlarx310 reproducen mejor el consumo y no estiman valores negativos del consumo.

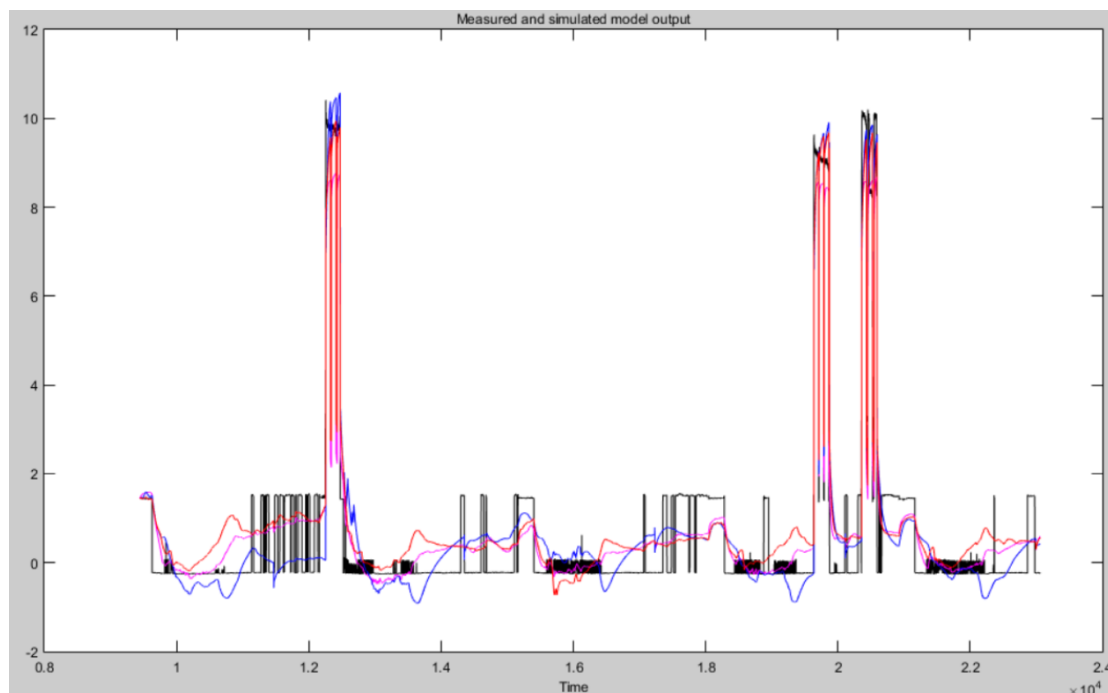


Fig. 3.70. Consumo real del invernadero (curva negra) y predicha por los modelos no lineales sin la variable tiempo (narx310ST en rojo, narx111ST en magenta y narx510ST en azul), para el conjunto de datos de Validación 2

El resultado de la predicción en la validación 2 ha mejorado considerablemente, lo que podría indicar que la decisión de eliminar el tiempo ha sido adecuada. El nuevo modelo aportado narx510ST

3.5. Comparación de modelos

En los apartados anteriores se han generado diferentes tipos de modelos, considerando diferentes estructuras, así como diferentes entradas. De los modelos generados para cada tipo de estructura se van a escoger aquellos que presentan mejores predicciones por ser los más interesante. Estos modelos se va a comparar estos modelos entre ellos. En la figura 3.71 se muestra un diagrama clasificando las estructuras generadas, así como los modelos más óptimos para cada una.

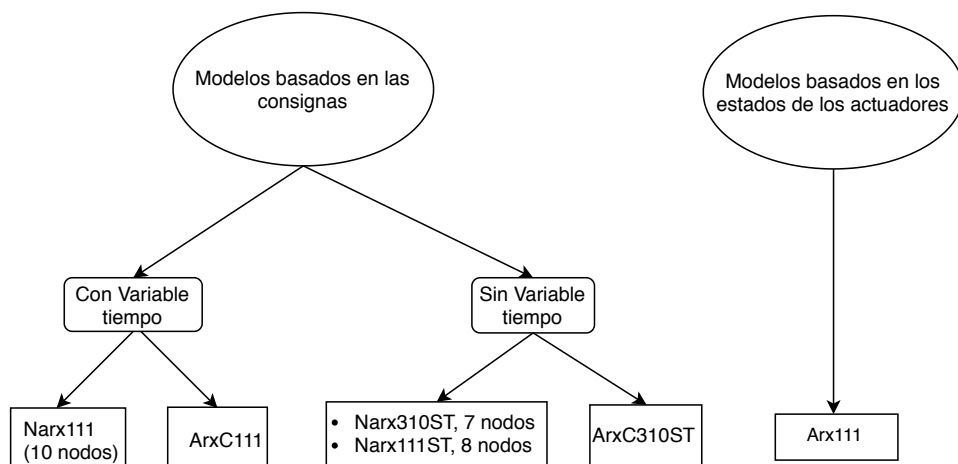


Fig. 3.71. Clasificación de las diferentes estructuras probadas y los mejores modelos de cada una

En la figura 3.71 se muestran los 5 mejores modelos estimados basados en consignas. Estos modelos son los más interesantes ya que podrían ser herramientas útiles si los resultados de la predicción son buenos. Para ver que modelos son los mejores, y hasta que punto son válidos, se van a comparar las estimaciones de consumo que arrojan para diferentes horizontes de predicción.

