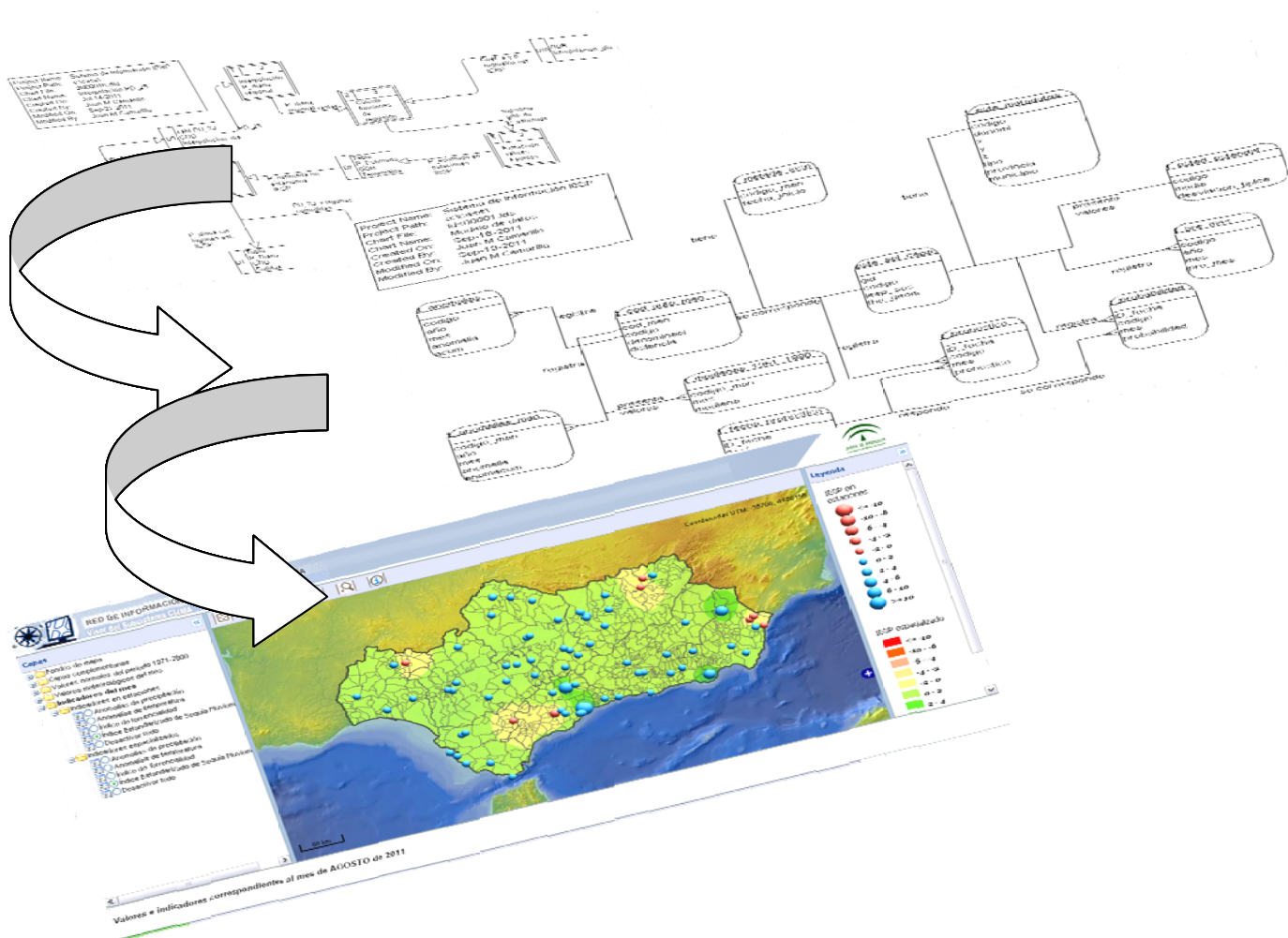


PROYECTO DE TESIS DOCTORAL

Diseño, metodología y modelo de datos para el cálculo, seguimiento y predicción del Índice Estandarizado de Sequía Pluviométrica en Andalucía



DOCTORANDO: Juan Mariano Camarillo Naranjo

DIRECTORA: María Fernanda Pita López

DEPARTAMENTO: Dpto. de Geografía Física y AGR

PROGRAMA: Cambios climáticos y morfogenéticos. Evolución de los paisajes mediterráneos

UNIVERSIDAD DE SEVILLA



Agradecimientos:

A la Dra. María Fernanda Pita López, por su magisterio profesional y humano y, sobre todo, por haberme respetado siempre.

Al Dr. José Ojeda por su visión compartida de la Geografía y su confianza en mi trabajo.

A José Ignacio Álvarez por su trabajo diario en el proyecto de configuración de este sistema. A él se deben el diseño material de los procesos en ArcGis, la configuración del servidor de mapas y la configuración del visor cartográfico web. Y por la facilidad con la que siempre he trabajado con él.

A los profesores Mónica Aguilar Alba, Dr. Ismael Vallejo, Dr. Joaquín Márquez y Dr. Enrique Santos por la compañía de todos estos años.

A la institución universitaria como engranaje básico de la educación pública.

A la Red de Información Ambiental de Andalucía y a la Consejería de Medio Ambiente de la Comunidad por su apuesta en este proyecto y la financiación para la ejecución del mismo.

A Carmen Rodríguez que me ayudó con la maquetación de la bibliografía y por ser una gran bibliotecaria por vocación y una gran amiga por devoción.

A Laura León, por su ayuda con los programas de diseño y con la maquetación del trabajo.

Los agradecimientos personales los dejo para la intimidad. Ellos y ellas saben por qué.

PARTE PRIMERA: PLANTEAMIENTOS GENERALES 1

CAPÍTULO I. INTRODUCCION, ANTECEDENTES, ESTADO DE LA CUESTION Y JUSTIFICACION DEL TRABAJO 1

- I.1. Introducción al fenómeno de la sequía. Concepto y tipologías..... 1*
- I.2. Los índices de sequía pluviométrica como indicadores ambientales de estado..... 7*
- I.3. Los sistemas de alerta temprana de sequía (Drought Early Warning Systems)..... 14*
- I.4. Justificación y oportunidad del trabajo de investigación 23*

CAPÍTULO II. OBJETIVOS DE LA INVESTIGACION: OBJETIVOS GENERALES Y OBJETIVOS ESPECIFICOS 27

- II.1. Objetivos generales..... 27*
- II.2. Objetivos específicos 34*

PARTE SEGUNDA: PLANTEAMIENTOS METODOLOGICOS Y DESARROLLO DE LAS FASES DE INVESTIGACION..... 36

CAPÍTULO III. HIPOTESIS DE PARTIDA Y PROPUESTA METODOLOGICA 36

- III.1. Hipótesis de partida 36*
- III.2. Propuesta metodológica y fases del diseño del Sistema de Información propuesto 37*

CAPÍTULO IV. ANALISIS Y DEFINICION DE LOS REQUISITOS DEL SISTEMA: MODELADO DEL FLUJO DE DATOS, MODELADO LOGICO DE LOS DATOS, DICCIONARIO DE ENTIDADES Y ARQUITECTURA DEL SISTEMA	55
<i>IV.1. Modelado del flujo de datos (Data Flow Modelling)</i>	<i>55</i>
<i>IV.2. Modelado lógico de datos (Logical Data Modelling –LDM) o diseño del modelo Entidad-Relación.....</i>	<i>140</i>
<i>IV.3. Arquitectura del sistema.....</i>	<i>157</i>
 <i>PARTE TERCERA: RESULTADOS, CONCLUSIONES Y LINEAS DE TRABAJO FUTURAS</i>	 <i>160</i>
 CAPÍTULO V: RESULTADOS DE LA INVESTIGACION OBTENIDOS	 160
<i>V.1. Resultados concretos y objetivos alcanzados por el trabajo de investigación.....</i>	<i>160</i>
<i>V.2. Satisfacción de las hipótesis propuestas.....</i>	<i>164</i>
 CAPITULO VI: CONCLUSIONES, DISCUSION Y LINEAS DE TRABAJO FUTURAS.....	 166
<i>VI.1. Conclusiones</i>	<i>166</i>
<i>VI.2. Líneas de trabajo de futuro.....</i>	<i>169</i>
 REFERENCIAS BIBLIOGRAFICAS.....	 171
 ANEXOS.....	 178

PARTE PRIMERA: PLANTEAMIENTOS GENERALES

CAPITULO I. INTRODUCCION, ANTECEDENTES, ESTADO DE LA CUESTION Y JUSTIFICACION DEL TRABAJO

La sequía es uno de los fenómenos naturales que ha despertado un mayor seguimiento por parte de la comunidad científica y una preocupación creciente por parte de los organismos implicados en la gestión de la misma. Los últimos episodios acontecidos en las dos últimas décadas y que han afectado a multitud de regiones y de espacios en grado diferente, han puesto de manifiesto la importancia de establecer mecanismos, herramientas y protocolos de monitorización y seguimiento de la evolución del fenómeno. Estrategias y esfuerzos coordinados a nivel nacional e internacional han sido desarrollados en los últimos diez años produciendo un salto cualitativo importante en la capacidad de gestión y de adaptación de un fenómeno cada vez más global y que, en nuestros contextos, contrariamente a lo deseable, ha afectado a sociedades cada vez más vulnerables.

Este es el encuadre de este proyecto de tesis que a continuación se desarrolla y que tiene por título: “Diseño, metodología y modelo de datos para el cálculo, seguimiento y predicción del Índice Estandarizado de Sequía Pluviométrica en Andalucía”.

1.1. Introducción al fenómeno de la sequía. Concepto y tipologías

El inicio de un trabajo sobre el fenómeno de la sequía debe obligatoriamente abordar la naturaleza conceptual del mismo. Especialmente si dicho fenómeno se presenta de forma compleja en el conjunto de interrelaciones que se producen en torno al mismo y que afectan a las variables físico-ambientales, económicas, sociales e institucionales de los sistemas y subsistemas implicados.

Es por ello que se ha abordado una revisión del propio concepto de sequía y de las clasificaciones que han sido desarrolladas por los distintos autores con el objetivo, por un lado, de indagar en la naturaleza del fenómeno y, por otro, de enmarcar de forma precisa este trabajo de investigación dentro de la estructura compleja de las diferentes esferas que intersectan de forma transversal o tangencial el núcleo conceptual del fenómeno estudiado. Si bien hemos de advertir que la búsqueda de una definición universal de sequía es una opción no realista (Wilhite, 2005), fundamentalmente debido a la compleja naturaleza del fenómeno, así como a la multitud de factores físicos, ambientales, biológicos, económicos, sociales, políticos e institucionales que se ven involucrados en la génesis, la ocurrencia y los impactos del mismo. Adicionalmente, la conceptualización de un fenómeno como la sequía puede variar según la región o área geográfica siendo también dependiente del contexto institucional (Quiring S.M, 2009).

Las distintas definiciones del fenómeno de sequía que han sido aportadas por el conjunto de autores consultados presentan una serie de características comunes:

- Reflejan un intento de aproximación conceptual que responda a la complejidad intrínseca del fenómeno, por lo que encontramos multitud de aproximaciones basadas en enfoques disciplinares, escalas, métodos de estudio y categorías diferentes. Wilhite y Glantz (1985) encontraron 150 definiciones del término sequía en la revisión que llevaron a cabo a principios de los años 80 y propusieron el establecimiento de cuatro categorías tipológicas de sequía: meteorológica, hidrológica, agrícola y socioeconómica, ampliamente aceptadas en la actualidad. Nos encontramos, pues, ante un fenómeno complejo que permite un abordaje desde aproximaciones, métodos y análisis muy diferentes.
- Enmarcan de forma muy común a la sequía en la esfera de los riesgos naturales. Las diferentes aproximaciones que pueden realizarse a un fenómeno dotado de gran complejidad quedan bien articuladas en la esfera de la sequía como riesgo natural. La evaluación de la peligrosidad

del fenómeno, a partir de los análisis y métodos estadísticos orientados a sacar a luz el carácter anómalo (desviación), la intensidad y la duración de la sequía se asocian al tipo de sequía meteorológica, mientras que los métodos orientados a la evaluación de la exposición y de la vulnerabilidad, se asocian a las tipologías hidrológica, agrícola y socio-económica. Es la inclusión del fenómeno de la sequía dentro de la esfera de los riesgos naturales la que permite la integración de diferentes métodos, análisis, enfoques y clasificaciones en los estudios sobre este fenómeno.

Recientemente el Grupo para la Discusión sobre la Sequía dentro de la Estrategia Internacional para la Reducción de Desastres (*ISDR* en sus siglas en inglés) ha propuesto un nuevo paradigma para mejorar el conocimiento del riesgo –en su acepción inglesa de *risk*-. Este nuevo paradigma enfatiza la necesidad de un mayor conocimiento y descripción de ambos elementos, la naturaleza física del riesgo por un lado y los factores sociales que influyen en la vulnerabilidad social por otro (Wilhite, 2005). Este paradigma pone especial acento en la necesidad de definiciones conceptuales, enfoques y métodos orientados a la existencia de impactos sobre cualquiera de las esferas interesadas.

- De forma generalizada se aprecia un importante peso de la componente espacial y geográfica en los estudios de sequía en tanto que se observa una interrelación muy estrecha entre dicha componente espacial y los mecanismos de génesis, impactos generados y sistemas de gestión de la sequía.
- La mayor parte de las definiciones consultadas hacen referencia a las condiciones deficitarias anormales en el ciclo de la precipitación y/o en la disponibilidad del recurso, así como en la potencial influencia en términos de impacto en algunas de las esferas ambiental, económica y social.

Conceptualmente podemos esquematizar esta serie de consideraciones a partir de la Figura I.1, que recoge de forma sintética el conjunto de interrelaciones y enfoques anteriormente especificados.

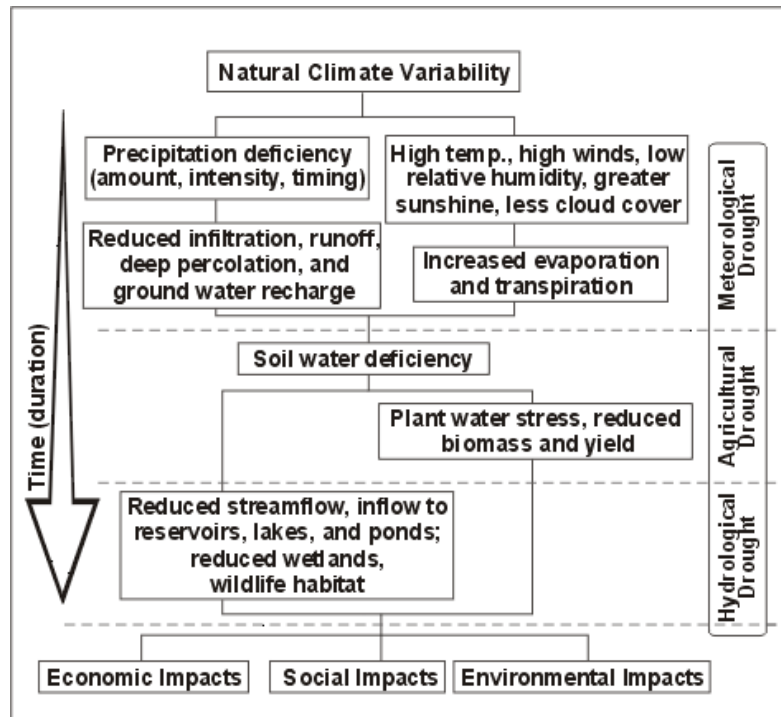


Figura I.1. Esquema conceptual de la sequía

Fuente: <http://www.drought.unl.edu/whatis/concept.htm>

Para nuestro trabajo hemos adoptado un enfoque que se corresponde con dos definiciones de sequía aportadas por distintos autores y que nos parecen que enmarcan el fenómeno de forma precisa y que reflejan la complejidad del mismo.

Pita López (2009) define la sequía como un “*déficit hídrico inusual, una anomalía pluviométrica negativa lo suficientemente intensa y prolongada como para generar impactos adversos en la sociedad que la padece*”, enmarcando claramente el fenómeno en la determinación de los “*dos hechos definitorios esenciales de la sequía, su origen natural y su carácter extremo, siendo los impactos un corolario que la acompaña como posibilidad aunque no como rasgo esencial*”.

Termina enfrentando la sequía con el concepto a veces erróneo de escasez, en tanto en cuanto se ha venido produciendo una igualación de ambos términos que en demasiadas ocasiones ha supuesto una confusión terminológica y conceptual en los distintos debates y enfoques. Especialmente si otorgamos al término escasez un

“origen predominantemente antrópico, un carácter no por fuerza coyuntural y, sobre todo, un protagonismo incuestionable de los impactos negativos”.

Esta definición se corresponde con la primera de las tipologías de sequía aceptadas, la sequía meteorológica, conectando el fenómeno con el origen natural del mismo y entendido este como un riesgo natural que es una parte normal de los climas de virtualmente todas las regiones (Wilhite, 2005). Entendido de esta forma, es la potencialidad de generar daños de este comportamiento anormal y coyuntural de una variable meteorológica como la precipitación, la que conecta con la esfera más amplia de los impactos, la exposición y la vulnerabilidad de una sociedad. Por lo tanto no podemos asignar, desde esta acepción, un carácter intrínsecamente perjudicial y aún menos el grado de dichos perjuicios a la sequía, entendida esta como sequía meteorológica.

Será este el marco en el que se sitúa este trabajo de investigación, lo cual obviamente no implica la asunción del marco teórico más amplio en el que obligatoriamente deben insertarse los estudios sobre sequía; si bien, a efectos de situar con la mayor precisión posible el lugar del mismo, hemos de indicar que este abordará la sequía en tanto que *sequía meteorológica*, abundando por tanto en los valores de peligrosidad que podemos asociarle en virtud de la metodología que se adoptará en el transcurso del mismo.

La adscripción de este trabajo a la definición conceptual anterior se completa mediante la necesaria ampliación del concepto de sequía desde el punto de vista operacional. Desde esta perspectiva hay dos tipos principales de definiciones de sequía: conceptuales y operacionales. Las primeras son formuladas en términos generales y son utilizadas para ayudar a enmarcar y entender la naturaleza del fenómeno tal y como hemos apuntado a través de la definición de Pita López (Pita, 2009). Por el contrario las segundas son muy específicas y se utilizan principalmente para identificar las características y las dimensiones de cada una de las sequías, su comienzo, su finalización y el grado de severidad que pueden alcanzar (Quiring S.M, 2009).

Abundando en dicha cuestión, los estudios y análisis de la sequía presentan una vertiente operacional en términos de sistemas de alerta (*warning systems*) que podemos definir como el conjunto de procedimientos de medición, seguimiento, monitorización y predicción –en los casos más avanzados- del fenómeno (Svoboda *et al*, 2002).

La necesidad de establecer sistemas de monitorización del comportamiento temporal y espacial de la sequía en tiempo real ha incidido en el desarrollo, el diseño y la producción de sistemas de indicadores específicos para dicho propósito. Estos sistemas de indicadores se basan en la aplicación de lo que podemos englobar genéricamente en el concepto de “índices de sequía”.

Las diferencias de enfoques posibles en virtud de la definición conceptual y tipología de sequía en la que nos enmarcamos, determinarán en gran medida la selección y aplicación de los indicadores que deben implementarse para un territorio concreto dentro del marco operacional y de la monitorización del fenómeno. En nuestro caso, al habernos circunscritos a la dimensión meteorológica de la sequía, adoptaremos, como podrá comprobarse en la definición de los objetivos del trabajo de investigación, la aplicación de un índice operacional de seguimiento de la sequía pluviométrica, concretamente el Índice Estandarizado de Sequía Pluviométrica (IESP) desarrollado por Pita López, M.F (Pita, 2001).