En la gráfica 3.72 y se muestra la predicción de los modelos con un horizonte de predicción de 30 segundos, para un periodo de 72 horas pertenecientes a los datos de validación.

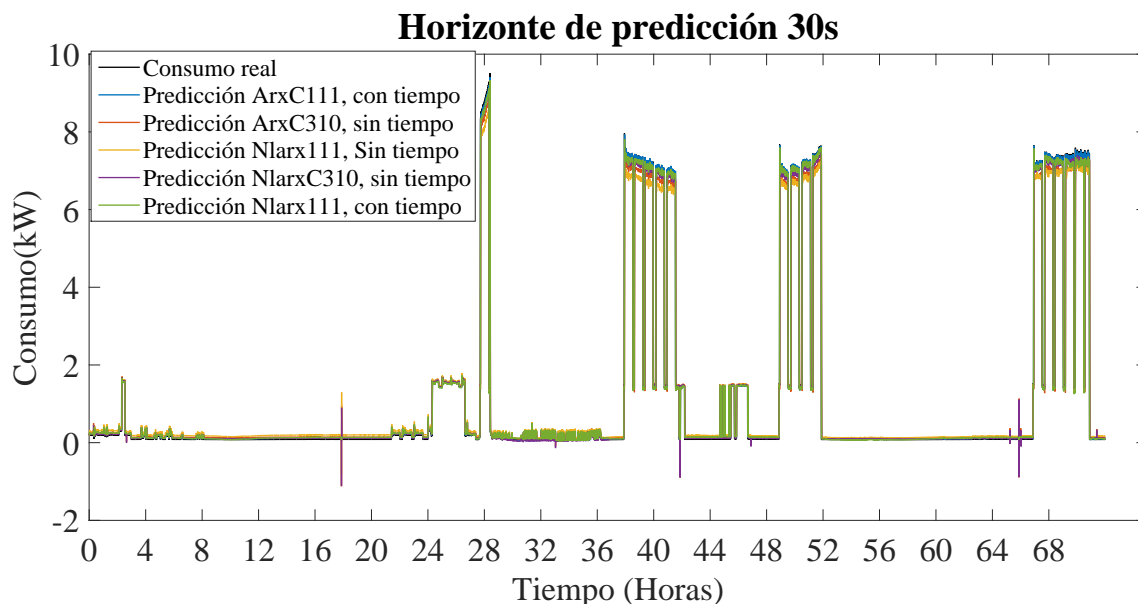


Fig. 3.72. Consumo real y su predicción con los modelos indicados, aplicados a los datos de validación. Horizonte de predicción 30 segundos

. Como se observa en esta primera gráfica la predicción se ajusta bastante bien al consumo real. En gran parte se debe a que el horizonte de predicción es el mínimo posible, ya que los datos están tomados cada 30 segundos, con bastante frecuencia. Se observa que existen 3 Outlier

en algunas de las medidas. Sin embargo no todos los modelos se ven afectados por este valor atípico en las medidas.

En la tabla 3.17 se muestran los datos estadísticos de los residuos así como el ajuste de la predicción basado en el error cuadrático medio.

TABLA 3.17. Estadísticos de la diferencia entre los valores de consumo reales y estimados por los mejores modelos generados con un horizonte de predicción de 30 segundos, para los datos de validación.

	Ajuste	Máx. error	Media error	Varianza	Desviación estándar
ArxC111	94.91	3.85	0.0414	0.0136	0.1167
ArxC310ST	82.14	8.961	0.1248	0.0173	0.416
Narx111ST	82.91	8.50	0.153	0.15	0.386
NlarxC310ST	83.08	8.65	0.0835	0.162	0.403
Narx111	94.60	3.94	0.057	0.014	0.118

De la tabla 3.17 se puede deducir que aunque todos los modelos presentan un buen ajuste para un horizonte de predicción de 30 segundos, los modelos ArxC111 y Narx111 presentan claramente los mejores resultados.

En la siguiente gráfica se muestra la predicción con un horizonte de 5 minutos. Esto permitirá distinguir mejor que modelos estiman con mayor precisión la salida real.