Al mismo tiempo, será también objetivo de este trabajo de investigación el análisis de las necesidades funcionales a las que debe responder un sistema de monitorización del comportamiento de la sequía pluviométrica en una región determinada, así como en las fases y requisitos funcionales de la configuración de un sistema de información para hacer efectiva la monitorización del fenómeno.

1.2. Los índices de sequía pluviométrica como indicadores ambientales de estado

El seguimiento de un fenómeno como la sequía exige el diseño de métodos que consigan medir y cuantificar los tres componentes principales de los que participa: la magnitud (déficit pluviométrico acumulado por debajo de un determinado umbral), la duración (número de meses o años consecutivos con valores por debajo de dicho umbral) e intensidad (amplitud de los déficits respecto al umbral considerado). Junto a estos, otros parámetros de caracterización serían la velocidad de implantación o tiempo que tarda una sequía en alcanzar su máxima intensidad, la extensión espacial a la que va a afectar el fenómeno o la variabilidad espacial que presenta.

En este sentido, múltiples indicadores o índices pueden ser encontrados en la literatura científica, habiendo sido algunos de ellos adoptados en su vertiente operacional en sistemas de alerta. De hecho, cualquier propuesta de planificación o gestión de este riesgo natural se sustenta, en primera instancia, en la definición de los índices que van a ser utilizados como indicadores expresivos del comportamiento del fenómeno.

Existen numerosas propuestas de índices de sequía, diseñados para algunos de los diferentes tipos que quedaron expuestos en el epígrafe conceptual –sequía meteorológica, sequía hidrológica, sequía agrícola y sequía social-. Entre los más utilizados y testados por la comunidad científica en la esfera de los índices meteorológicos, objeto de este trabajo, podemos destacar el Índice de *Palmer (PDSI)*, no estrictamente pluviométrico, utilizado ampliamente en las décadas de 1970, 1980 y 1990, el *Rainfall Anomaly Index* (Van Rooy, 2005), el *Bhalme & Mooley Index* (Bhalme & Mooley, 1980) o el *Standardized Anomaly Index* (Katz y Glantz, 1986), si bien en la actualidad, el índice de más amplia difusión y aplicación es el *Standardized Precipitation Index* (McKee et al, 1993 y 1995). Una revisión y evaluación completa de los principales índice de sequía en cualquiera de sus vertientes puede consultarse en Marcos (2001) o Heim (2002).

En Andalucía y España comienza a extenderse en los últimos años la aplicación del Índice Estandarizado de Sequía (IESP) (Pita, 2001), fundamentalmente debido a una mejor adaptación del mismo al tipo de sequías largas que suelen darse en amplias regiones de la Península Ibérica, así como a la validez de un indicador que es independiente en su cálculo de la duración de la sequía, en contraposición al SPI (Pita, 2001). Una definición y valoración del IESP será abordada en el siguiente epígrafe.

Todos estos indicadores e índices deben ser enmarcados en un concepto más amplio como es el de indicador ambiental de estado. Según la Agencia Europea de Medio Ambiente (EAMA) un indicador es un *“valor observado representativo de un fenómeno determinado. En general, los indicadores cuantifican la información mediante la agregación de múltiples y diferentes datos. La información resultante se encuentra pues sintetizada. En resumen, los indicadores simplifican una información que puede ayudar a revelar fenómenos complejos”*.

En la publicación de la Dirección General de Calidad y Evaluación Ambiental del Ministerio de Medio Ambiente, titulada *“Indicadores Ambientales. Una propuesta para España”*, un indicador ambiental es definido como *“una variable que ha sido socialmente dotada de un significado añadido al derivado de su propia configuración científica, con el fin de reflejar de forma sintética una preocupación social con respecto al medio ambiente e insertarla coherentemente en el proceso de toma de decisiones”*.

Así pues un indicador ambiental es una variable que, mediante la síntesis de la información ambiental, pretende reflejar el estado del medio ambiente, o de algún aspecto de él, en un momento y en un espacio determinados, y que por ello adquiere gran valor como herramienta en los procesos de evaluación y de toma de decisiones políticas sobre los problemas ambientales. Un indicador ambiental debe por lo tanto cumplir una serie de requisitos fundamentales:

- Ser científicamente válido, estar basado en un buen conocimiento del sistema descrito.

- Ser representativo del conjunto.
- Ser sensible a los cambios que se produzcan en el medio o en las actividades humanas relacionadas con él.
- Estar basado en datos fiables y de buena calidad.
- Ofrecer información relevante para el usuario, además de simple y clara para facilitar la comprensión de la misma por parte del usuario no especializado.
- Ser predictivo, de manera que pueda alertar sobre una evolución negativa.
- Ser comparable.
- Presentar un buen equilibrio coste-efectividad.

Los indicadores dependen de la noción de medio ambiente que se adopte; ésta se encuentra frecuentemente asociada a la noción de problema y por ello los indicadores tienden a evaluar la situación de los principales problemas ambientales. También se desarrollan con frecuencia indicadores según el ámbito de aplicación, fundamentalmente los dirigidos a la evaluación de los resultados de las políticas ambientales y aquellos referidos a la integración de los aspectos ambientales en las políticas económicas y sectoriales. Los indicadores siguen diferentes marcos organizativos, que pueden ser causales, temáticos, sectoriales, ecosistémicos,.... Estos marcos no son en absoluto excluyentes, y de hecho se cruzan frecuentemente.

Desde esta perspectiva creemos necesario enmarcar el objetivo del este trabajo de investigación, las necesidades de diseño de un sistema de información para el cálculo y el seguimiento del IESP en Andalucía, dentro de este concepto más amplio de indicador ambiental de estado. El propio contexto institucional así lo determina facilitando de esta forma la configuración de herramientas que puedan dar respuesta a las necesidades de gestión y a las obligaciones normativas que existen en Andalucía, maridando la investigación y la transferencia efectiva de conocimiento aplicado.

I.2.1. El Índice Estandarizado de Sequía Pluviométrica (IESP)

La selección, entre la batería de indicadores de sequía pluviométrica existentes en la literatura, del Índice Estandarizado de Sequía Pluviométrica (IESP) está sustentada en la necesidad de utilizar, en el marco del proyecto, un indicador de fácil aplicación y bien adaptado a nuestro medio, capaz de realizar el seguimiento de la sequía pluviométrica en todo el territorio andaluz.

Tal y como ha sido mencionado, en los procesos de gestión del riesgo se hace necesario disponer de indicadores de sequía meteorológica “naturales” y diferentes de los indicadores de “escasez”, que son los que se utilizan básicamente para hacer el seguimiento de la relación aportes/demandas de agua y establecer los umbrales y sistemas de alerta consecuentes, para llevar a cabo, en suma, la gestión del agua.

Es por ello, que integrar herramientas que permitan mejorar el conocimiento y el seguimiento del inicio, la implantación, el alcance y la intensidad de ciclos de sequía es uno de los objetivos y obligaciones de los diferentes organismos de gestión.

En Andalucía, la integración del cálculo en tiempo real del índice propuesto, formaría parte esencial del Subsistema de Climatología Ambiental (CLIMA) y sería también de aplicación en fases posteriores en todas las cuencas, subcuencas y sistemas de explotación y abastecimiento de agua.

El cálculo del Índice debe, además, aplicarse de forma automática y en rutina a escala mensual tanto al territorio andaluz, en una primera fase, como a las distintas cuencas vertientes que lo componen o a los sistemas de abastecimiento existentes.

Al ser el IESP un índice de sequía meteorológico participa de una serie de elementos valorables en los procesos de seguimiento del fenómeno de la sequía:

- Aunque la sequía es el resultado de un balance hídrico en el que intervienen la precipitación como aporte de agua y la evapotranspiración

como demanda fundamental, la precipitación es el elemento esencial que interviene en su génesis dado que es el más variable, de manera que puede afirmarse que la inmensa mayoría de la sequías son atribuibles a un descenso de la precipitación en mucha mayor medida que a un aumento de la demanda evaporativa. En la práctica, el conocimiento de la sequía pluviométrica suele ser suficiente para conocer el grado de sequía general del territorio.

- La precipitación es el parámetro más abundantemente medido en el territorio y los índices pluviométricos son los más sencillos e inmediatos de aplicación, con lo cual, además, el coste/beneficio de estos índices es inmejorable.
- Suele ser el paso previo a la sequía hidrológica y la económica y, por tanto, tiene un valor anticipatorio que puede ser muy útil para la gestión.
- En la Directiva Marco del Agua se establece la necesidad de delimitar con nitidez, para una situación de sequía, la parte de responsabilidad natural de la que es atribuible total o parcialmente al hombre, de forma tal que sólo las sequías de origen estrictamente natural y de una intensidad y duración excepcionales permiten justificar prácticas de desembalse susceptibles de alterar el comportamiento de los ecosistemas fluviales.
- Los sistemas de seguros y la asistencia del Estado en casos de fenómenos extremos, también exigen precisar el carácter extremo y estrictamente natural del fenómeno para responsabilizarse o no de la asistencia a los damnificados.

I.2.1.1. Ventajas del IESP

Se propone el Índice Estandarizado de Sequía Pluviométrica (IESP), ideado por Pita López (Pita, 2001), aplicado con muy buenos resultados para todo el ámbito español

y contrastado con el *Standardised Precipitation Index (SPI)* para 6, 12, 24 y 36 meses, arrojando mucho mejores resultados que este último, especialmente para las sequías largas (Pita, M.F., 2001), normales y frecuentes en nuestra región.

Efectivamente, el *SPI* es un índice basado en la acumulación de las anomalías pluviométricas mensuales, pero en el que se establece un límite temporal apriorístico para esas acumulaciones (3, 6, 12, 24... meses) independientemente de la realidad de la sequía que se esté produciendo. Ello determina que las acumulaciones para periodos de 3 o 6 meses sean útiles para identificar las sequías cortas, pero resulten inútiles para las largas. A la inversa, las acumulaciones de 24 o 36 meses, por ejemplo, describen bien las sequías largas, pero no las cortas. El índice que proponemos soluciona este problema y describe por igual las sequías de distintas duraciones.

En la actualidad el IESP está presente en todos los Informes anuales de Medio Ambiente de la Junta de Andalucía y en la información ambiental de su página web.

I.2.1.2. Cálculo del IESP

Los datos de partida utilizados por el índice son las precipitaciones mensuales y su cálculo se realiza en tres etapas sucesivas.

En la primera de ellas se calcula la anomalía pluviométrica de cada uno de los meses de la serie, a partir de la expresión:

$$AP_i = P_i - PMED$$

Donde:

AP_i = Anomalía pluviométrica mensual.

P_i = Precipitación mensual

PMED = Precipitación mediana del mes.

En la segunda fase se calculan las anomalías pluviométricas acumuladas, desde el primer mes de la serie. En el momento en que se encuentra una anomalía acumulada negativa se iniciaría una secuencia seca, que concluiría con la aparición de una anomalía acumulada positiva, dando paso a una secuencia excedentaria en agua; durante esta secuencia excedentaria, las anomalías siguen acumulándose hasta que aparezca de nuevo una anomalía pluviométrica negativa, momento en que se iniciaría una nueva secuencia seca, que se obtiene por el mismo método, recomenzando el cálculo de las acumulaciones a partir de ese valor negativo de anomalía pluviométrica. En consecuencia, el cálculo de esta segunda fase se resume de la forma siguiente:

$$APAi = \sum APi$$

Desde $i = 1$ hasta $APi < 0$ y $APAi-1 \geq 0$

Donde:

$APAi$ = Anomalía pluviométrica acumulada del mes.

Por último, en la tercera fase se estandarizarían estas anomalías acumuladas mediante su conversión en puntuaciones z.

$$ZAPAi = (APAi - APA) / \sigma_{APA}$$

Donde:

$ZAPAi$ = Anomalía pluviométrica acumulada estandarizada del mes.

APA = Valor medio de las anomalías pluviométricas acumuladas de todos los meses de la serie.

σ_{APA} = Desviación típica de las anomalías pluviométricas acumuladas de todos los meses de la serie.

I.3. Los sistemas de alerta temprana de sequía (Drought Early Warning Systems)

La severidad creciente de los impactos de las sequías en sociedades desarrolladas con un alto grado de exposición a las mismas, especialmente en aquellas zonas en las que el fenómeno presenta una recurrencia cíclica, que ha sido observada en las últimas décadas, han determinado el creciente desarrollo paralelo de estrategias de monitorización del fenómeno.

En España, en las dos últimas décadas hemos asistido a la ocurrencia de dos ciclos extraordinarios de sequía (figura I.2). Especialmente destacable ha sido la de la década pasada, con una intensidad muy extrema -11 millones de ciudadanos sufrieron cortes de suministro- y que ha coincidido con un aumento en la exposición de la sociedad española cuyo volumen poblacional y tasas de crecimiento económico imponían un alto grado de presión sobre los recursos hídricos. La constatación del aumento de nuestra vulnerabilidad, junto a la naturaleza recurrente que el fenómeno tiene en nuestros entornos climáticos es coincidente con el de otras regiones del mundo que han sufrido situaciones similares en los últimos 20 años –EEUU, Australia, Israel,...-.

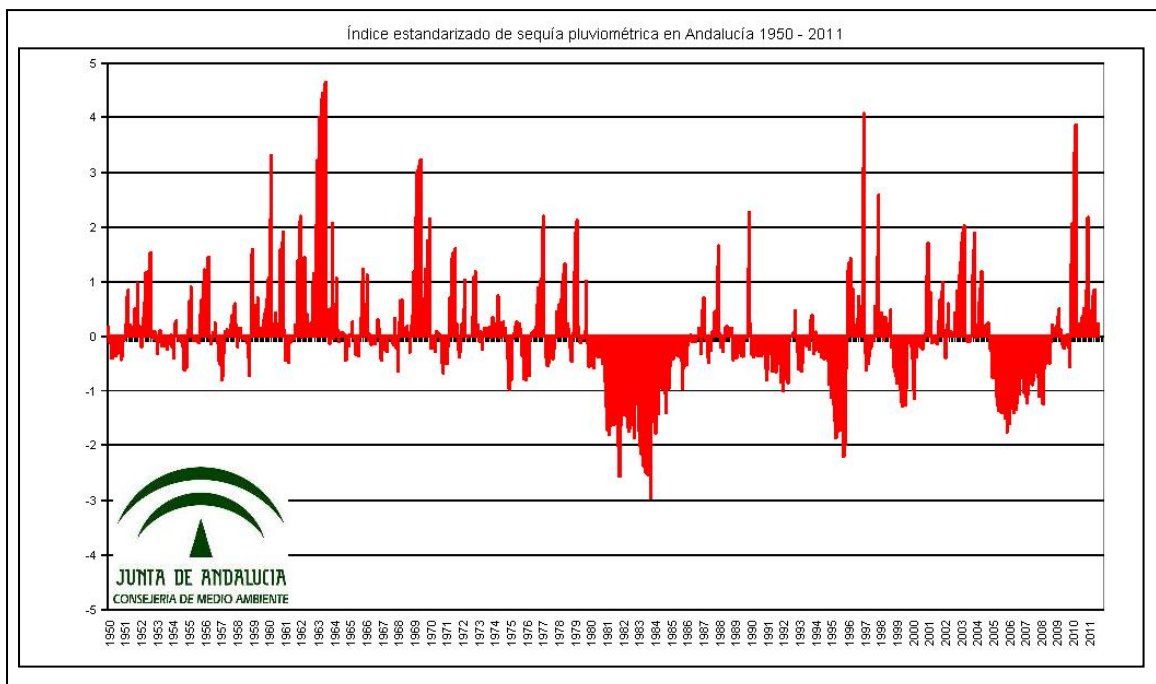


Figura I.2. IESP en Andalucía (1950-2011)

Fuente: Consejería de Medio Ambiente. Junta de Andalucía

La complejidad ya puesta de manifiesto del propio fenómeno, su capacidad potencial de generar impactos muy extensos en el conjunto de todas aquellas esferas relacionadas con los recursos hídricos (Wilhite, 2005) tales como la coyuntura económica, el estado del medio ambiente, la disponibilidad de suministro humano, etc..., así como la dispersión de los impactos a áreas geográficamente extensas, han llevado a las instituciones públicas de gestión a una preocupación creciente por la implantación de sistemas de alerta de sequías.

Frente a este fenómeno intenso, la importancia creciente observada sobre los sistemas de alerta temprana de sequías –*Drought Early Warning Systems* en sus siglas en inglés- viene determinada por una serie de necesidades que se han puesto de manifiesto en las últimas experiencias de manejo y gestión de las mismas en estos últimos años.

En España es destacable la iniciativa del Ministerio de Medio Ambiente de crear un Comité de Expertos en Sequía, que comenzó su labor en el año 2005 y que produjo un documento “La sequía en España. Directrices para minimizar su impacto” que puso las bases, entre otras cuestiones ligadas a la planificación, para la implantación progresiva de sistemas de indicadores de seguimiento del fenómeno (Pita, 2005).

En dicho documento, se pone especial énfasis en la absoluta necesidad de un sistema de indicadores para el seguimiento de la sequía en el conjunto de España.

Tal y como se afirma en el mismo documento, muchas son las razones que abogan en favor del establecimiento de un sistema de indicadores capaz de realizar el seguimiento de la sequía en el territorio (Pita, 2005):

- En primer lugar, la necesidad de anticiparse a la sequía y de preverla, dentro de la tendencia actual a considerarla como un riesgo más y no como una crisis pasajera a la que hubiera que hacer frente mediante mecanismos de resolución de emergencias.

- Esta necesidad se ve además favorecida por dos caracteres de la sequía que hacen especialmente favorable y posible esta anticipación: su duración prolongada y su lenta velocidad de implantación, que determinan que el déficit hídrico se vaya excavando progresivamente, permitiendo en este lapso temporal todo tipo de medidas de ajuste capaces de minimizar los impactos negativos.
- También apunta en la misma dirección la Directiva Marco del Agua cuando señala la necesidad de justificar el carácter extremo y estrictamente natural de una sequía para justificar cualquier descenso en los parámetros de calidad de las aguas o en el estado ecológico de los ríos. Tal justificación sólo puede realizarse a partir de un sistema de indicadores capaz de reflejar con claridad ambos caracteres.
- Los indicadores son también esenciales como instrumentos de ayuda en el establecimiento de las primas a pagar por los seguros frente a la sequía y en la toma de decisiones respecto al pago de las correspondientes indemnizaciones cuando existen dudas acerca de la naturaleza del fenómeno. En la actualidad es la actividad agraria la que hace uso masivo de estos seguros, pero es previsible que se extiendan hacia actividades cada vez más diversas, como sucede también con otros tipos de riesgos.

En la esfera internacional, sin duda es Estados Unidos el país que más esfuerzos y recursos ha dedicado y dedica al desarrollo de sistemas de alerta temprana por sequías en el mundo. Especialmente destacable es la firma por parte del Presidente americano del Acta pública 109-430 de Diciembre del año 2006 que obliga a las instituciones federales a desarrollar *“un sistema de información de sequías dinámico y accesible que proporcione a los usuarios la habilidad y la capacidad de determinar los potenciales impactos de la sequía y de los riesgos asociados que conlleva, así como las herramientas para la toma de decisiones necesarias para una mejor adaptación y mitigación del riesgo”*.