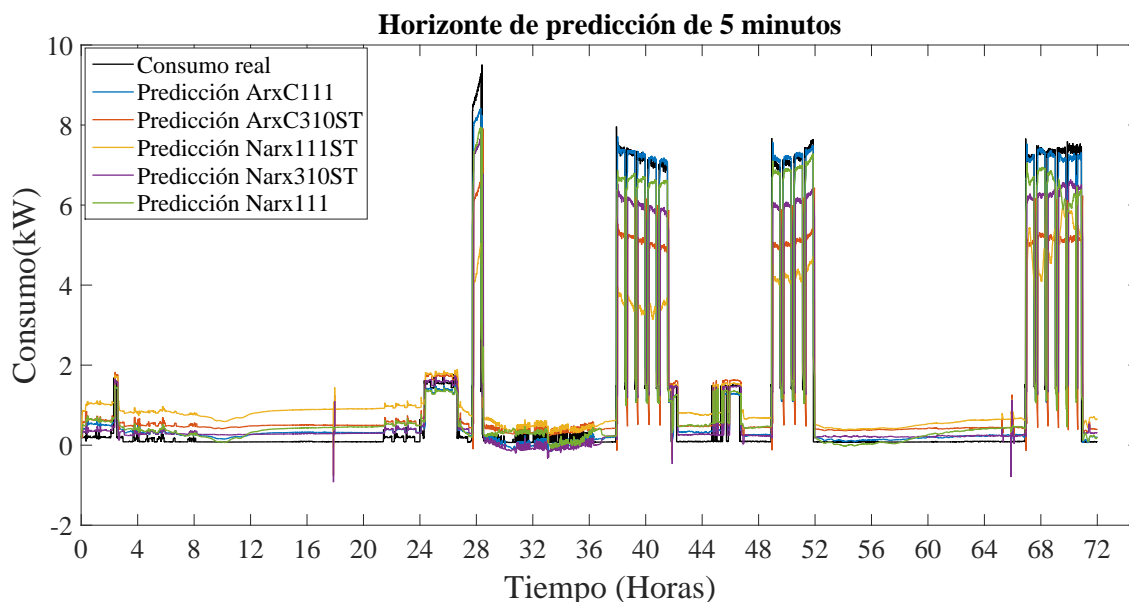


Fig. 3.73. Consumo real y su predicción con los modelos indicados aplicados a los datos de validación. Horizonte de predicción 5 minutos

Para poder comparar cuantitativamente se presenta la tabla 3.18 con los ajustes de la predicción así como los datos estadísticos de los residuos para el horizonte de predicción de 5 minutos.

TABLA 3.18. Estadísticos de la diferencia entre los valores de consumo reales y estimados por los mejores modelos generados con un horizonte de predicción de 5 minutos, para los datos de validación.

	Ajuste	Máx. error	Media error	Varianza	Desviación estándar
ArxC111	87.48	2.88	0.20	0.0525	0.2291
ArxC310ST	44.06	8.6113	0.74721	1.291	1.136
Narx111ST	40.39	7.476	0.961	1.1766	1.085
Narx310ST	53.35	8.2132	0.4814	1.0545	1.027
Narx111	82.81	2.7855	0.312	0.0772	0.2779

Se observa como la predicción de los modelos ArxC310ST, Narx111ST y Narx310ST se ha empobrecido muchísimo y no arroja resultados válidos.

Los modelos ArxC111 y Narx111, ambos con la variable tiempo presentan resultados aceptables. Sin embargo no se recomienda usarlos con un horizonte de predicción mayor de 5 minutos, ya que los estadísticos de los residuos indican un empeoramiento de los resultados. La media del error por ejemplo ya no es cercana a 0.

Aunque no se recomienda un horizonte de predicción mayor de 5 minutos, en la gráfica 3.74 así como en la tabla 3.19 se muestran los resultados de la predicción de los modelos ArxC111 y Narx111 para un horizonte de predicción de 10 minutos

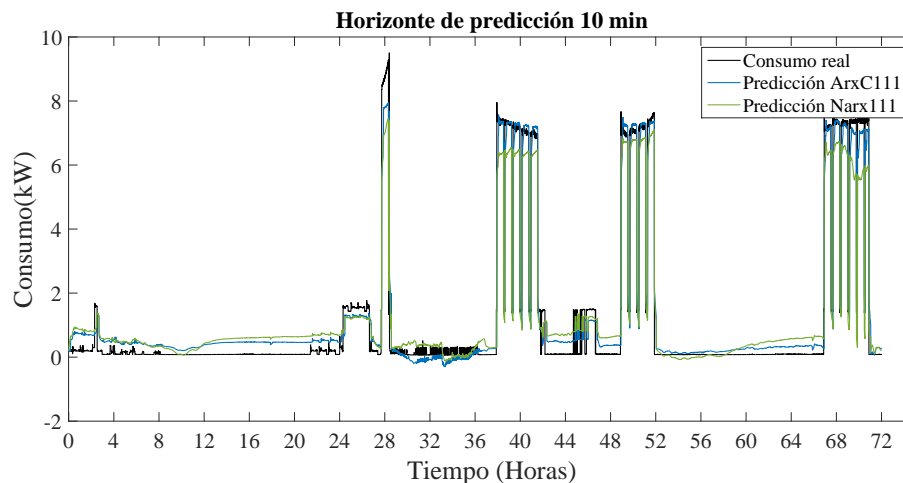


Fig. 3.74. Consumo real y su predicción con los modelos indicados aplicados a los datos de validación. Horizonte de predicción 10 minutos

TABLA 3.19. Estadísticos de la diferencia entre los valores de consumo reales y estimados por los mejores modelos generados con un horizonte de predicción de 10 minutos, para los datos de validación.

	Ajuste	Máx. error	Media error	Varianza	Desviación estándar
ArxC111	83.36	2.81	0.31	0.069	0.263
Narx111	77.05	2.74	0.44	0.115	0.339

De estos resultados, y de los anteriores con horizontes de predicción menor, se puede concluir que el mejor modelo resulta el ArxC111 y Narx111, respectivamente. Su horizonte de predicción máximo se estima en 5 minutos.

Puesto que estos modelos han sido generados pensando en estimar el consumo debido al funcionamiento de la calefacción y la ventilación, que eran los datos de consignas disponibles, se van a confirmar los resultados anteriores empleando un conjunto de datos donde el funcionamiento de estos equipos es mayoritario.

Para ello, en la gráfica 3.75 se muestra la predicción de los modelos con un horizonte de predicción de 30 segundos, para un periodo de 96 horas muy cercano al periodo de los datos de validación.

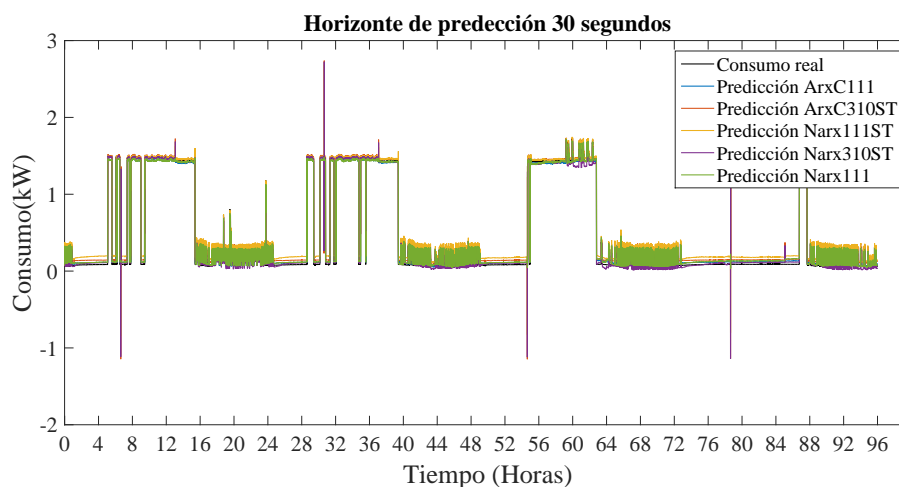


Fig. 3.75. Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción 30 segundos

En la tabla 3.20 se observan los resultados cuantitativos.

TABLA 3.20. Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 30s, para un periodo de funcionamiento de calefacción y ventilación

	Ajuste	Máx. error	Media error	Varianza	Desviación estándar
ArxC111	81.14	1.388	0.0515	0.0104	0.102
ArxC310ST	80.31	1.423	0.0701	0.0093	0.0965
Narx111ST	77.88	1.516	0.0866	0.0105	0.1022
Narx310ST	81.404	1.405	0.048	0.0104	0.1018
Narx111	81.11	1.382	0.053	0.0103	0.1013

Se puede apreciar como los modelos ArxC111 y Narx111 siguen siendo de los mejores. Pero para este horizonte de predicción el modelo Narx310ST presenta mejores resultados. Se amplía el horizonte de predicción a 5 minutos para seguir la evolución, cuyo resultado se presenta en la gráfica 3.76.