I.3.1. Objetivos, necesidades y diseño de los sistemas de alerta de temprana de sequías

Los objetivos básicos que debe cumplir un sistema de alerta temprana de sequías deben responder a las necesidades de planificación, estrategias de adaptación, evaluación de impactos y opciones de mitigación indispensables desde la óptica de gestión pública de un riesgo como la sequía (Wilhite, 2005).

Por tanto las opciones de diseño de estos sistemas deben obligatoriamente observar la inclusión de una serie de componentes básicos (Svoboda, 2006):

- Monitorización y predicción
- Evaluación del riesgo: indicadores y “disparadores” –*triggers*–
- Planificación ante el riesgo de sequías y adaptación y preparación al mismo
- Portal de sequías
- Comunicación y educación

La integración de estos componentes en un sistema de información de alerta temprana debe, de esta forma, asegurar la coordinación en términos de información necesaria procedente de la multitud de organismos implicados en el fenómeno: centros meteorológicos nacionales y/o regionales, organismos de gestión de cuencas hidrográficas, organismos de gestión de suministros, organismos nacionales y/o regionales implicados en la gestión del medio ambiente, la agricultura, etc... La complejidad del fenómeno necesita pues de datos procedentes de fuentes de información muy diversa que deben integrarse de forma efectiva para el correcto funcionamiento del sistema: datos meteorológicos –precipitación, temperatura, humedad, evapotranspiración...-, imágenes de satélite de contenido de humedad en el suelo y construcción de índices como el NDVI, datos de embalses y aforos, datos de piezómetros, datos de consumo de agua por sectores de actividad y abastecimiento humano, etc...

Al mismo tiempo debe garantizar la configuración de un ágil centro distribuidor de información sobre la sequía que sea capaz de proporcionar tanto al ciudadano como a los planificadores productos y servicios útiles de gestión, entre los que cabe destacar especialmente la configuración de un portal web unificado, así como productos estandarizados como bases de datos, mapas, predicciones, servicios interoperables, etc...

En nuestro caso, el objetivo de este trabajo se centra en el primero de los componentes señalados, cuyo elemento principal es el desarrollo de sistemas de información para la monitorización, el seguimiento y la predicción del fenómeno de la sequía y más concretamente de su vertiente meteorológica.

Desde esta óptica de monitorización y seguimiento de la sequía, varias son las necesidades a las que debe responder el diseño de sistemas de información de alerta de sequías y que han sido puestas de manifiesto por diferentes autores (Gibbs, 1987, Sazonov, 1989, Svoboda, 2002, Carbone 2008):

- Debe incluir una variedad de índices de sequía, adaptados a la naturaleza compleja del fenómeno y aplicados cada uno de ellos a los distintos tipos de sequía coexistentes: pluviométrica, hidrológica, agrícola y social.
- Los índices adoptados deben reflejar el estado actual del fenómeno, acercándose a lo que se conoce con el nombre de *near real time response* o respuesta en tiempo casi real.
- Debe presentar el comportamiento de la sequía a partir de su distribución espacial, ofreciendo cartografías de estado del fenómeno que permitan observar su localización y alcance.

- Debe generar predicciones del comportamiento futuro de la sequía a partir de los índices utilizados para cumplir con un doble objetivo: conocer los escenarios futuros de planificación y dotarse de una herramienta de evaluación del impacto de las medidas correctoras que vayan adoptándose.

Estas necesidades, a las que debe dar respuesta el diseño del componente de monitorización y seguimiento de las sequías dentro de estos sistemas de alerta, están acompañadas de una serie de dificultades o problemáticas comunes en todos los entornos de producción. Entre estas, varias son las que debemos destacar ya que, en su definición y resolución, van a encontrarse algunos de los objetivos de este trabajo de investigación:

- En primer lugar, la dificultad asociada a la disponibilidad de los registros meteorológicos a escala diaria. La fuente de datos principal de registros diarios procede de las distintas redes de estaciones meteorológicas automáticas, con una menor cobertura espacial del territorio. Por el contrario, los datos procedentes del conjunto de redes manuales, en la mayoría de las ocasiones como es el caso español, presentan un lapso de registro importante entre la observación en la estación, el envío del conjunto del mes a la AEMET, la validación de los registros y la grabación en el servidor de datos (Rodríguez, 2009).

Por lo tanto, la necesaria utilización de los registros de las EMA introduce una serie de dificultades, la más importante de las cuales es el importante número de días sin registro que invalidan el seguimiento pluviométrico del mes, además de los errores aleatorios, sistemáticos o aberrantes que caracterizan estas series.

- En segundo lugar, la dificultad de coordinar datos procedentes de distintas medición, fuentes y organismos. En este sentido, la implementación de sistemas capaces de integrar datos de diferente procedencia es una herramienta necesaria como punto de partida de la construcción de un sistema de monitorización de la sequía. Un ejemplo de ellos es el Subsistema CLIMA de información de climatología ambiental integrado en la Red de Información Ambiental de Andalucía REDIAM, que recolecta sistemáticamente los datos diarios procedentes de las distintas redes de estaciones automáticas que operan en Andalucía. De esta forma, en relación a las observaciones meteorológicas, se dispone en Andalucía de una magnífica herramienta que integra, de forma centralizada, todos los datos necesarios para el seguimiento de la sequía.

Una alternativa de futuro será la progresiva generación de servicios web de acceso a datos que serán incorporados en los años venideros en los sistemas de almacenamiento de datos de los diferentes organismos productores y que permitirán una conexión directa a las bases de datos de producción.

- En tercer lugar, la dificultad derivada de la propia complejidad del fenómeno, con una extrema amplitud tanto temporal como espacial y que obligará al diseño de herramientas de largo alcance con un alto grado de esfuerzos de coordinación. En palabras de Sazonov (1989), *“la cuestión capital para los estudiosos de la sequía y de su formación en el tiempo y el espacio es cómo definir las y seguir su evolución. La determinación del riesgo de sequía exige disponer de técnicas y procedimientos confiables y viables, que faciliten identificar el inicio de la sequía, caracterizar su evolución espacio-temporal y precisar el fin de la misma”*.

I.3.2. El ejemplo del sistema de monitorización de Estados Unidos -US Drought Portal-

La necesidad, especialmente en nuestros ámbitos, de caminar hacia la configuración de sistemas de monitorización de la sequía tanto en la escala nacional como en la escala autonómica, obliga a la revisión de alguna de las experiencias más avanzadas en la esfera internacional.

Es sin duda la experiencia estadounidense la que nos puede servir de ejemplo en el camino emprendido y en los avances necesarios que debemos dar con el objetivo indispensable de dotarnos como sociedad de una herramienta de información y gestión ante el riesgo de sequía.

Sin querer realizar un análisis exhaustivo de la experiencia estadounidense, materializada en el portal nacional del -US Drought Portal-, sí que nos gustaría destacar algunos aspectos relacionados con este trabajo de investigación; es decir, con las herramientas presentes en dicho sistema que han sido desarrolladas para el seguimiento del fenómeno en su vertiente meteorológica.

En primer lugar se trata de un portal que se sustenta en un esfuerzo importantísimo de coordinación entre las administraciones y las instituciones de investigación entre otras. El resultado de estos esfuerzos ha dado lugar al *National Integrated Drought Information System –NDIS-*, responsable de dar respuesta a los objetivos mencionados en el epígrafe anterior.

En segundo lugar, y para nosotros el elemento principal que concierne a este trabajo de investigación y a los trabajos desarrollados para la Consejería de Medio Ambiente de la Junta de Andalucía, la monitorización de la sequía se basa en tres elementos principales:

- El cálculo de uno o más índices de sequía en tiempo real de seguimiento –en el caso concreto de *NIDIS*, el *Standardized Precipitation Index (SPI)* y el índice de sequía de *Palmer-*.

- La disponibilidad de pronósticos sobre el comportamiento de esos índices.
- La difusión del comportamiento de estos índices en un visor geográfico web, recientemente incorporado, con acceso a los datos en el territorio capaces de ofrecer una visión espacial de un fenómeno tan complejo como la sequía. Este hecho es de vital importancia ya que garantiza una visión espacial “viva” del comportamiento y evolución temporal de la sequía de en las distintas zonas de Estados Unidos.

Un ejemplo de estas cartografías en tiempo real (se ofrecen los mapas diarios del comportamiento del SPI) pueden ser observados en la figura I.3 –distribución espacial del valor del SPI el día 22 de Septiembre de 2011- y en la figura I.4 –pronóstico del indicador-.

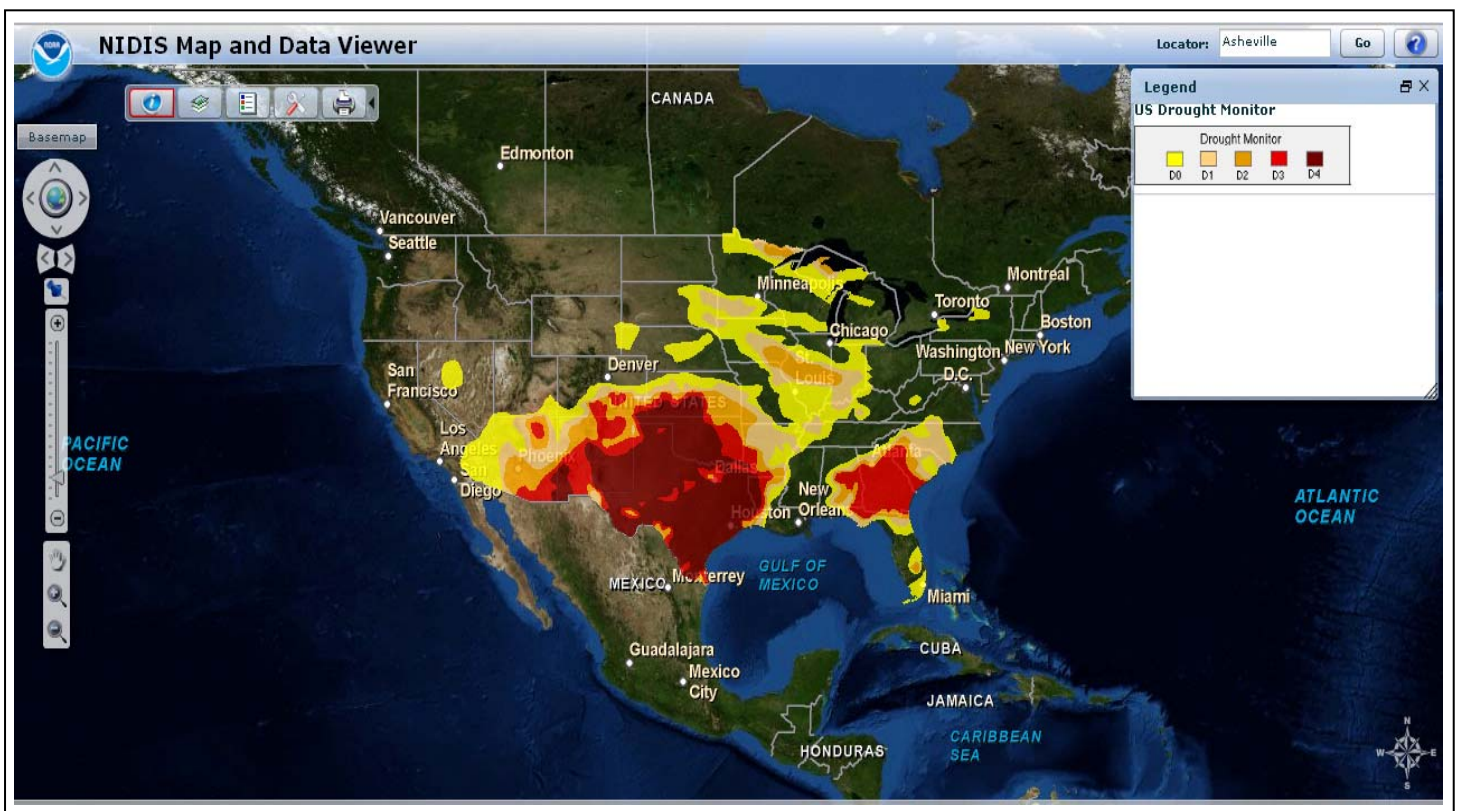


Figura I.3. Captura de NIDIS

Fuente: http://www.drought.gov/portal/server.pt/community/drought.gov/map_and_data_viewer

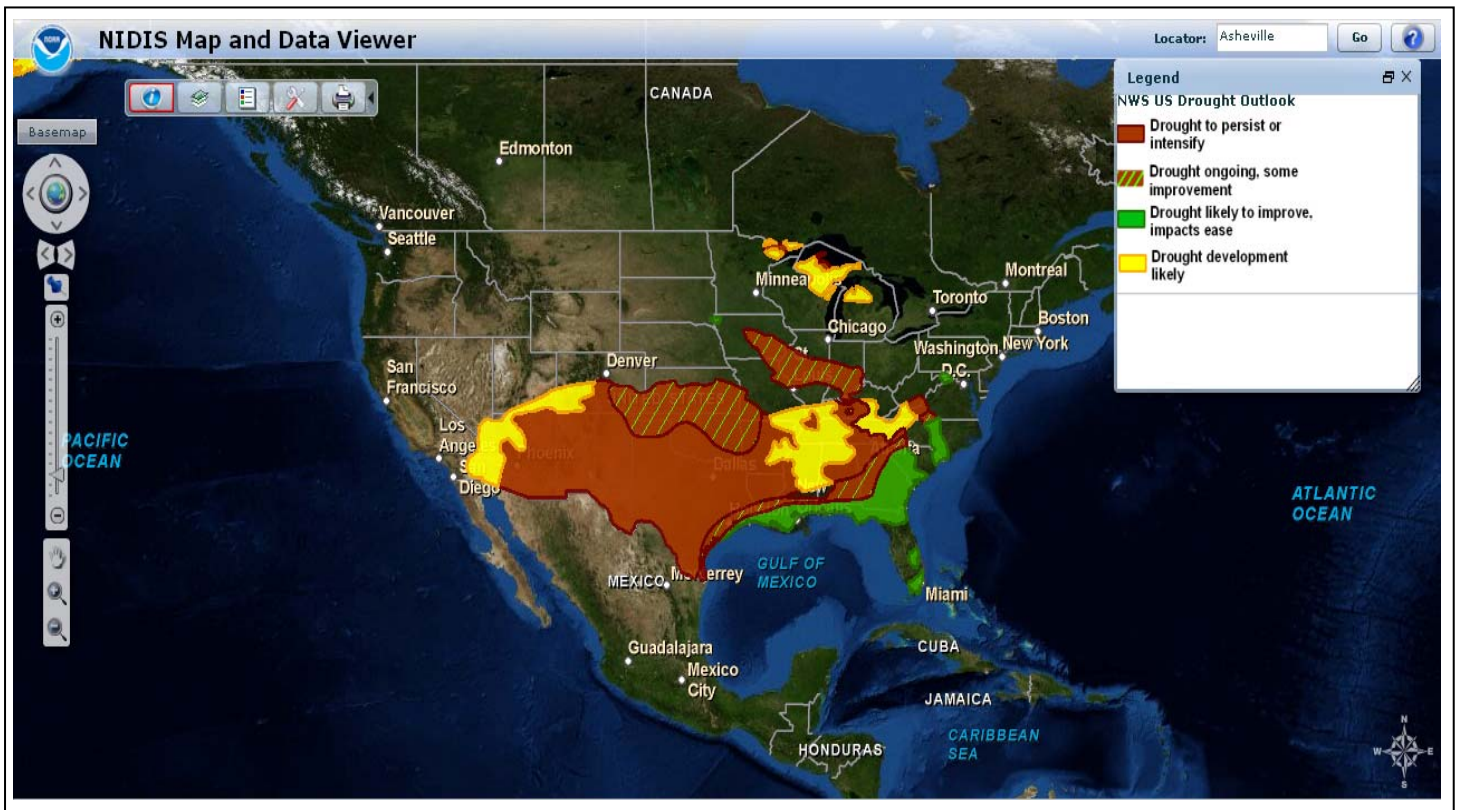


Figura I.4. Captura de NIDIS

Fuente: http://www.drought.gov/portal/server.pt/community/drought.gov/map_and_data_viewer

Pues bien, la posibilidad de implementar en Andalucía un sistema de información para el seguimiento de la sequía que aúne estos tres elementos anteriormente mencionados es el objetivo central de este trabajo de investigación y de los trabajos desarrollados. Analizar la naturaleza de la información de partida en los sistemas que almacenan la información meteorológica, sus dificultades y problemáticas, analizar los requisitos funcionales y operativos que son imprescindibles para llevar a cabo el diseño y posterior puesta en funcionamiento, van a ser los grandes objetivos de este trabajo.

1.4. Justificación y oportunidad del trabajo de investigación

El trabajo de investigación que aquí se presenta como candidato a la obtención del título de Doctor por la Universidad de Sevilla se justifica a partir de una serie de razonamientos y oportunidades que se enmarcan en dos esferas o rangos diferenciados: por un lado, los que se encuentran dentro del marco operacional en el

que nos estamos moviendo y, por otro, los que pueden ser vinculados a los aspectos de naturaleza más científica que sustentan a los primeros.

Dentro de las justificaciones de orden operacional, debemos destacar por su importancia las siguientes:

- La necesidad evidente puesta de manifiesto en todos los documentos de gestión y en gran parte de la literatura científica de desarrollar sistemas de alerta temprana para la gestión de un riesgo natural como es la sequía. Y dentro de estos sistemas, de la necesidad de contar con indicadores fiables de seguimiento basados en el cálculo y difusión de índices de sequía en tiempo real o casi real. En la revisión realizada de los sistemas de monitorización y seguimiento de la sequía existentes tanto en la escala nacional como internacional, es precisamente este uno de los elementos que queda sin resolver. En España no existe ni a nivel estatal ni a nivel autonómico ninguna información en este sentido. Tan solo Andalucía, en el contexto de la REDIAM presenta algún indicador mensual medido en tiempo real, si bien la cartografía que se ofrece es de carácter estático -imágenes-. El contexto internacional presenta las mismas características con tan solo algunos casos que presentan cartografías de índices estáticas, salvo el caso estadounidense anteriormente mencionado que acaba de incorporar mapas dinámicos que muestran el comportamiento espacial y los valores de los índices de sequía que difunden.

Avanzar en un diseño operacional y puesta en funcionamiento de un sistema para la monitorización de la sequía en Andalucía, a partir del cálculo del IESP mensual en tiempo real y su difusión en cartografías dinámicas y visores web, como primer paso necesario en la configuración de un sistema completo de alerta temprana de sequías, justifica, desde nuestro punto de vista, la oportunidad de este trabajo de investigación.

- Este trabajo, adicionalmente, se enmarca dentro de los contratos de investigación que el grupo de climatología adscrito a este Departamento ha suscrito y desarrollado por encargo de la Consejería de Medio Ambiente de la Junta de Andalucía. Entre estos últimos debemos mencionar: *“Evaluación de la incidencia del clima sobre la sequía en Andalucía. Elaboración de índices climáticos de sequía”*, en el que se sentaron las bases conceptuales y científicas en términos de naturaleza y característica de los fenómenos de sequía en la Comunidad y revisión de los índices disponibles y *“Diseño e implementación del Índice Estandarizado de Sequía Pluviométrica IESP”*, que tenía como objetivo la puesta en marcha de un sistema de monitorización regional de la sequía a partir del cálculo y difusión del IESP en tiempo real.

Desde esta perspectiva la relevancia creciente que los procesos de transferencia de conocimiento deben jugar en el contexto actual, animan a seguir avanzando en la relación entre los equipos de investigación y los agentes, en este caso institucionales, en marcos de I+D+i como los que, desde la Consejería de Medio Ambiente han sido alentados y apoyados. Este proyecto de investigación creemos que es un buen ejemplo de ello.

- Un tercer aspecto a destacar es la necesidad de ofrecer y publicar un esquema metodológico y práctico en el diseño de sistemas para el seguimiento del IESP independientemente del lugar en el que se implante, puesto que las necesidades y los requisitos funcionales necesarios para su cálculo son invariables. De esta forma pretendemos ofrecer una herramienta global para la implantación de estos sistemas cada vez más necesarios en los entornos de gestión de sequías.

Desde la segunda perspectiva, algunos son los elementos principales que justifican la elección de un trabajo de investigación como el que se presenta:

- Avanzar en el conocimiento de la naturaleza de los indicadores de sequía, particularmente en su vertiente operacional, lo que nos lleva a profundizar en la investigación de la naturaleza de la información meteorológica disponible en términos de calidad, disponibilidad, etc..., en la naturaleza y procedimientos de cálculo de estos indicadores, así como en los requisitos y procedimientos analíticos de orden espacial y geográfico que implica su cálculo y su inserción operacional en sistemas de información y contextos tecnológicos actuales.

- Obtener como resultado de la investigación una vez delimitado como objetivo de la misma un modelo de datos universal para el cálculo del indicador propuesto que pueda ser revisado y adoptado en marcos operacionales en la gestión de las sequías.

- Evaluar el papel del experto en información climática en los procesos de diseño de sistemas de información climatológicos como el propuesto. Especialmente en las aportaciones que pueden ser realizadas derivadas de los requisitos y las necesidades particulares de las que la información climática participa y que son tremendamente sensibles en el éxito operacional de los mismos. Estos sistemas no solo deben almacenar información, sino que su diseño debe responder a las preguntas que, de forma apriorística, deben ser formuladas. Entendemos que precisamente, en la fase de formulación inicial, se encuentra un nicho profesional que debe ser resaltado en este trabajo de investigación. El manejo del lenguaje y procedimientos utilizado por la ingeniería de sistemas facilita además, entendemos, el proceso de trabajo en términos de productividad y, sobre todo, en términos de satisfacción operativa de los sistemas, evitando de esta forma errores de diseño que en muchas ocasiones limitan la operatividad de los mismos.

CAPITULO II. OBJETIVOS DE LA INVESTIGACION: OBJETIVOS GENERALES Y OBJETIVOS ESPECIFICOS

La definición de los objetivos generales y específicos del trabajo de investigación, dentro del marco conceptual abordado en el epígrafe anterior, son el elemento clave para entender el contenido, el alcance y las dimensiones del trabajo presentado. En términos prácticos definen por si mismos el cuerpo del trabajo de investigación, la metodología o metodologías adoptadas y desarrolladas así como, por supuesto, las conclusiones obtenidas en términos de satisfacción y consecución de dichos objetivos. La definición precisa y diferenciada de los mismos es, por lo tanto, una de las partes más importantes, sino la que más, de todo el proceso de producción del trabajo de investigación, convirtiéndose de esta forma en los testigos y referentes de la evaluación del mismo.

Los objetivos planteados definen apriorísticamente el lugar de llegada de la investigación desarrollada, la meta que pretende ser alcanzada a partir de las fases metodológicas, analíticas y operacionales que van a ser diseñadas. Es por ello que una definición nítida, sucinta y correcta de los objetivos del trabajo de investigación es el mejor vehículo que podemos seguir para un adecuado y eficaz diseño de los caminos en términos de pasos y fases ordenadas que debemos diseñar. Definen por tanto intrínsecamente el conjunto del trabajo de investigación, desde sus fases embrionarias a la presentación final de resultados.

II.1. Objetivos generales

Los objetivos principales que se establecen en este trabajo de investigación son los siguientes:

II.1.1. Diseñar un Sistema de Información (SI) operacional para la monitorización y seguimiento de la sequía pluviométrica en Andalucía

Este es el gran objetivo del trabajo de investigación, insertar el cálculo del indicador de sequía pluviométrica IESP en el contexto actual de las Tecnologías de la Información (TI), produciendo un Sistema de Información operacional completo, de validez universal, que permita el control de todo el proceso de diseño conceptual, diseño físico a partir de un SGBD (Sistema de Gestión de Bases de Datos), cálculo del indicador, cálculo de probabilidades, producción de la cartografía mensual del mismo y difusión a través de Internet. Tal y como se puso de manifiesto en el capítulo introductorio, existen varios ejemplos internacionales de sistemas de información operacionales de monitorización de sequías, entre los que cabe destacar aquellos desarrollados en el seno del Centro Nacional para la Mitigación de la Sequía -*National Drought Mitigation Center*- y del Sistema de Información Nacional Integrado para la Sequía -*National Integrated Drought Information System*- en los Estados Unidos o el Centro Nacional del Clima -*National Climatic Centre*- australiano, no existiendo ningún ejemplo operacional activo en las administraciones estatales o autonómicas españolas.

Desde esta óptica de pretender configurarse como sistema integral, por tanto, el diseño final que debe alcanzarse debe responder a los siguientes requerimientos funcionales:

- debe garantizar el control sobre todas las fases del proceso, desde el diseño conceptual a la publicación en Internet
- debe garantizar el control de calidad necesario sobre los datos de entrada en el sistema -la precipitación diaria procedente de las estaciones meteorológicas automáticas del conjunto de redes de observación de Andalucía-
- debe garantizar y efectivamente producir el cálculo del IESP en su escala mensual como indicador ambiental de estado.

- debe incorporar un método para el cálculo de la probabilidad y pronóstico del comportamiento del indicador en un horizonte temporal futuro obteniendo, de esta forma, predicciones fiables y estadísticamente significativas del fenómeno
- debe incorporar procedimientos para la producción de cartografías que revelen el comportamiento espacial del fenómeno
- debe permitir la elaboración de servicios interoperables de mapas (WMS) para su difusión universal en el contexto de Internet.

II.1.2. Avanzar en el análisis de la naturaleza de la información climática y de la información climática derivada en el contexto actual de las Tecnologías de la Información y la Comunicación (TIC) y de la sociedad de la información¹

Este segundo objetivo es un vector transversal a todo el trabajo de investigación ya que pretende desgranar elementos tales como las necesidades actuales que son exigibles a la información de naturaleza climática para ser integrada en Sistemas de Información operados con las tecnologías presentes, la potencialidad de análisis, explotación y difusión actuales de la misma, los requisitos de homogeneidad necesarios o el modelo de relaciones implícitas que permita su explotación sistemática.

Importante será también determinar, en el contexto metodológico de análisis y diseño de sistemas de información, las especificidades y requisitos particulares que deben contemplarse, obligatoriamente, en las distintas fases de ese diseño en relación a la naturaleza de la información climática.

¹ La UNESCO en 1982 definió las Tecnologías de la Información y la Comunicación como “un conjunto de disciplinas científicas, tecnológicas, de ingeniería y de técnicas de gestión utilizadas en el manejo y procesamiento de la información: sus aplicaciones, las computadoras y su interacción con los hombres y máquinas; y los contenidos asociados de carácter social, económico y cultural.”. El sociólogo Manuel Castells en su trilogía “La Era de la Información” (2001-2006) enmarca en una teoría sistémica completa los conceptos de tecnología y de sociedad informacional que sustentan el marco conceptual actual de estos términos.

II.1.3. Establecer un modelo de datos relacional para el cálculo del Índice de Sequía Pluviométrica IESP con validez universal y por lo tanto adaptable y utilizable en cualquier entorno de producción del mismo

Según la definición más aceptada, un modelo es una simplificación de la realidad que debe incorporar las relaciones funcionales existentes entre los distintos elementos básicos que lo componen.

En el caso que nos ocupa –un modelo de datos- será necesario que su diseño e implementación responda a una serie de cuestiones básicas y esenciales:

- establecimiento de las necesidades de almacenamiento de la información generada
- definición de los objetos o entidades a partir de los cuales se estructurará la información o los datos extraídos de la realidad. Estos objetos o entidades del modelo serán tablas alfanuméricas en un contexto de SGBD.
- establecimiento de los requisitos informativos a los que debe responder ese modelo y en especial a la información derivada o calculada –vistas o consultas- que debe producir
- definición de las relaciones existentes entre los distintos objetos o entidades a partir de los cuales se estructura la información contenida en el sistema y que harán posible la interrelación de los datos o atributos asociados a cada una de las unidades de observación en forma de variables.
- definición de los requisitos de acceso por parte de los usuarios en términos de rapidez de respuesta y de interoperabilidad con los clientes tanto humanos como clientes de aplicaciones o Web.

Al mismo tiempo será necesario tener en cuenta una serie de consideraciones específicas en virtud de la naturaleza de la información hacia la que va diseñado el modelo de datos que pretende diseñarse:

información de naturaleza geográfica y climática. Entre estas podemos destacar los siguientes:

- incorporación de las geometrías de las unidades de observación espaciales que necesariamente deberán estar presentes en el modelo de datos: los observatorios meteorológicos.
- establecimiento de las relaciones y cardinalidades entre dichas geometrías y el resto de entidades u objetos –tablas alfanuméricas- que componen el modelo
- definición de las vistas o consultas a las que el modelo deberá responder y que, en nuestro caso, hacen referencia directamente a nuevas variables con una alta significación y contenido climático.

Desde esta perspectiva, por tanto, el sistema propuesto se conceptualiza como un ***sistema de información climático sustentado por una base de datos de naturaleza geográfica.***

Por último el modelo debe pretender tener validez universal en su propuesta y definición; es decir, se trata de diseñar un modelo de datos válido para el cálculo del IESP, independiente de los valores meteorológicos medidos o del ámbito de producción. Lo importante, desde este punto de vista, es el diseño de la estructura organizada de las entidades del modelo y de las relaciones entre ellas. Ello permite configurar un esquema independiente de los valores concretos almacenados en el modelo que va más allá y que se sustenta en la naturaleza de la información climática y en las relaciones intrínsecas que existen entre las tablas que contienen la información. Ambos elementos nos garantizan por tanto la universalidad del modelo (figura II.1).

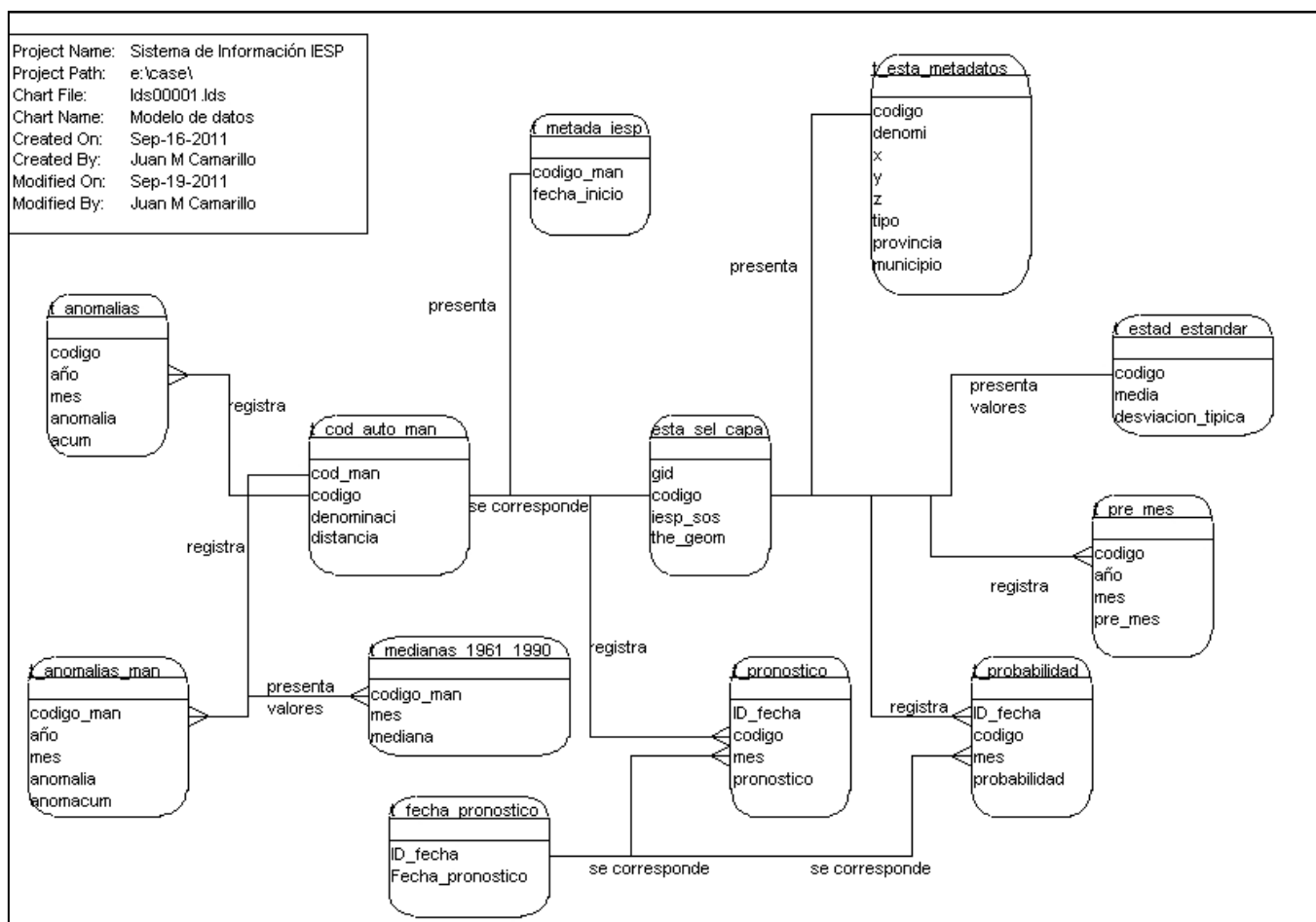


Figura II.1. Modelo de datos IESP

Fuente: elaboración propia

II.1.4. Diseñar una metodología operacional completa que sirva de referencia para el diseño de sistemas de alerta sobre sequías pluviométricas, poniendo especial atención en la jerarquía y organización de las fases que deben ir cubriéndose para alcanzar el funcionamiento del Sistema.

Cada una de las fases diseñadas, que se corresponderán con la propuesta metodológica presente en este trabajo de investigación, pretende cubrir el esquema completo necesario para el diseño de estos sistemas de alerta, desde el análisis funcional del mismo a la fase de establecimiento de los procedimientos necesarios para la difusión de los indicadores como servicios interoperables en la Web, pasando por la necesaria definición del modelo relacional de los datos. Al mismo tiempo, en cada una de las fases será importante analizar las implicaciones que cada una de ellas conlleva en relación con la naturaleza de la información climática, sus dificultades de

tratamiento o la potencialidad en el uso y difusión de la misma, pretendiéndose de esta forma conseguir un corpus metodológico completo que pueda servir de referencia y asistencia en el diseño de sistemas de alerta o de sistemas de indicadores climáticos.

II.1.5. Conseguir enmarcar este trabajo en un esquema conceptual más amplio de investigación aplicada.

Sin querer entrar en disquisiciones sobre las diferencias existentes entre la investigación básica y la investigación aplicada y partiendo de la base de que ambas son absolutamente necesarias y en la mayoría de los casos complementarias, es objetivo de este trabajo intentar lograr la puesta en marcha y funcionamiento de un sistema de alerta por sequía pluviométrica operativo en el marco de la Consejería de Medio Ambiente de la Junta de Andalucía. La consecución de este objetivo en cuyos cimientos existen factores muy importantes de conocimiento e investigación acerca de los procesos de registro, naturaleza y explotación de la información climática, pretende cubrir un nicho adicional: no solamente se persigue poner las bases metodológicas necesarias para el diseño de estos sistemas en el contexto tecnológico actual, sino que además se pretende que, como resultado de lo anterior, se produzca una transferencia real y tangible del conocimiento acumulado a los sistemas de gestión de un organismo público al que compete la gestión de los recursos naturales.

Esta perspectiva, normalmente infrecuente en los trabajos de investigación de nuestra disciplina, pasa ahora a ser un objetivo primordial que por sí mismo, en tanto que obtenga como resultado un sistema operativo y funcional, servirá de evaluación y test de los procedimientos diseñados sobre el papel.

II.2. Objetivos específicos

Como objetivos extendidos de los expuestos anteriormente caben destacar aquellos que particularmente se relacionan con alguna de las fases de diseño del Sistema y que, por su especial relevancia, deben merecer un apartado en este epígrafe.

II.2.1. Diseño y evaluación de un procedimiento y propuesta metodológica para la interpolación de lagunas de precipitación diaria en Andalucía.

Y decimos procedimiento y propuesta metodológica porque participa de ambas esferas o perspectivas; es, por un lado, un procedimiento funcional en rutina que se inserta dentro del esquema operacional del Sistema y, al mismo, tiempo está sustentado por el diseño de una propuesta metodológica completa que interpola los valores diarios de precipitación necesarios para el cálculo del valor mensual de la precipitación en las estaciones de referencia sobre las que se calcula el IESP.

II.1.2. Determinación de un procedimiento de cálculo de la probabilidad y pronóstico del comportamiento del IESP en una ventana temporal de 1 a 12 meses.

Este objetivo añadido se sustenta en la necesidad de responder, en el contexto de un sistema de alerta de sequía pluviométrica, a la proyección en el tiempo que puede esperarse del comportamiento del fenómeno analizado. Por lo tanto se aportará una propuesta de cálculo de pronósticos de comportamiento del IESP y de la probabilidad de estos, así como de los procedimientos necesarios para la integración de esos cálculos en el modelo relacional de datos y en el esquema operativo del Sistema.

II.2.3. Integración del Sistema de Información de Alerta de Sequías pluviométricas (SIAS) propuesto dentro de un marco operacional más amplio como el que configura la propuesta del futuro Sistema de Indicadores Climático-Ambientales de la Consejería de Medio Ambiente de la Junta de Andalucía.

PARTE SEGUNDA: PLANTEAMIENTOS METODOLOGICOS Y DESARROLLO DE LAS FASES DE INVESTIGACION

CAPITULO III. HIPOTESIS DE PARTIDA Y PROPUESTA METODOLOGICA

III.1. Hipótesis de partida

Las hipótesis de partida del trabajo de investigación son las siguientes:

- A) El cálculo y difusión de un Indicador ambiental de estado en tiempo real como es el Índice Estandarizado de Sequía Pluviométrica es imposible si no se enmarca en un Sistema de Información que recolecte los datos necesarios para su cálculo, los organice en una estructura previamente diseñada, tenga diseñados los procedimientos necesarios para garantizar el flujo de información y, finalmente, tenga establecidos los protocolos necesarios para la difusión del mismo.

- B) La Teoría General de los Sistemas es un marco teórico adecuado, así como una herramienta de análisis útil para integrar y explotar la información climática, abarcando el ciclo vital completo de la misma, desde su producción al nivel básico del registro meteorológico hasta la fase final de difusión de los productos derivados de la misma como son los indicadores climáticos.

- C) Los procesos implicados en el cálculo de indicadores climáticos comparten una serie de requerimientos procedimentales entre los que cabe destacar por su importancia y necesidad los siguientes:

- Diseño de un modelo de datos previo que garantice el correcto flujo de información orientado al cálculo de un indicador climático.
- Incorporación a ese modelo de datos de la componente espacial, convirtiéndolo, de esta manera, en un modelo de datos geográfico.
- Necesidad de diseño de procesos de interpolación espacial específicos que deben quedar insertados en el sistema.
- Necesidades de diseño de semiologías gráficas adecuadas para su correcta interpretación y difusión.