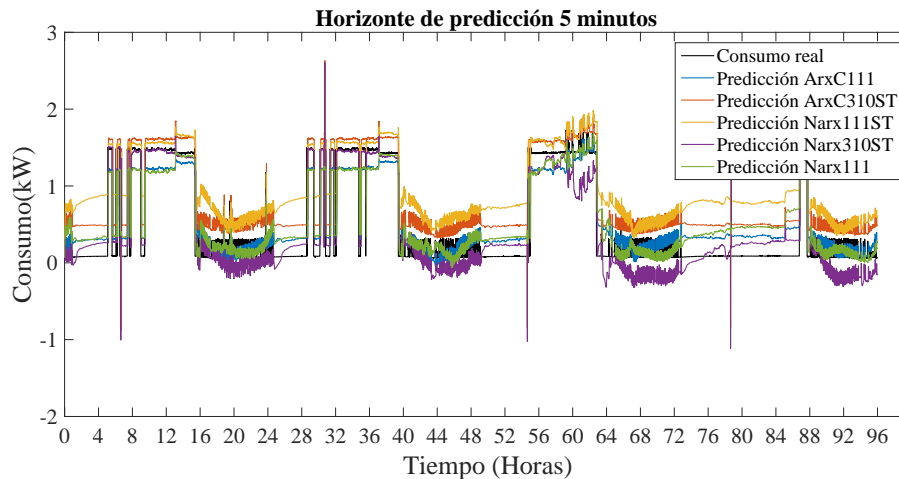


Fig. 3.76. Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción 5 minutos

Se observa en la gráfica 3.76 como aunque el modelo Narx310ST predice bien el consumo debido a la calefacción, presenta una tendencia a estimar valores negativos de consumo. Esto hace que rechacemos el modelo. Esta tendencia se pone aun más de manifiesto en la gráfica 3.77 donde se ha querido mostrar la predicción con un horizonte de 10 minutos.

Los modelos basados en la estructura Arx 111 son los que mejor captan la tendencia, aunque la predicción del consumo de calefacción podría ser mejor.

Los datos cuantitativos para la comparación del horizonte de 5 minutos están disponibles en la tabla 3.21 3.21

TABLA 3.21. Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 5 minutos, para un periodo de funcionamiento de calefacción y ventilación

	Ajuste	Máx. error	Media error	Varianza	Desviación estándar
ArxC111	56.40	1.335	0.205	0.0274	0.1656
ArxC310ST	36.47	1.596	0.3351	0.0355	0.1884
Narx111ST	10.05	1.751	0.474	0.0719	0.268
Narx310ST	53.55	1.494	0.188	0.0438	0.2093
Narx111	53.01	1.511	0.246	0.0342	0.1849

Observando los datos de la tabla 3.21, confirmamos que el modelo ArxC111 seguido del modelo Narx111 presentan los mejores resultados.

Aunque no se recomienda un horizonte de predicción mayor de 5 minutos, en la gráfica 3.74 así como en la tabla 3.19 se muestran los resultados de la predicción de los modelos ArxC111 y Narx111 para un horizonte de predicción de 10 minutos

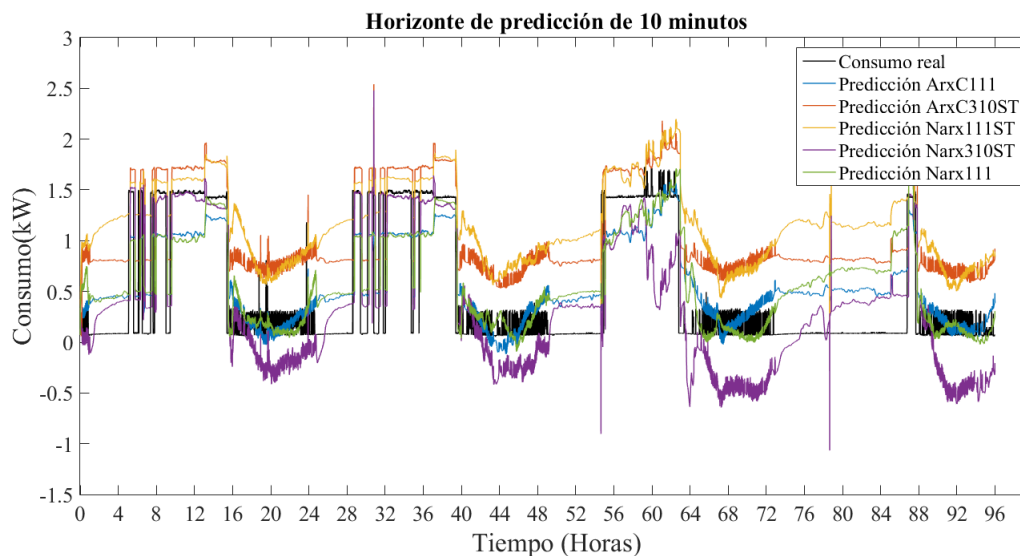


Fig. 3.77. Consumo real y su predicción con los modelos indicados aplicados a un periodo de 96 horas con calefacción y ventilación. Horizonte de predicción 10 minutos

TABLA 3.22. Estadísticos de los residuos entre los valores de consumo reales y estimados por los mejores modelos con un horizonte de predicción de 10 minutos, para un periodo de funcionamiento de calefacción y ventilación

	Ajuste	Máx. error	Media error	Varianza	Desviación estándar
ArxC111	38.01	1.314	0.314	0.042	0.20
ArxC310ST	-4.7	2.46	0.583	0.062	0.248
Narx111ST	-31.52	2.011	0.700	0.143	0.379
Narx310ST	27.96	2.39	0.33	0.08	0.284
Narx111	31.86	1.62	0.327	0.063	0.252

4. CONCLUSIONES Y FUTUROS TRABAJOS

De la toma de decisiones durante la elaboración del trabajo, así como de los resultados obtenidos se pueden obtener las siguientes conclusiones:

- El procesamiento de los datos es fundamental para obtener señales con buena calidad en la información que aportan. Los sensores de medición requieren un mantenimiento adecuado, ya que se encuentran en un ambiente con humedad, polvo y actividad humana, que pueden alterar su situación original
- La selección de variables permite reducir la dimensión del vector de entradas. En este trabajo se ha demostrado que solo la temperatura exterior es necesaria y se puede prescindir de la temperatura interior. No se puede asegurar que la temperatura del suelo sea prescindible ya que esta medición se realizó a 3 cm de profundidad en este trabajo.
- La variable tiempo u otra relacionada con el estado del cultivo puede ser prescindible, si no se modela el consumo debido riego y enriquecimiento de CO₂ como en este caso.
- Los modelos lineales, concretamente los modelos Arx, a pesar de ser estructuras sencillas proporcionan buenas predicciones del consumo del invernadero.
- Los modelos no Nlarx pueden mejorar la predicción, pero es fundamental partir de un buen modelo lineal para establecer el orden de los regresores.
- En este trabajo el mejor modelo obtenido ha sido mediante un modelo lineal. El modelo no lineal obtenido a partir de él no ha mejorado los resultados. Puesto que un modelo no lineal siempre puede mejorar la predicción del modelo lineal del que parte, se considera que se necesita un mejor entrenamiento de las RNA generadas o considerar otras estructuras con mayor número de capas ocultas u nodos.

En cuanto a las posibles líneas de trabajo futuro, se considera que es necesario la generalización de los modelos, aplicándolo a invernaderos diferentes así como profundizar en los resultados obtenidos y estudiar mejor las estructuras RNA. También sería necesario modelar el resto de equipos que aquí no se ha podido debido a la indisponibilidad de datos. Concretamente, modelar la actuación automática de la deshumificadora, aerotermo y enriquecimiento de CO₂. También se ve oportuno investigar más acerca de la adecuación o no de incluir la variable tiempo.

BIBLIOGRAFÍA

- [1] M. Berenguel Soria, L. J. Yebra y F. Rodríguez, “Adaptive control strategies for greenhouse temperature control”, en *European Control Conference ECC*, 2003.
- [2] M. Ruiz Arahál, M. Berenguel Soria y F. Rodríguez Díaz, *Técnicas de predicción con aplicaciones en ingeniería*. Sevilla: Universidad de Sevilla, Secretariado de publicaciones, 2006.
- [3] X. Blasco, M. Martínez, J. M. Herrero, C. Ramos y J. Sanchis, “Model-based predictive control of greenhouse climate for reducing energy and water consumption”, *Computers and Electronics in Agriculture*, vol. 55, n.º 1, págs. 49-70, 2007.
- [4] J. B. Oliveira, J. Boaventura-Cunha y P. B. M. Oliveira, “A feasibility study of sliding mode predictive control for greenhouses”, *Optimal Control Applications and Methods*, vol. 37, n.º 4, págs. 730-748, 2016.
- [5] P. J. M. van Beveren, J. Bontsema, G. van Straten y E. J. van Henten, “Optimal control of greenhouse climate using minimal energy and grower defined bounds”, *Applied Energy*, vol. 159, págs. 509-519, 2015.
- [6] J. C. Moreno, M. Berenguel, F. Rodríguez y A. Baños, “Robust control of greenhouse climate exploiting measurable disturbances”, *IFAC Proceedings Volumes*, vol. 35, n.º 1, págs. 271-276, 2002.
- [7] F. R. Díaz, “Modelado y control jerárquico de crecimiento de cultivos en invernadero”, Tesis doct., Universidad de Almería, 2002.
- [8] M.-H. Liang, L.-J. Chen, Y.-F. He y S.-F. Du, “Greenhouse temperature predictive control for energy saving using switch actuators”, *IFAC-PapersOnLine*, vol. 51, n.º 17, 2018.
- [9] A. Ramírez-Arias y col., “Improving efficiency of greenhouse heating systems using model predictive control”, *IFAC Proceedings Volumes*, vol. 38, n.º 1, págs. 40-45, 2005.
- [10] *Junta de Andalucía - La superficie de invernaderos de Andalucía oriental crece hasta las 35.489 hectáreas, un 1,7 % más que en la pasada campaña*. (visitado 22-06-2019).
- [11] *CHROMAE Project (DPI2017-85007-R)*. dirección: <http://www2.ual.es/chromae/> (visitado 22-06-2019).
- [12] European Commission, “Plan de Acción para la Eficiencia Energética”, inf. téc., 2006.
- [13] *Estación Experimental Cajamar | Fundación Cajamar*. dirección: <https://bit.ly/2KByk4p> (visitado 24-04-2019).
- [14] F. G. Ación, *Apuntes de Estudio de Casos Prácticos y Análisis Económicos de Proyectos*, Master en Energía Solar. Universidad de Almería, Curso 2017/2018.
- [15] C. M. Leon, “Diseño de una instalación fotovoltaica de autoconsumo para un invernadero de alta tecnología de la provincia de Almería.”, Tesis de mtría., Universidad de Almería, 2018.