D) La validez y aportación del papel de geógrafo como especialista de la información climática, de su naturaleza, de los procesos involucrados en su explotación y en la producción de indicadores climáticos, así como en el diseño completo de sistemas de información climática basados en la tecnología actual.

III.2. Propuesta metodológica y fases del diseño del Sistema de Información propuesto

La elección de la metodología así como el establecimiento de las hipótesis de partida que van a aplicarse y que van a sustentar el desarrollo del trabajo es seguramente una de las etapas más importantes en la definición de un trabajo de investigación. El concepto de metodología hace referencia a los métodos de investigación que permiten lograr ciertos objetivos en una ciencia. La metodología es parte del proceso de investigación (**método científico**) que posibilita la sistematización de los métodos, de las fases y de las técnicas necesarias para llevarla a cabo. En otras palabras, la metodología es una etapa específica que procede de una posición teórica y epistemológica para la selección de técnicas concretas de investigación. La metodología, entonces, depende de los postulados que el investigador crea que son válidos, ya que la acción metodológica será su herramienta para analizar la realidad estudiada (Gutiérrez, 2006).

Según esta definición, la estructura metodológica de cualquier trabajo de investigación, como lo es una tesis doctoral, debe además sustentarse en un marco teórico y epistemológico más amplio que signifique la validez del propio método empleado. Es la inserción de ese conjunto de fases secuenciales y procedimientos analíticos que conforman la metodología del trabajo de investigación en cuestión, en un marco teórico aceptado y validado, la que va a posibilitar la validación de los resultados alcanzados en la fase final de la tesis.

III.2.1. *El marco teórico y metodológico*

La definición del objeto de estudio planteado en este proyecto de tesis doctoral, un Sistema de Información Climática, exige, en primer lugar, proceder a una definición ajustada del mismo en aras a centrar la naturaleza intrínseca del fenómeno.

Se entiende por Sistema de Información (SI) un conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar la toma de decisiones (Laudon y Laudon, 2004); o también, un conjunto de personas, datos, procesos y tecnologías de la información que interactúan para recoger, almacenar y proveer la información necesaria para el correcto funcionamiento de una organización (Whitten, Bentley y Dittman, 2004).

El exponencial aumento de las necesidades de información en las sociedades actuales y, especialmente, en los procesos de toma de decisiones y de gestión, ha sido sin duda, según numerosos autores (Castells, 2001), (Fugueras *et al*, 2004), (Mattelart, 2002) el vector principal que caracteriza el desarrollo de las sociedades modernas actuales.

Esta necesidad de información se ha visto paralelamente acompañada por un aumento sin precedentes en el desarrollo de los niveles de producción de ésta en términos de satisfacción de las crecientes demandas y necesidades. Este es el contexto de la aparición de los Sistemas de Información (SI) como vehículo y herramienta teórica, conceptual y física, diseñados para responder a las necesidades

de producción, almacenamiento, interrelación, análisis, explotación y difusión de la información en la era digital.

Analizar los elementos claves de las definiciones anteriores, nos ayudará a definir y centrar la elección del marco teórico general en el que se sustenta este trabajo de investigación aplicada. Sistema e información son los dos conceptos claves que aparecen en la definición de los objetivos anteriormente marcados.

El primero de ellos, Sistema, nos lleva al marco teórico general de la Teoría General de los Sistemas –TGS- formulada originariamente por Bertalanffy a partir de numerosas publicaciones entre 1948 y 1968. La definición derivada de los estudios de Bertalanffy considera un sistema como el “conjunto de dos o más elementos interrelacionados entre sí que trabajan para lograr un objetivo común”. El corpus teórico de la Teoría General de los Sistemas –en adelante TGS- fue ampliamente acogido por la comunidad científica debido fundamentalmente a varias razones:

- supuso un intento exitoso de superar algunos fundamentos mecanicistas de la física que no explicaban el funcionamiento de procesos complejos en el ámbito de la biología
- incorporaba el papel preponderante de las interrelaciones entre los distintos componentes del Sistema y de la noción de funcionalidad asociada a las mismas, convirtiéndola en un marco –metalenguaje- ideal para su aplicación en disciplinas tan diversas como la biología, el conjunto de las ciencias naturales, las teorías de la información, la psicología social, etc...
- permitió la posibilidad de extender su aplicación, ya que la TGS surgió debido a la necesidad de abordar científicamente la comprensión de los *sistemas concretos* que forman la realidad, generalmente complejos y únicos, resultantes de una historia particular, en lugar de *sistemas abstractos* como los que estudia la física (figura III.1).

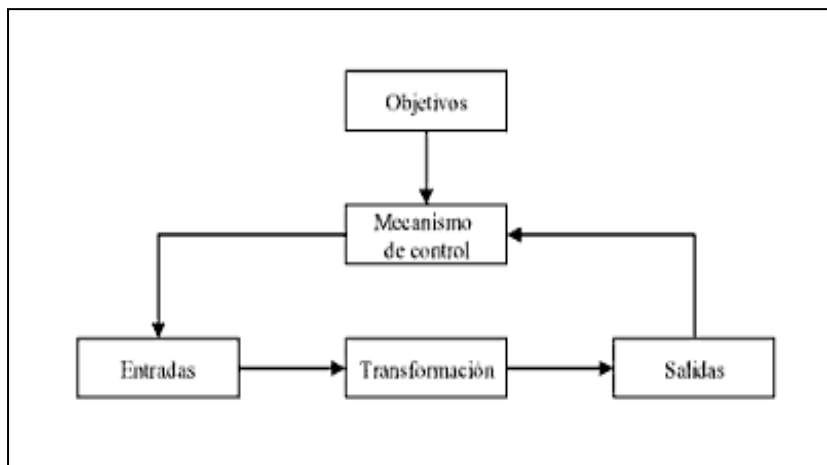


Figura III.1. Esquema conceptual de sistema

Fuente: Fernández Alarcón, 2006

El segundo concepto implicado es el de información. Este es un concepto polisémico de difícil abordaje y para el que existen numerosas aproximaciones. De hecho, incluso en los trabajos pioneros de C.E. Shannon, especialmente en su libro "La teoría matemática de la comunicación" de 1949, escrito en colaboración con W.E. Weaver y que da inicio a la ciencia moderna de la información, la cuestión de la definición del concepto no queda resuelta. Tampoco M. Castell en su monumental trilogía "La Era de la Información" ofrece una definición propia sino que se remite en una nota del primer volumen (Pág. 43) a dos definiciones ajenas: la de Machlup (1962): "La comunicación del conocimiento" y la de Porat (1977): "La información son los datos que se han organizado y comunicado".

De ambas definiciones es la segunda la que presenta una serie de matices que encajan bien en la concepción global de este trabajo, que se va a sustentar en la colección de unos datos organizados a partir del diseño de una estructura –modelo de datos- y que van a producir elementos nuevos que serán comunicados o difundidos. Para completar estas acepciones podemos incorporar la definición que realiza el Diccionario de Negocios (*Business Dictionary*) de la *WebFinance, Inc* (2011) que define información como "un conjunto de datos sobre los que se ha verificado su

fiabilidad y actualidad, que están específicamente organizados para un propósito o función, que se presentan dentro de un contexto que les da su significado y relevancia y que proporcionan un incremento del conocimiento y una disminución en las incertidumbres”.

Desde ambas perspectivas, la aplicación de la TGS al concepto de información supuso de hecho uno de los campos de mayor desarrollo que alcanzó la extensión de esta teoría general y que ha dado lugar a nuevas perspectivas en el desarrollo epistemológico de la misma.

Esta aplicación, de forma sucinta, se basa en los siguientes elementos claves, que serán la base teórica sobre el que descansa el esqueleto de este trabajo de investigación:

1. Definición de Sistema de Información (SI) como un conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar la toma de decisiones. Se trata de incidir en el concepto de diseño inteligente en el que la presencia del especialista en la información, en este caso un climatólogo, juega un papel esencial.
2. La importancia de la detección y concreción de los subsistemas implicados en la definición del sistema de información en su conjunto, así como de la intensidad y el sentido de los flujos de intercambio que, entre ellos, van a determinar la componente funcional del mismo. Aplicados estos conceptos a las ciencias de la información, es el diseño del modelo de datos basado en las componentes entidad-relación, el elemento principal que garantiza el enfoque sistémico aplicado.
3. Inherentes a la teoría general, existirán una serie de conceptos subyacentes que nos ayudarán a la conceptualización necesaria del sistema adoptado:

- *Propósito u objetivo*: se trata de uno de los conceptos más importantes encuadrado en la teoría de sistemas y que incluso aparece en la definición original de Bertalanffy. No es posible atender al marco teórico de la TGS sin proceder a la definición del objetivo común que preside la propia existencia de los sistemas. En nuestro caso, un sistema de información, su definición, se sustentará en el objetivo y propósito de garantizar el almacenamiento, la interrelación y la producción y difusión de nueva información a partir de un modelo de la realidad que integre de forma efectiva todos los flujos de información necesarios entre los distintos componentes del mismo.
- *Globalismo o totalidad*: basado en la concepción orgánica de los sistemas, se formula según la asunción de que cualquier proceso que afecte o modifique alguna de las partes del sistema, tendrá consecuencias en el conjunto del mismo. En nuestro caso, los cambios que puedan introducirse en el modelo de datos adoptado, producirán invariablemente un cambio en los flujos de información que conllevarían un funcionamiento automáticamente diferenciado del conjunto del mismo.
- *Entropía*: la entropía, concepto primeramente derivado de la segunda ley de la Termodinámica (Clausius, 1950), y extendido después a la mecánica estadística por Ludwig Boltzmann y a la teoría de la información (Shanon, 1949), se concibe como una medida del desorden de los sistemas. Desde esta perspectiva y aplicada a los sistemas de información, a medida que aumenta la información disminuye la entropía, alejándose por tanto los sistemas del desorden y disminuyendo el grado de aleatoriedad de los mismos. Surge de esta forma el concepto de neguentropía aplicado a los sistemas de

información bien desarrollados y maduros, los cuales consiguen disminuir la aleatoriedad en los flujos informativos con la consiguiente disminución de la entropía como medida del caos o desorden del sistema.

- *Retroalimentación*: los sistemas de información se caracterizan por la presencia de fenómenos de retroalimentación *-feed-back-* que garantizan el “aprendizaje” del sistema, así como su proceso de maduración y la búsqueda de disminución del grado de entropía.

 - *Homeostasis*: es el equilibrio dinámico entre las partes del sistema, la búsqueda de un equilibrio interno que responde a las necesidades de adaptación frente a los cambios externos del medioambiente y que se relaciona con las fases de crecimiento en el contexto de la información y de los sistemas que la gestionan y que deben, por tanto, tenerse en cuenta en las fases embrionarias de diseño de los mismos.
4. Metodología de diseño y análisis de sistemas complejos: desde una perspectiva aplicada, el marco teórico de la Teoría General de los Sistemas ha sido extendido al análisis y diseño de sistemas complejos de información, desarrollándose, de esta forma, toda una serie de propuestas metodológicas que desgranaremos en el siguiente capítulo y que constituyen la base del presente trabajo de investigación.

III.2.2. Propuesta metodológica y fases

Como ya se ha afirmado, el núcleo central de un trabajo de investigación es la selección y validación de la metodología seleccionada para cumplir los objetivos de dicha investigación. Es por ello que este apartado merecerá especial atención en el desarrollo de la tesis.

Si entendemos la metodología como los diferentes métodos de investigación que permiten lograr los objetivos de una actividad científica, debemos obligatoriamente enmarcar a esta dentro de una posición teórica y epistemológica apriorística como la que ha sido defendida en el epígrafe anterior. Además, de esta posición o marco teórico adoptado es de donde se deriva la selección de técnicas concretas de investigación. La metodología, entonces, depende de los postulados que el investigador crea que son válidos, ya que la acción metodológica será su herramienta para analizar la realidad estudiada (Gutiérrez, 2006).

Por lo tanto, si la metodología a adoptarse y el marco teórico son indisociables, será necesario realizar un breve recorrido previo por el conjunto de metodologías desarrolladas en el contexto del análisis y diseño de sistemas de información, entendidos ambos como rama de aplicación derivada del desarrollo de la Teoría General de los Sistemas.

El importante desarrollo de las metodologías de ingeniería de software y de sistemas ha producido una amplia panoplia de opciones de diseño. En la actualidad podemos agruparla en cuatro categorías básicas:

A) Metodologías Estructuradas o en cascada (*waterfall models*)

Basadas en una concepción lineal del conjunto de fases que integran el desarrollo de los proyectos de ingeniería de sistemas. Son las metodologías originarias que derivaron de la aplicación de la TGS a los sistemas de información. Se conciben como un proyecto completo en el que deben verificarse cada una de las fases del mismo antes de pasar a la siguiente, normalmente implican un desarrollo largo en el tiempo de ejecución, suelen

abarcar varias áreas de negocio de la organización y están orientadas al manejo de un gran volumen de datos y de procesos.

La adopción de este enfoque estructurado ha dado lugar al desarrollo de metodologías específicas de entre las que cabe destacar las siguientes, así como a los siguientes autores:

- *SSADM (Structured System Analysis and Design Methods, National Computing Center, UK, 1995).*
- *STRADIS (Structured Analysis, Design and Implementation of Information Systems, Gane and Sarson, 1979).*
- *YSM (Yourdon Systems Method, Yourdon, 1993).*
- *MERISE, Quang and Chartier-Kastler, 1991.*
- *METRICA, Ministerio de Administraciones Públicas, España.*
- Kendall & Kendall (2005).
- Long (1997)

Todos ellos abarcan el conjunto de fases del proyecto, desde los estudios de factibilidad a la implementación física del sistema y su validación posterior, pasando por el análisis de los requisitos del sistema, el diseño propiamente dicho, la documentación o el control de proyectos.

B) Evolutivas-Incrementales:

Se deriva de las primeras en cuanto a la definición de las fases propuestas, si bien permiten seguir secuencias ascendentes o descendentes en las etapas de desarrollo, permitiendo, de esta manera, cumplimentar fases de desarrollo del proyecto en paralelo (figura III.2).

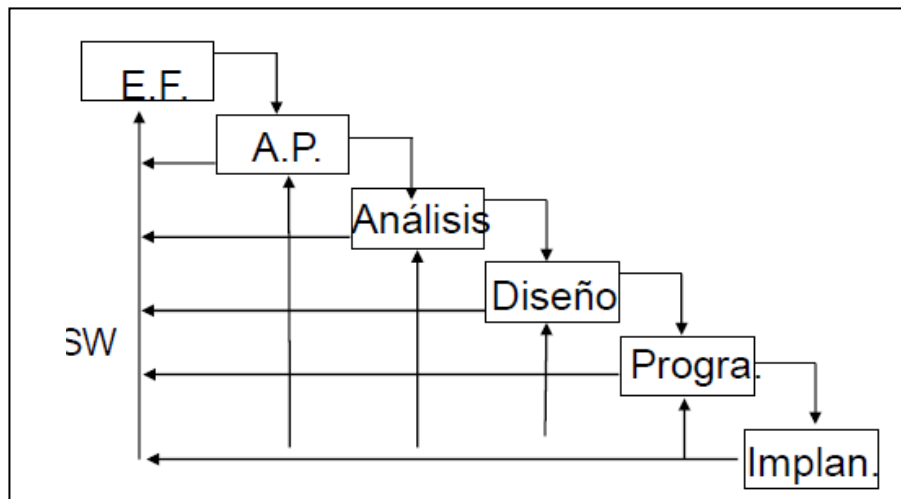


Figura III.2

Fuente: Fernández Alarcón, 2006

C) Prototipos:

Basados en el desarrollo de un modelo a escala del sistema propuesto (prototipo) que identifica las necesidades de información. Este modelo maneja las transacciones básicas de información y los procesos básicos en los que aquella se encuentra implicada. Permite consultas básicas al sistema y se sustenta en una mejora y desarrollo cada vez más avanzado de los prototipos originales.

De entre ellas destaca el avance del método *DSDM* (*Dynamic Systems Development Method*) que constituye un marco de trabajo completo para gestión y desarrollo de proyectos basado en la tecnología *RAD* (*Rapid Application Development*) cuyo objetivo principal es la reducción de costes y la revisión y actualización permanente de las distintas fases de desarrollo del sistema (figura III.3).

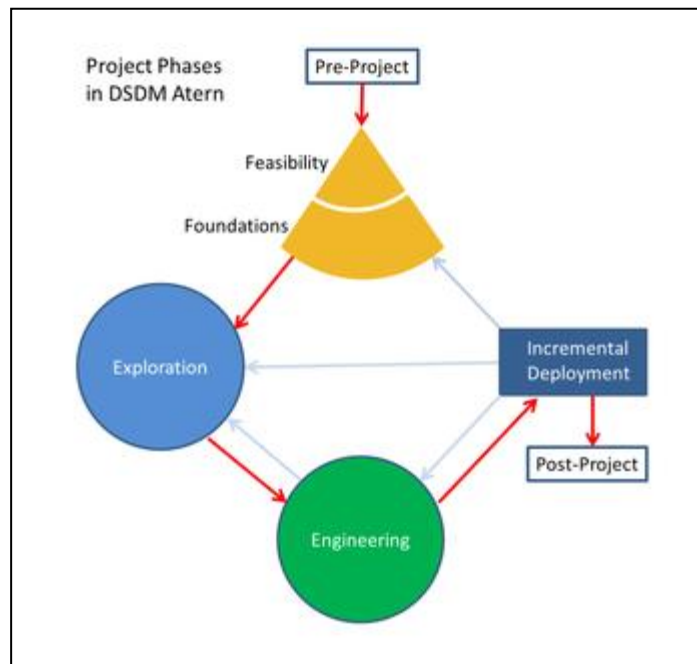


Figura III.3

Fuente: Cockburn, 2011

D) Orientada a objetos (*Object Oriented*):

Seguramente la metodología más avanzada y en mayor auge que existe en la actualidad debido a la gran versatilidad que ofrece así como en los altos niveles de reutilización, adaptabilidad y grado de abstracción que subyacen en su propuesta.

Se trata de una metodología que no modela la realidad, sino la forma en que las personas comprenden y procesan la misma. Basada en un proceso ascendente que se sustenta en la abstracción de clases en aumento. Sus conceptos básicos son:

- Objeto: módulo que contiene los datos e instrucciones que operan sobre los datos
- Mensaje: solicitud que se hace al objeto para que se comporte de alguna forma

- Métodos: determinan cómo actúa el objeto al recibir un mensaje
- Clase: descripción de un conjunto de objetos casi idénticos, que consta de métodos y datos que resumen sus características comunes
- Herencia: mecanismo para compartir métodos y datos entre clases, subclases y objetos

En este sentido son de común utilización las herramientas *UML –Unified Modelling Language–*, no exclusivas del diseño orientado a objetos si bien muy focalizadas en el mismo.

En este punto es importante recordar los objetivos del trabajo de investigación ya que la decisión de la metodología a adoptarse se encuentra íntimamente ligada a los mismos.

Y el principal de ellos es, además del diseño de un sistema de información de alerta para sequías meteorológicas en Andalucía, la investigación de las especificidades que, en el diseño del mismo, introduce, en forma de requisitos, la información de naturaleza climática y/o geográfica.

De este hecho se deriva la diferenciación que debemos hacer en estos momentos entre la materialización práctica de un proyecto de ingeniería de sistemas –que corresponde a la esfera del encargo práctico que realizó la Consejería de Medio Ambiente de la Junta de Andalucía- y las consideraciones en materia de investigación que conciernen a este proyecto de tesis doctoral.