- [16] *Software de hojas de cálculo - Microsoft Excel*. dirección: <https://products.office.com/es-es/excel> (visitado 25-04-2019).
- [17] *MATLAB Documentation - MathWorks España*. dirección: <https://es.mathworks.com/help/index.html> (visitado 23-04-2019).
- [18] L. Ljung, *System identification toolbox. User's Guide*, MathWorks, 2018.
- [19] K. Berk, *Modeling and forecasting electricity demand: a risk management perspective*, ép. BestMasters. Wiesbaden: Springer Spektrum, 2015.
- [20] R. Mena, "Desarrollo de un modelo neuronal para la predicción de la demanda energética del edificio bioclimático CIESOL", Tesis de mtría., Universidad de Almería, 2012, pág. 287.
- [21] A. Savitzky y M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures.", *Analytical Chemistry*, vol. 36, n.º 8, págs. 1627-1639, 1964.
- [22] A. Guerra de la Corte, *Técnicas de selección de variables en minería estadística de datos Trabajo Fin de Grado*, spa. Sevilla: Universidad de Sevilla, 2016.
- [23] S. C. Yusta, "Algoritmos meméticos para selección de variables en el análisis discriminante", vol. 49, págs. 333-347, 2007.
- [24] D. Peña, *Introducción a la estadística para las ciencias sociales*. McGraw-Hill, 1997.
- [25] D. C. Montgomery, *Introducción al análisis de regresión lineal*. México: Compañía Editorial Continental, 2004.
- [26] K. Baba, R. Shibata y M. Sibuya, "Partial correlation and conditional correlation as measure of conditional independence", *Australian & New Zealand Journal of Statistics*, vol. 46, págs. 657-664, dic. de 2004.
- [27] I. Jolliffe, *Principal Component Analysis*. Springer Verlag, 1986.
- [28] G. Udny Yule, "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers", *Philosophical Transactions of the Royal Society of London Series A*, vol. 226, págs. 267-298, 1927.
- [29] G. Box y G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.
- [30] W. Mcculloch y W. Pitts, "A Logical Calculus of Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, vol. 5, págs. 127-147, 1943.
- [31] D. O. Hebb, *The organization of behavior: A neuropsychological theory*. New York: Wiley, jun. de 1949.
- [32] P. J. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences", Tesis doct., Harvard University, 1974.
- [33] D. E. Rumelhart, G. E. Hinton y R. J. Williams, "Learning Representations by Back-propagating Errors", *Nature*, vol. 323, n.º 6088, págs. 533-536, 1986.