Es por ello, que, para centrar la investigación en aras de cumplir los objetivos propuestos, vayamos a focalizarnos en el núcleo central de todo proyecto de ingeniería de sistemas, que no es otro que la fase de *análisis y de diseño* del mismo (figura III.4). La razón de esta elección es que, en dicha fase es en la que se hace más evidente el maridaje de las dos esferas que constantemente aparecen desde el inicio de este trabajo, la información de naturaleza climática –el IESP- y su encaje operacional en el contexto indispensable de un sistema de información.

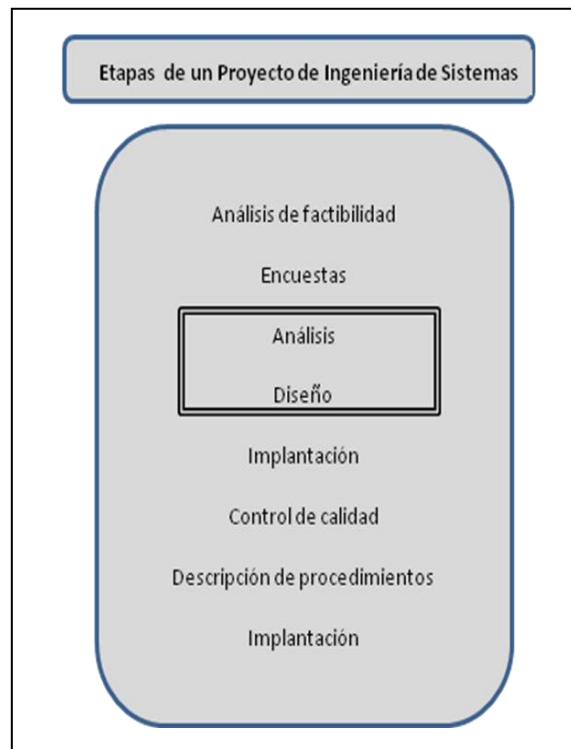


Figura III.4 Fuente: elaboración propia siguiendo a Yourdon. Metodología estructurada

La elección de la metodología que va a servir de base para la consecución de los objetivos propuestos, de entre las anteriormente mencionadas, es uno de los métodos estructurados clásicos, el *SSADM* (*Structured System for Analysis and Design Methods*, NCC, 1995).

La elección de la misma se sustenta en las siguientes premisas:

- Se trata de una metodología consolidada en los entornos de ingeniería de sistemas y supone un estándar internacional.
- Presenta un alto grado de abstracción, especialmente en las fases de análisis y diseño del sistema, lo cual se adapta perfectamente al objetivo de investigación principal que no es sino poner de manifiesto las especificidades de la información de naturaleza climática en el diseño de sistemas de información.
- Se trata de una metodología estructurada en el que las fases quedan bien definidas

- Utiliza herramientas gráficas que permiten sintetizar la lógica del sistema, sus funcionalidades, procesos y flujos de información implicados en estos. De entre ellas cabe destacar los *DFD (Data Flow Diagram)* como técnica gráfica principal que será utilizada en las distintas fases de desarrollo del sistema.
- Es una metodología soportada por la utilización de herramientas *CASE (Computer Aided Software Engineering)*, conjunto de aplicaciones informáticas específicamente diseñadas para la asistencia a los procesos de diseño de sistemas de información o de desarrollo de *software*.

III.2.2.1. La metodología *SSADM* y las fases de desarrollo de sistemas de información

La metodología *SSADM* se basa en la secuenciación de 6 etapas o fases en el desarrollo de un sistema de información. Estas etapas son las siguientes:

- A) Etapa 0. Estudio de viabilidad (*Feasibility Study*): se trata de la fase de investigación acerca de la viabilidad técnica y económica del sistema propuesto. Esta etapa es la de definición general del proyecto, la identificación de usuarios, la definición de las problemáticas asociadas y las discusiones generales relativas a las opciones técnicas y económicas que harán posible la configuración del sistema.
- B) Etapa 1. Investigación del entorno actual (*Investigation of the Current Environment*): en esta etapa se definen los requisitos generales, así como la estructura de los datos y de los procesos que deberán ser tenidos en cuenta en la fase de análisis y diseño. Comienza al mismo tiempo a definirse el proyecto, a calcularse los costes en tiempo y económicos que consumirá la puesta en marcha del sistema de información.

- C) Etapa 2. Opciones de negocio del sistema (*Business System Options*): o plan de negocio del sistema, se basa en la formulación de los requisitos de negocio del sistema, así como en la evaluación de las implicaciones y beneficios de las opciones propuestas.
- D) Etapa 3. Especificación de requisitos (*Requirements specification*): en esta etapa se producen la identificación de los requisitos funcionales y no funcionales detalladamente. Aparecerán ahora las propuestas de los métodos y las técnicas necesarias para describir el conjunto de los procesos y procedimientos implicados en el flujo de información, así como la estructura de los datos. Culmina con la producción del diseño lógico del sistema propuesto. Es quizás la etapa crucial del conjunto de fases implicadas en el diseño de un sistema de información ya que el resultado de la misma no es otro que la formulación lógica del funcionamiento del sistema en su conjunto. Para ello se utilizan técnicas de modelado que mantienen el suficiente nivel de abstracción como para representar la realidad implicada en el sistema concreto que se está diseñando.
- E) Etapa 4. Opciones técnicas del sistema (*Technical Systems Options*): en esta etapa se procede a la definición y selección de las opciones técnicas específicas, así como de los diferentes métodos de implementación de las mismas.
- F) Etapa 5. Diseño lógico del sistema (*Logical Design*): puede realizarse simultáneamente a la etapa 4. Se definen en esta etapa cuestiones como las interfaces de usuarios, los procesos de actualización de información o los procesos de consultas a los datos contenidos en el sistema.

- G) Etapa 6. Diseño físico del sistema (*Physical Design*): se trata en esta etapa final de la implementación física del sistema, basada en los requisitos analizados, el diseño lógico realizado y las opciones técnicas seleccionadas.

Para asistir a la ejecución y desarrollo de todas estas fases o etapas, la metodología *SSADM* implementa una serie de herramientas y de técnicas específicas que podrán ser utilizadas en función de los criterios y objetivos que marcan cada una de esas etapas. Sucintamente son las siguientes:

1. Modelado del flujo de datos (*Data Flow Modelling –DFM–*): Esta herramienta se centra en la identificación, modelado y documentación de los distintos flujos de información que suceden alrededor y dentro del sistema de información. El resultado será la producción de un Diagrama de Flujo de Datos (*Data Flow Diagram –DFD–*) que mostrará la conexión de los distintos procesos y procedimientos que tendrán lugar dentro del sistema de información. Este diagrama mostrará el almacén de los datos, la forma en la que estos son accedidos, la descripción de las entidades o actores externos, en términos de intercambio de datos, que interactúan con el sistema así como los flujos entre todos estos componentes.
2. Modelado lógico de los datos (*Logical Data Modelling –LDM–*): dentro de este proceso los requisitos de los datos son investigados, definidos y modelados. El elemento central de esta herramienta es la producción de la estructura lógica de los datos (*Logical Data Structure –LDS–*). Esta *LDS* es la nomenclatura específica que, dentro de la metodología *SSADM*, adopta el modelo entidad-relación (E-R) de modelado de bases de datos.

La mecánica y reglas del modelo entidad-relación, por su importancia y por constituir la base de la representación de los datos manejados por el sistema, así como de las relaciones entre entidades (tablas alfanuméricas), será abordado de forma detallada más adelante cuando se defina el modelo de datos en el desarrollo de la fase de requisitos según la metodología adoptada.

3. Modelado de eventos de entidades (*Entity Event Modelling –EEM–*): el elemento central de esta herramienta es la producción de los diccionarios de datos a modo de colección y registro de cada proceso, flujo de datos, almacén de datos y entidad externa que componen el sistema. También llamados *Entity Life History –ELH–*.

Como ya se ha señalado anteriormente, los objetivos de este trabajo de investigación se centran, para su consecución, en las fases de análisis de las especificaciones de los requisitos del sistema (etapa 3) y de opciones técnicas del sistema (etapa 4) ya que son las fases que cierran el ciclo de análisis y diseño que sirven a nuestros propósitos. El diseño lógico del sistema (etapa 5), en el que debe reflejarse la operativa del mismo en términos de herramientas administrativas, interfaces, herramientas de conexión, etc... que permitirán al administrador del mismo aplicar los procesos operacionales – comprobación datos, lanzamiento de procedimientos, salvaguarda de datos, conexión a servidor remoto, etc...-, queda al margen de este trabajo de investigación.

Serán los modelos abstractos producidos en el contexto de cada una de estas tres fases las que nos permitan investigar la inserción de los datos, los procedimientos de cálculo y los flujos de información climática necesarios para el cálculo de un indicador de estado como es el IESP. Esquemáticamente la metodología de análisis adoptada quedaría reflejada en la siguiente figura (Figura III.5).

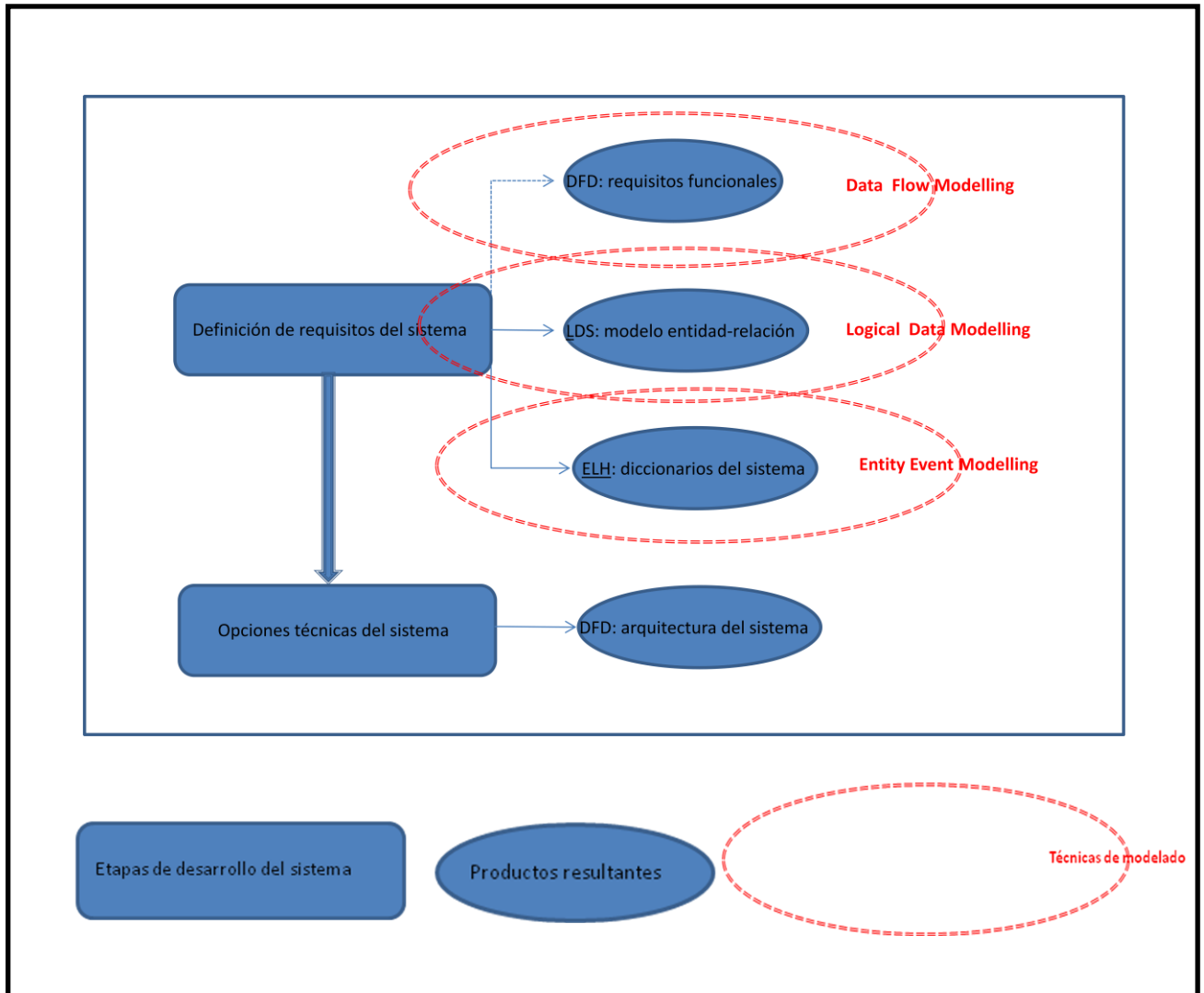


Figura III.5

Fuente: elaboración propia adaptado de metodología SSADM de análisis y diseño de sistemas de información

CAPITULO IV. ANALISIS Y DEFINICION DE LOS REQUISITOS DEL SISTEMA: MODELADO DEL FLUJO DE DATOS, MODELADO LOGICO DE LOS DATOS, DICCIONARIO DE ENTIDADES Y ARQUITECTURA DEL SISTEMA

Esta fase es la que supone un nivel más abstracto en la definición de las necesidades y requisitos en términos de información, flujo de la misma y procesos que se encuentran implicados en el Sistema.

Se basa en la ejecución de tres procesos de modelado:

- Modelado del flujo de datos (*Data Flow modelling*), que supone la generación de un Diagrama de Flujo de Datos que representa el *modelo funcional* del Sistema. En este proceso diseñaremos, mediante herramientas de asistencia *CASE* específicamente diseñadas para la metodología *SSADM*, la arquitectura funcional del sistema, poniendo especial énfasis en los procesos específicos en los que se encuentra involucrada la información de naturaleza climática y que en cuya evaluación se centran los objetivos de este trabajo de investigación.
- Modelado lógico de los datos (*Logical Data Modelling*), cuyo resultado es la generación de la estructura lógica de los datos basada en el modelo entidad-relación (E-R) clásico en la actividad de modelado de la base de datos en el que va a sustentarse la organización de la información así como la intensidad y el sentido en el flujo e intercambios de la misma entre entidades (tablas alfanuméricas).
- Modelado de eventos de entidades (*Entity Event Modelling*), cuyo resultado es la documentación de los diccionarios de entidades y eventos del sistema.

IV.1. Modelado del flujo de datos (Data Flow Modelling)

La utilización de Diagramas de Flujo de Datos es una de las herramientas metodológicas más utilizadas en el análisis y diseño de sistemas de información basadas en el modelado de la realidad. Mediante ellas se representa las relaciones entre la realidad y la forma en la que esta puede ser comprendida para trasladarse a un entorno de sistemas de computación. Se trata por tanto de una forma de comprender la realidad y llevar la manifestación física de esta a una visión lógica de la misma.

El principal propósito de esta técnica es conseguir una comunicación entre el analista y el usuario a nivel lo suficientemente abstracto como para permitir la descomposición de los problemas y del funcionamiento global y detallado del funcionamiento del sistema.

La dinámica de producción de Diagramas de Flujo de Datos se basa también en la descomposición o expansión de diferentes niveles de detalle en la modelización funcional del sistema. Esta concepción permite ir desarrollando, en la medida de las necesidades de la misma, a nivel cada vez más detallado, el funcionamiento particular de cada proceso involucrado en el sistema mediante una separación progresiva de procesos en subprocesos jerarquizados (Shumacher, 2001).

En nuestro caso se utilizarán los siguientes tipos de Diagramas de Flujos de Datos:

- Diagrama de contexto: que servirá para explicar el contexto completo del Sistema así como los componentes básicos del mismo.
- Diagramas Funcionales: a partir de los cuáles será representado el comportamiento funcional del sistema en los distintos niveles de jerarquización.

Cada uno de estos *DFD* será explicado a partir de un esquema común en forma de ficha que presenta los siguientes elementos:

- **Objetivos:** descripción de los objetivos funcionales a cubrir por el *DFD* desarrollado.
- **Datos de partida:** descripción de los datos de partida sobre los que se sustenta el *DFD*.
- **Requisitos funcionales** necesarios: descripción de los requisitos que son necesarios para el correcto flujo de información y funcionamiento de los procesos.
- **Componentes del sistema:** descripción de los componentes que aparecen en el diagrama de contexto.
- **Procesos de datos:** descripción de los procesos que aparecen en los *DFD* en las diferentes escalas producidos.
- **Flujos de datos:** descripción del flujo de datos entre componentes del sistema (diagrama de contexto) o involucrados en procesos funcionales (modelos funcionales).
- **Almacén de datos:** descripción de los almacenes de datos necesarios para el funcionamiento del sistema.
- **Entidades externas:** descripción de las entidades (bases de datos o sistemas) externas que se comunican con el sistema.
- **Requisitos y procesos de naturaleza geográfica y/o climática:** descripción de los requisitos y procesos específicos de la información de naturaleza geográfica y/o climática que fluyen por el sistema.

Estas fichas son un compendio del Diccionario de Entidades que serán incluidas como explicación de los *DFD* generados. Entendemos que, al presentarlo de esta forma, facilita la comprensión de los mismos a los miembros del Tribunal.

Al mismo tiempo, la lectura de los *DFD* generados se sustenta en la siguiente leyenda estandarizada y normalizada por el gestor de la metodología *SSADM* y desarrollada para la herramienta *CASE* utilizada (figura IV.1).

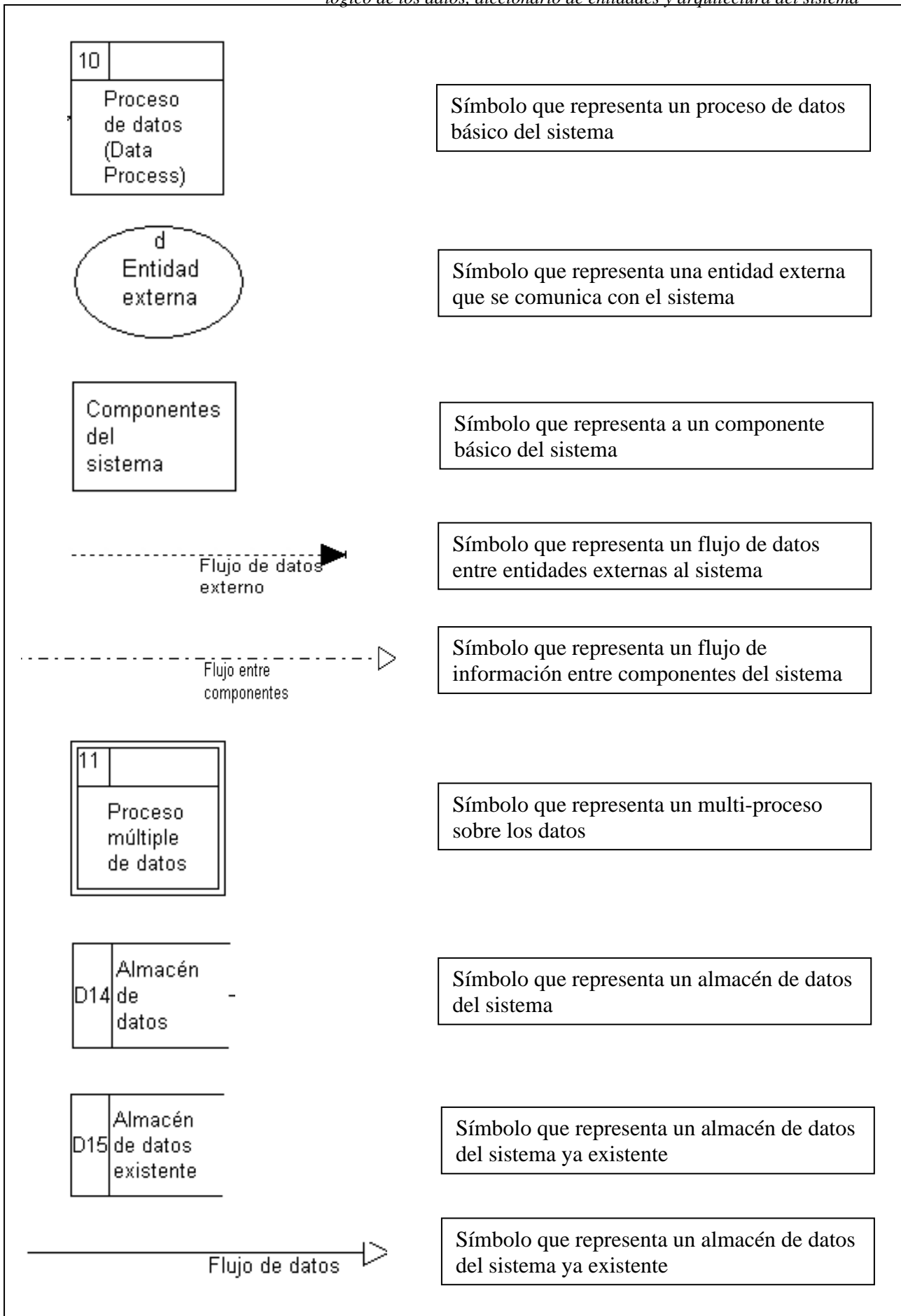


Figura IV.1. Legenda de los Diagramas de Flujo de Datos (DFD)

Fuente: elaboración propia

El esquema funcional del sistema, basado en el análisis de requisitos funcionales del mismo, queda representado por los Diagrama de Flujo de Datos que se presentan en las figuras IV.2, IV.3, IV.18, IV.19, IV.21, IV.22 y IV.23).