- [34] J. Nijskens, J. Deltour, S. Coutisse y A. Nisen, "Radiation transfer through covering materials, solar and thermal screens of greenhouses", *Agricultural and Forest Meteorology*, vol. 35, n.º 1-4, págs. 229-242, oct. de 1985.
- [35] T. Boulard y S. Wang, "Greenhouse crop transpiration model from external climate conditions", *Agricultural and Forest Meteorology*, vol. 100, págs. 25-34, ene. de 2000.
- [36] G. Bot, "Physical modeling of greenhouse climate", *IFAC Proceedings Volumes*, vol. 24, n.º 11, págs. 7-12, sep. de 1991.
- [37] H. Frausto, J. G. Pieters y J. Deltour, "Modelling Greenhouse Temperature by means of Auto Regressive Models", 2003.
- [38] J. B. Cunha, A. E. Ruano y C. Couto, "Identification of greenhouse climate dynamic models", 1996.
- [39] J. Cunha, C. Couto y A. Ruano, "A greenhouse Climate Multivariable Predictive Controller", *Acta Horticulturae*, vol. 534, págs. 269-276, ago. de 2000.
- [40] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1986.
- [41] M. Taki, "Application of Neural Networks and multiple regression models in greenhouse climate estimation", *Agricultural Engineering International : The CIGR e-journal*, vol. 18, págs. 29-43, sep. de 2016.
- [42] A. Castañeda-Miranda y V. Castaño, "Smart frost control in greenhouses by neural networks models", *Computers and Electronics in Agriculture*, vol. 137, págs. 102-114, mayo de 2017.
- [43] W. Zou, F. Yao, B. Zhang, C. He y Z. Guan, "Verification and predicting temperature and humidity in a solar greenhouse based on convex bidirectional extreme learning machine algorithm", *Neurocomputing*, vol. 249, mar. de 2017.
- [44] J.-c. Fernández y col., "Fuzzy Modeling Vapor Pressure Deficit to Monitoring Microclimate in Greenhouses", *IFAC-PapersOnLine*, vol. 49, págs. 371-374, dic. de 2016.
- [45] C. E. Hernández Hernández, "Control and Supervision Strategies for the Integrated Management of Installations in Energy Efficient Environments", Tesis doct., Universidad de Almería, 2017.
- [46] H. Akaike, "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, vol. 19, n.º 6, págs. 716-723, 1974.
- [47] T. Hill, M. O'Connor y W. Remus, "Neural Network Models for Time Series Forecasts", *Manage. Sci.*, vol. 42, n.º 7, págs. 1082-1092, jul. de 1996.
- [48] K. Funahashi, "On the Approximate Realization of Continuous Mappings by Neural Networks", *Neural Netw.*, vol. 2, n.º 3, págs. 183-192, mayo de 1989.
- [49] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review*, vol. 65, n.º 6, págs. 386-408, 1958.

- [50] K. Hornik, M. Stinchcombe y H. White, “Multilayer Feedforward Networks Are Universal Approximators”, *Neural Netw.*, vol. 2, n.º 5, págs. 359-366, jul. de 1989.
- [51] I. Seginer, T. Boulard y B. Bailey, “Neural Network Models of the Greenhouse Climate”, *Journal of Agricultural Engineering Research*, vol. 59, págs. 203-216, nov. de 1994.
- [52] R. Mena-Yedra, F. Rodriguez, M. d. M. Castilla y M. Arahál, “A prediction model based on neural networks for the energy consumption of CIESOL bioclimatic building”, *Energy and Buildings*, vol. 82, jul. de 2014.
- [53] W. Hongkang y col., “Recurrent Neural Network Model for Prediction of Microclimate in Solar Greenhouse”, *IFAC-PapersOnLine*, vol. 51, n.º 17, págs. 790-795, 2018.
- [54] P. Ferreira, E. A. Faria y A. Ruano, “Neural Network Models in Greenhouse Air Temperature Prediction”, *Neurocomputing*, vol. 43, págs. 51-75, mar. de 2002.

La agricultura intensiva bajo plástico es una de las actividades de mayor impacto económico en la provincia Almería, la primera en el sector productivo. La rentabilidad económica de la actividad pasa por obtener el máximo beneficio de la venta de los productos en el mercado y la minimización de los costos de producción. El cada vez más notable cambio climático tiene un efecto directo sobre la agricultura. Las variables climáticas suponen el factor determinante en el crecimiento y producción de los cultivos. La disponibilidad de recursos hidrológicos es también un problema sensible. En este trabajo se aborda la predicción de la demanda eléctrica en invernaderos tecnificados en función de las variables climáticas exteriores y las condiciones climáticas requeridas al cultivo dentro del invernadero por parte del productor. Para tal fin se aplican métodos de identificación de sistemas de caja negra lineales y no lineales. Antes de la generación de los modelos y su calibración, se analizan y tratan los datos de consumo, así como se estudian las variables a incluir en los modelos mediante los métodos de selección de variables.

Palabras clave: Agricultura, Identificación de sistemas, Demanda eléctrica.

Intensive crops in greenhouses are one of the main economic activities of the province of Almería, the first in the productive sector. The economic profitability of the activity rely on obtaining the maximum incomes of the trade of products and the minimization of its production cost. The more and more noteworthy climate change, also has a direct effect on agriculture. Climatic variables are the decisive factor in the growth and production of crops. The availability of hydrological resources is also a sensitive problem. This work addresses the prediction of electrical demand in HiTech greenhouses, according to the external climatic variables and the climatic conditions required inside the greenhouse by the producer. To do so, linear and non-linear black box system identification methods are applied. Before the creation and calibration of the models, consumption data are analyzed and process, as well as the variables to be included in the models are studied by variable selection methods.

Key words: Agriculture, System Identification, Electric demand