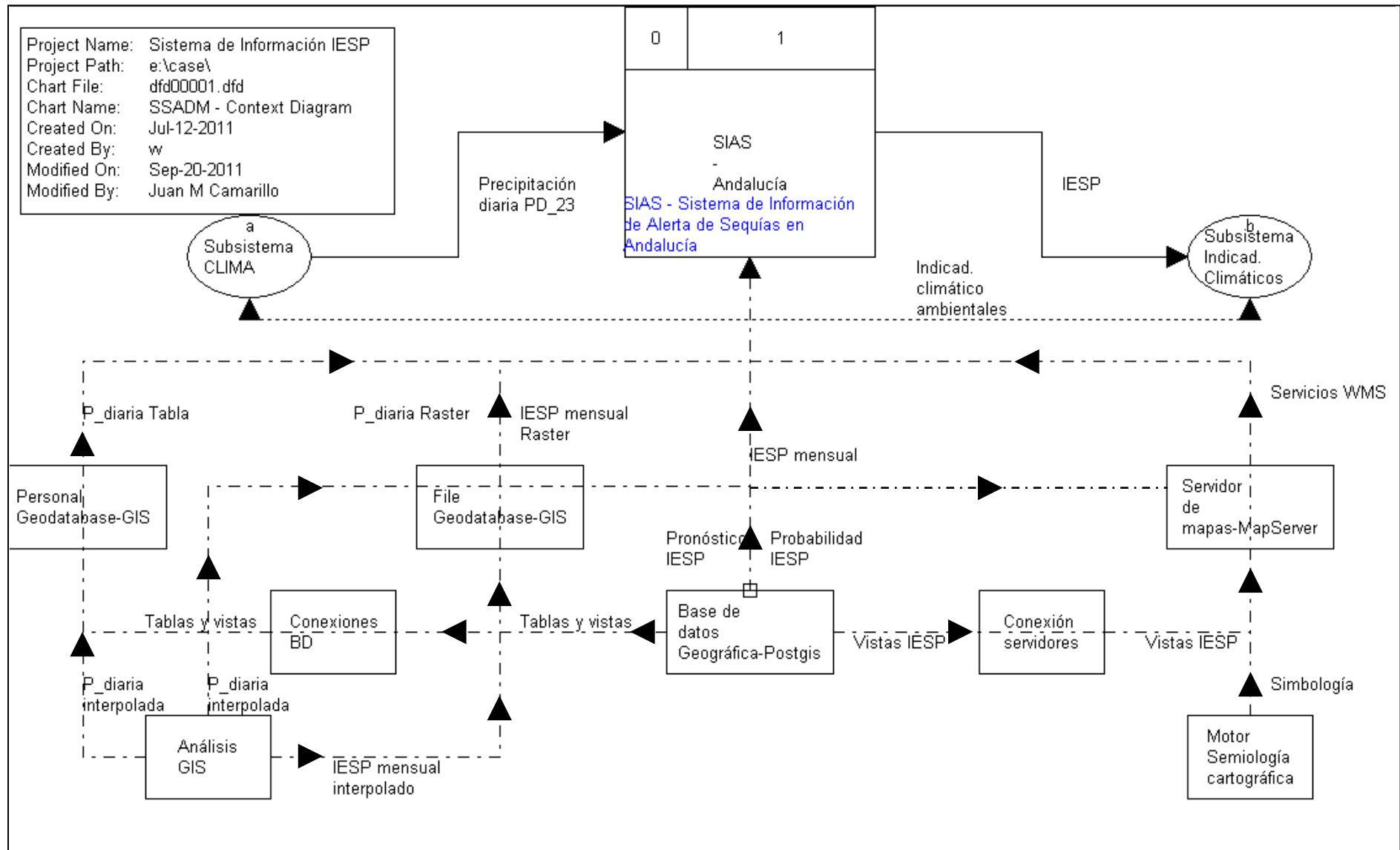
Los Diagramas de Flujo (*DFD*) desarrollados en este trabajo de investigación serán los siguientes:

- Diagrama de contexto del sistema

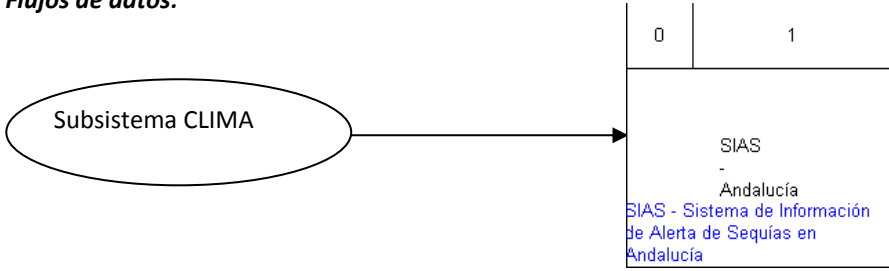
- Diagrama funcional del sistema

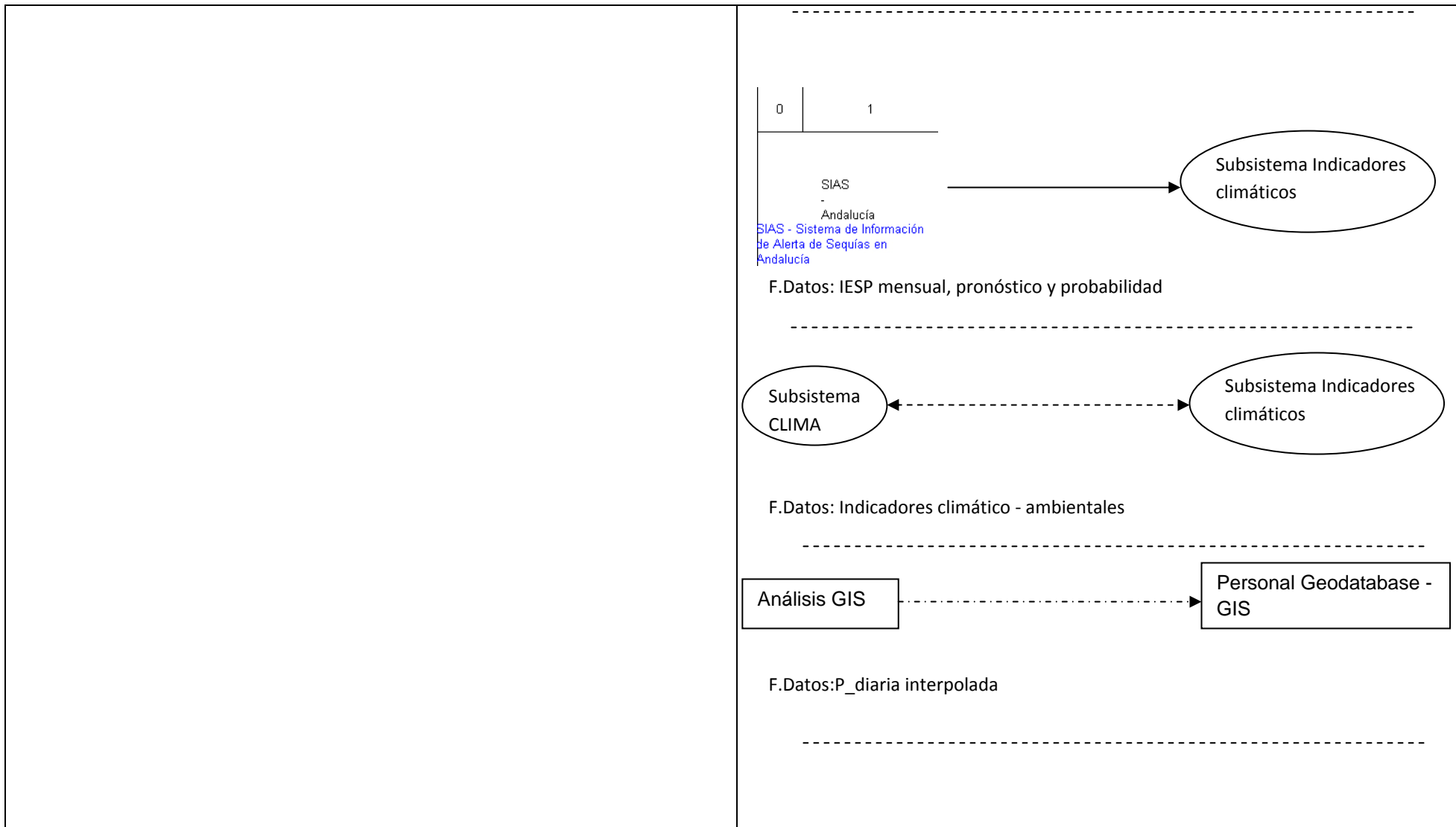
- Diagramas de los subprocesos implicados en aquellos procesos básicos del sistema que, por su especial significación para los objetivos de investigación de este trabajo y por su alta significación climática y/o geográfica, merecen ser detallados a partir de sus propios Diagramas de Flujo de Datos:
 - o Interpolación de los valores de precipitación diaria PD₂₃
 - o Espacialización del valor del IESP mensual
 - o Cálculo del pronóstico del comportamiento del IESP en 1,...,12 meses
 - o Cálculo de las probabilidades asociadas a los pronósticos
 - o Generación de los servicios WMS

IV.1.1. Diagrama de contexto del Sistema (Figura IV.2)

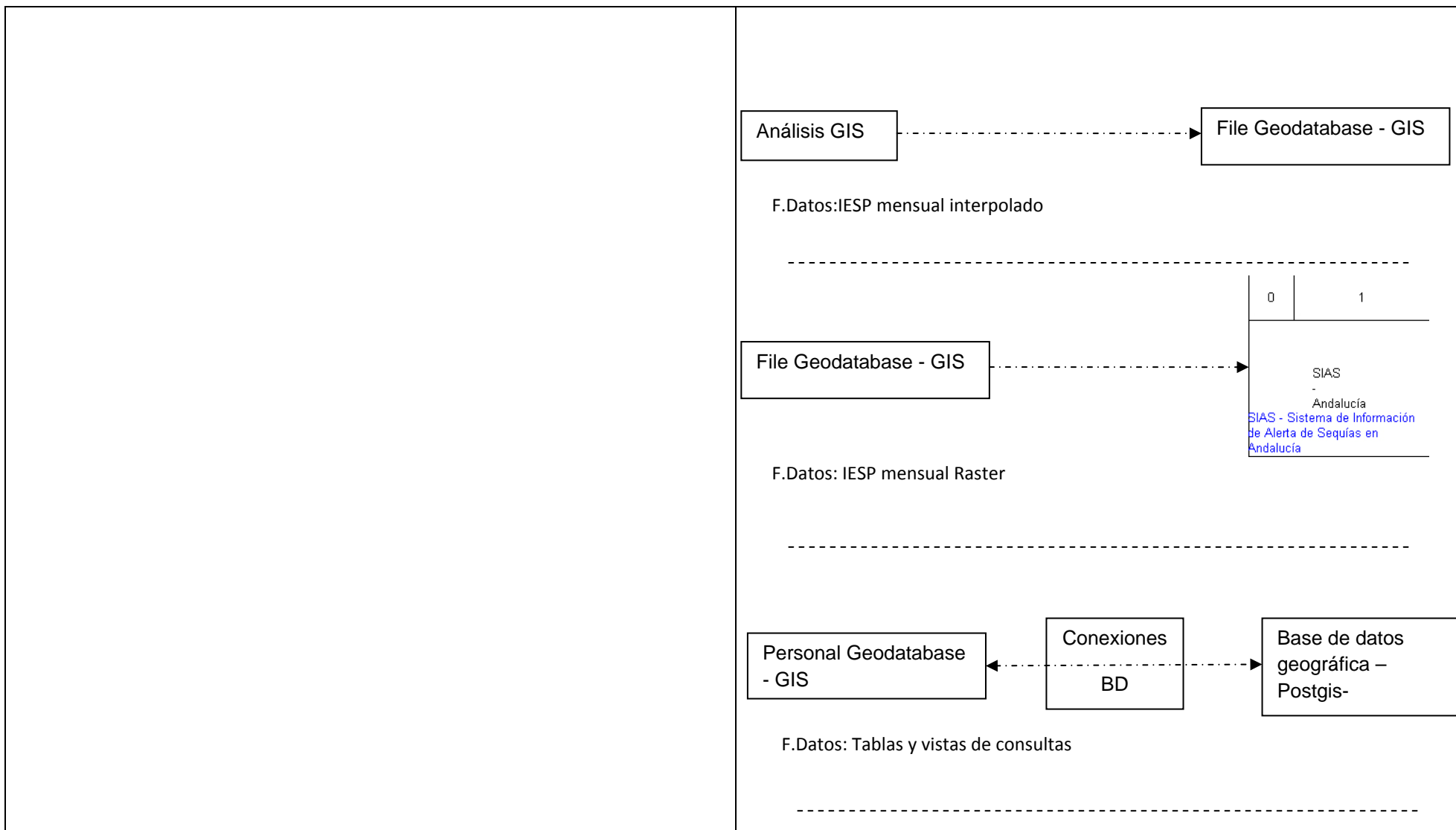


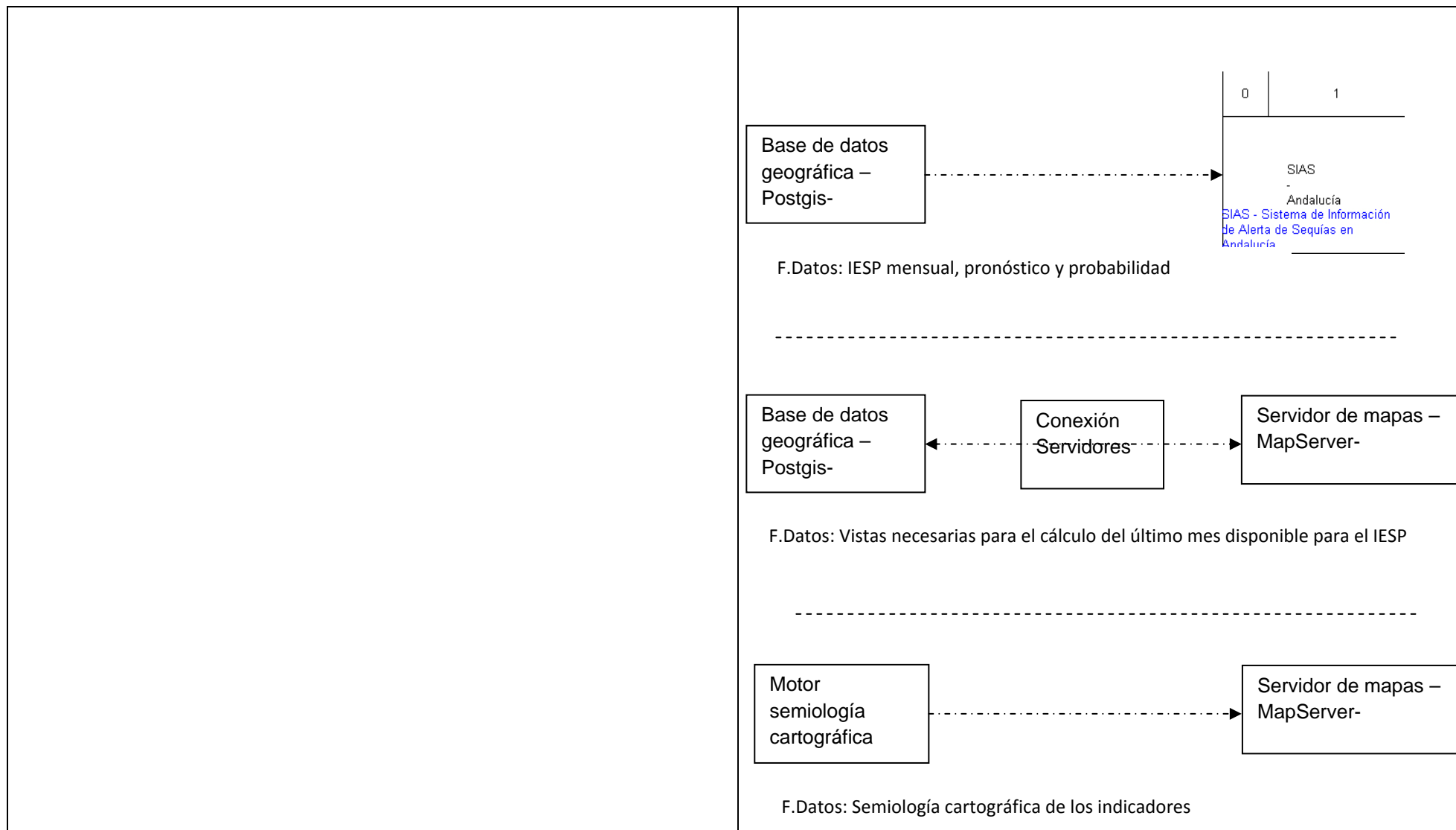
<u>DFD nº 00001 Diagrama de contexto del Sistema de Información de Alertas de Sequía Pluviométrica de Andalucía</u>	
<p>Objetivos: Representación de los componentes necesarios para el funcionamiento del Sistema y de las relaciones y flujos compartidos entre ellos. El Sistema pretende:</p> <ul style="list-style-type: none"> - Calcular el IESP mensual en tiempo real - Espacializar en Andalucía el comportamiento de dicho Índice - Calcular los pronósticos 1.....12 meses y la probabilidad asociada a los mismos - Generar las cartografías de esos Indicadores como servicios interoperables WMS 	<p>Datos de partida: Los únicos datos de partida son los valores de precipitación diaria registrados en un conjunto de 297 estaciones meteorológicas automáticas presentes en las distintas redes meteorológicas que operan en Andalucía, entre las que se encuentran las 84 estaciones automáticas seleccionadas para el cálculo del IESP. La selección de dichas estaciones se sustenta en el análisis previo de la calidad de los registros en términos de cercanía a estaciones manuales de la AEMET para las que existían datos homogéneos (Test de Alexanderson. <i>Proclim Tool</i>) y extendidos a 1950 y con un porcentaje de lagunas mensuales inferior al 20% que fueron posteriormente interpoladas mediante regresión múltiple con el grupo de estaciones mejor correlacionadas. El cálculo del IESP, mensual, necesita datos continuos de precipitación mensual que necesitaron la interpolación previa. Con ellos se procedió al cálculo del IESP en 288 estaciones manuales de referencia desde 1950 hasta la puesta en marcha del sistema. (Anexo 1).</p> <p>El cálculo del Indicador de Estado en tiempo real se realizará por tanto sobre la precipitación acumulada de los valores diarios registrados en 84 estaciones automáticas de diferentes redes almacenados en el Subsistema de Climatología Ambiental CLIMA de la Red de Información Ambiental (REDIAM) – Consejería de Medio Ambiente. Junta de Andalucía.</p>
<p>Requisitos funcionales:</p> <ul style="list-style-type: none"> - Conexión con CLIMA. - Conexión con Subsistema de Indicadores Climático-Ambientales. - Interpolación de lagunas en los valores de precipitación diaria para obtener valores mensuales. - Cálculo de IESP mensual (último mes disponible), pronósticos y probabilidades - Generar servicios interoperables wms - Generar cartografías de precipitación diaria en Andalucía para su incorporación en el Subsistema CLIMA. - Incorporar herramientas de software libre. 	<p>Componentes del sistema:</p> <ul style="list-style-type: none"> - Personal Geodatabase: la realización de procedimientos de análisis espacial (vectorial y raster) indispensables para la interpolación de las lagunas de registros en los valores de precipitación diaria procedentes de las estaciones automáticas obliga a la utilización de herramientas GIS conectadas a una base de datos de acceso y escritura de datos. El software seleccionado, ArcGis, conecta dichos entornos de análisis a una Personal Geodatabase que pasa a formar parte del Sistema como componente. - Análisis espaciales basados en tecnologías GIS: se ha incorporado como componente del Sistema ya que, independientemente de la arquitectura seleccionada, los procedimientos implicados en el cálculo del indicador IESP no podrían llevarse a cabo sin la disponibilidad de estas herramientas de análisis

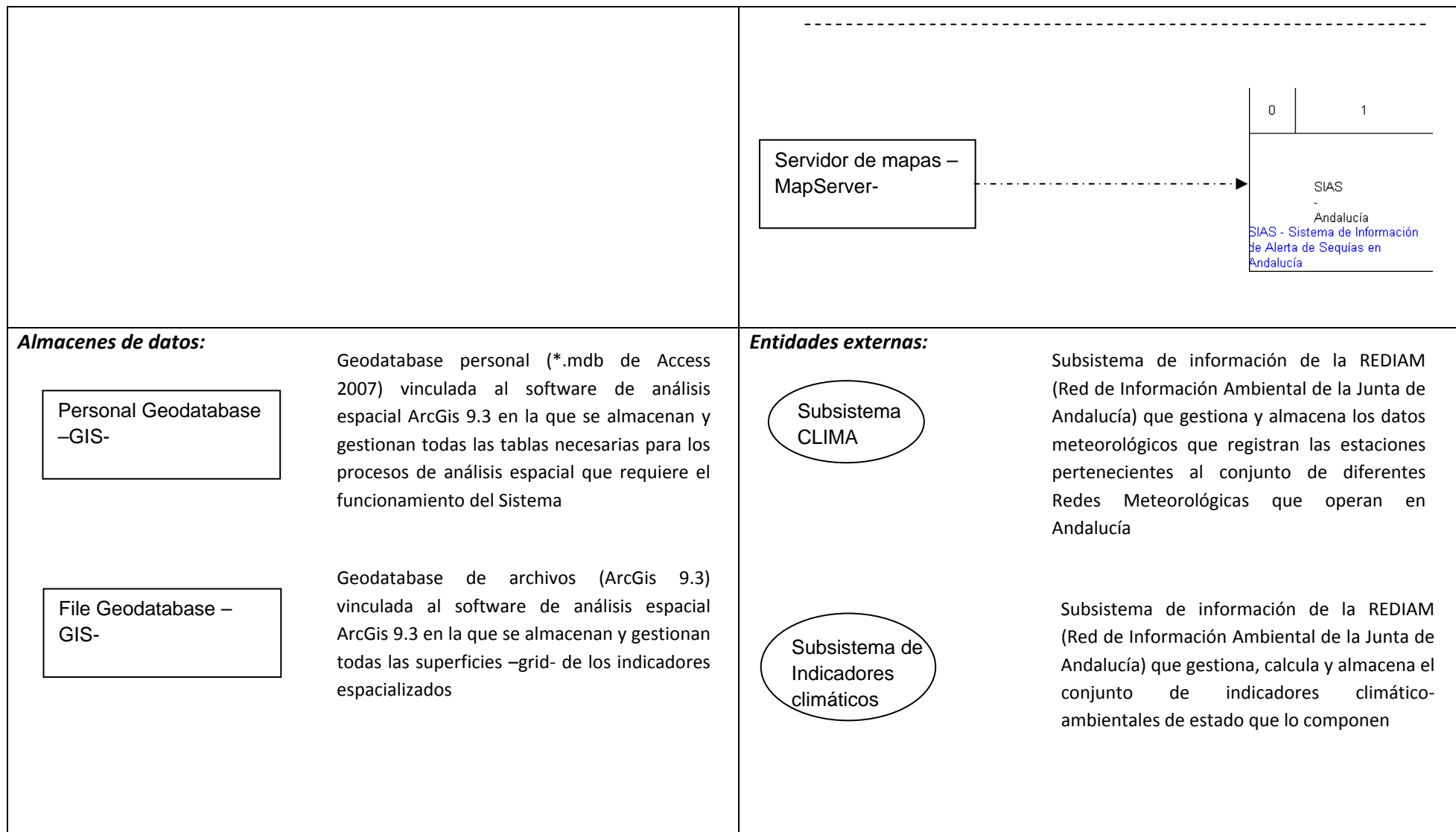
	<p>espacial, tanto vectorial como raster.</p> <ul style="list-style-type: none"> - File geodatabase: geodatabase de archivos que serán el almacén de las imágenes raster que almacenan los valores espacializados tanto de la P_diaria como del IESP mensual. - Base de datos geográfica: el cálculo del último IESP mensual disponible y su tratamiento como vista (SQL) con conexión viva a los datos de precipitación mensual de las 81 estaciones seleccionadas, necesita de un SGBD (Sistema de Gestión de Base de Datos) en el que el Lenguaje Estructurado de Consulta (SQL) disponible en las mismas sea lo suficientemente robusto en funciones. Al mismo debe permitir gestionar las geometrías puntuales de las estaciones así como la conexión con un servidor de mapas. - Conexiones Bases de Datos: se trata de un componente básico del Sistema por la necesidad de tratamiento analítico espacial en entornos GIS y de SGBD robustos en SQL y conexiones a servidores de mapas. - Servidor de mapas: componente encargado de construir los servicios interoperables WMS (Web Map Services) que serán mostrados en los clientes (visores geográficos) con los que el IESP será difundido. - Conexión base de datos a servidor de mapas: es el componente del Sistema que va a conectar la vista del IESP con el servidor de mapas productor del servicio interoperable WMS. - Semiología cartográfica: ha sido añadido como componente ya que la construcción de los servicios WMS necesitan de una correcta selección y configuración de la semiología gráfica de tratamiento del IESP previa.
<p>Procesos de datos: A partir de DFD funcionales</p>	<p>Flujos de datos:</p>  <p>F.Datos: Precipitación diaria PD_23</p>





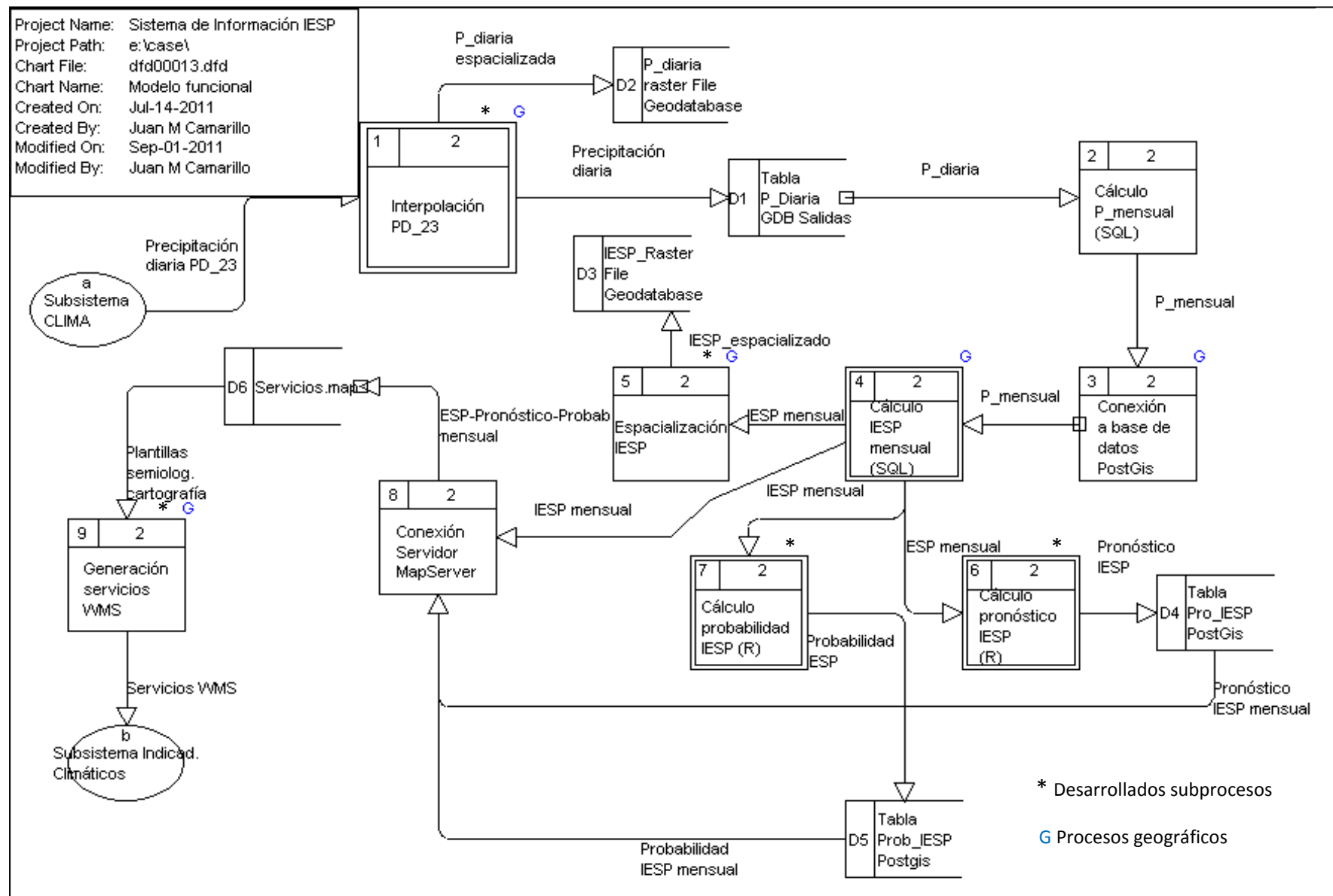






<p>Base de datos geográfica –PostGis-</p>	<p>Base de datos geográfica en la que se gestionan las tablas asociadas al modelo de datos propuesto y en la que se calculan (vistas SQL) los valores de los indicadores mensuales vinculados a las geometrías puntuales que los representan</p>	
<p>Requisitos de naturaleza geográfica y/o climática:</p> <p>La configuración de un Sistema de información de las características especificadas exige una serie de requisitos especiales vinculados al conocimiento temático tanto de la naturaleza de la información climática como de los procesos espaciales o geográficos que se relacionan con esta:</p> <ol style="list-style-type: none"> 1. Conocimiento de las problemáticas asociadas al registro diario de datos meteorológicos en estaciones automáticas (intervalos de medida, detección de <i>outlyers</i>, ausencia de valores registrados, ...). 2. Necesidad de diseño de un proceso automático de interpolación de lagunas en los valores de precipitación diaria basado en el comportamiento de la precipitación en el entorno de la estación y, por tanto, con necesidades de incorporar algoritmos de interpolación espacial y de funciones matemáticas (regresión simple y múltiple), que permitan la consecución de una serie temporal continua en los valores de precipitación diarios de los que se derivarán los mensuales necesarios para el Índice. 3. Conocimiento del funcionamiento, comportamiento y requisitos del IESP como índice estandarizado basado en la evaluación de las anomalías pluviométricas acumuladas. 4. Diseño de procesos automatizados de interpolación espacial del IESP con el objetivo de obtener cartografías continuas del mismo. 5. Exploración previa de la estructura empírica del comportamiento del IESP en su serie histórica con el objetivo de extraer en cada mes el pronóstico del mismo en una serie de 1 a 12 meses, así como la probabilidad asociada a estos. 6. Conocimiento del funcionamiento y manejo de geometrías en entornos de SGDB espaciales. 7. Diseño de semiologías gráficas que garanticen una correcta transmisión del comportamiento del Índice a los usuarios del Sistema. 8. Conocimiento del funcionamiento y manejo de la construcción de semiologías para servicios WMS (<i>Web Map Services</i>) en el contexto de servidores de mapas 9. Conocimiento del funcionamiento y diseño de visores web como ventana de difusión del indicador de estado. 		

IV.1.2. Diagrama Funcional del Sistema (Figura IV.3)



DFD nº 00013 Diagrama de Flujo de Datos del Sistema de Información de Alertas de Sequía Pluviométrica de Andalucía. Nivel 1

Objetivos: Representación de los procesos y requisitos funcionales necesarios para el funcionamiento del Sistema y de las relaciones y flujos compartidos entre ellos. El Sistema pretende:

- Calcular el IESP mensual en tiempo real (valor del mes anterior al de acceso en el Sistema por parte de los usuarios –mes en curso-1-).
- Espacializar en Andalucía el comportamiento de dicho Índice
- Calcular los pronósticos 1.....12 meses y la probabilidad asociada a los mismos
- Generar las cartografías de esos Indicadores como servicios interoperables WMS
- Utilizar en la arquitectura en la medida de lo posible software libre según especificaciones de la Junta de Andalucía

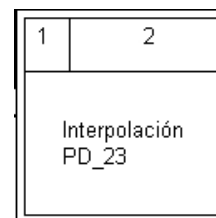
Datos de partida: Los únicos datos de partida son los valores de precipitación diaria registrados en un conjunto de 297 estaciones meteorológicas automáticas presentes en las distintas redes meteorológicas que operan en Andalucía, entre las que se encuentran las 84 estaciones automáticas seleccionadas para el cálculo del IESP. La selección de dichas estaciones se sustenta en el análisis previo de la calidad de los registros en términos de cercanía a estaciones manuales de la AEMET sobre las que existían datos homogéneos (Test de Alexanderson. Proclim Tool) y extendidos a 1950 y con un porcentaje de lagunas mensuales inferior al 20% que fueron posteriormente interpoladas mediante regresión múltiple con el grupo de estaciones mejor correlacionadas. El cálculo del IESP, mensual, necesita datos continuos de precipitación mensual que necesitaron la interpolación previa. Con ellos se procedió al cálculo del IESP en 288 estaciones manuales de referencia desde 1950 hasta la puesta en marcha del sistema. (Anexo 1).

El cálculo del Indicador de Estado en tiempo real se realizará por tanto sobre la precipitación acumulada de los valores diarios registrados en 84 estaciones automáticas de diferentes redes almacenados en el Subsistema de Climatología Ambiental CLIMA de la Red de Información Ambiental (REDIAM) – Consejería de Medio Ambiente. Junta de Andalucía.

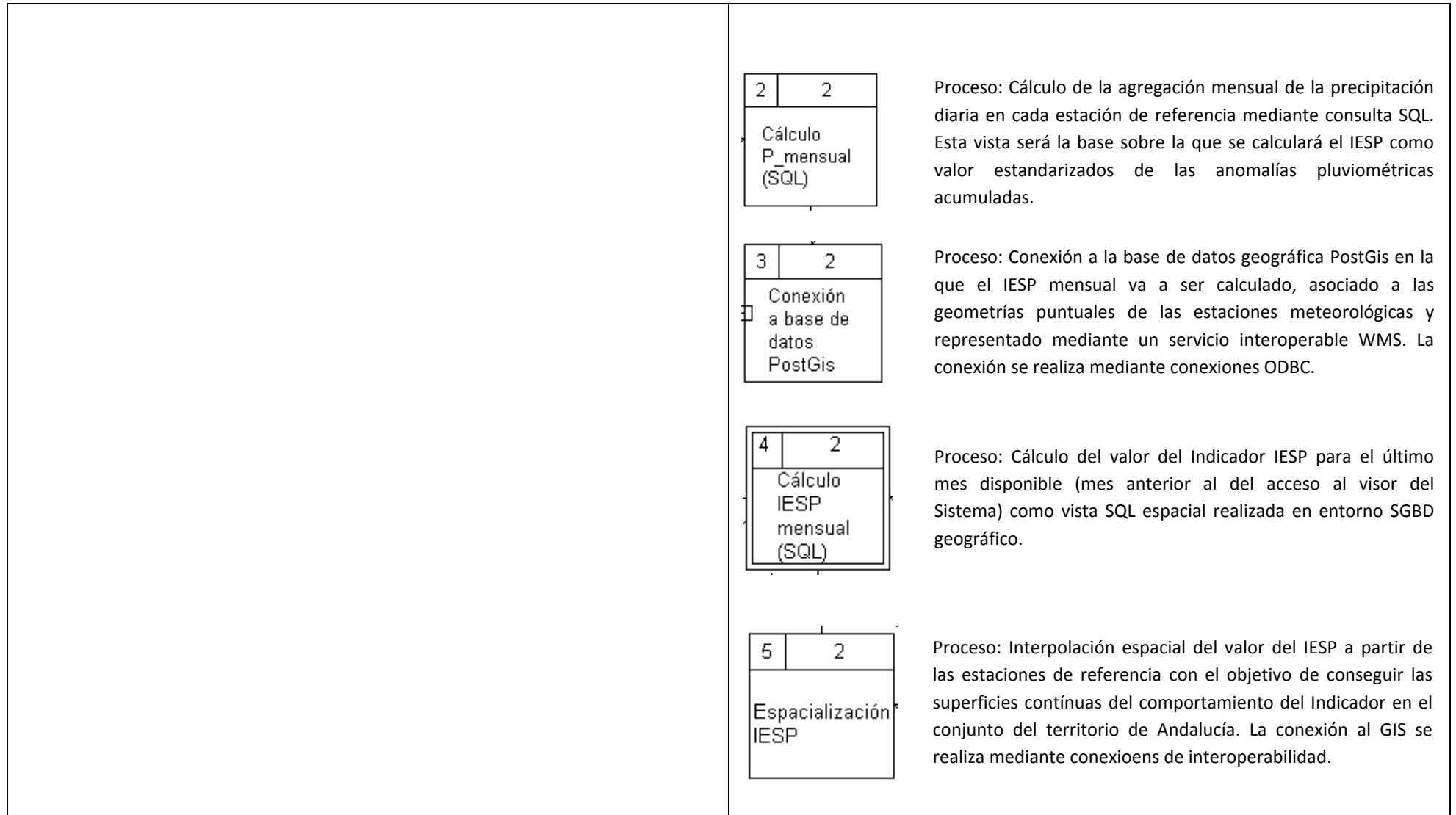
Requisitos funcionales:

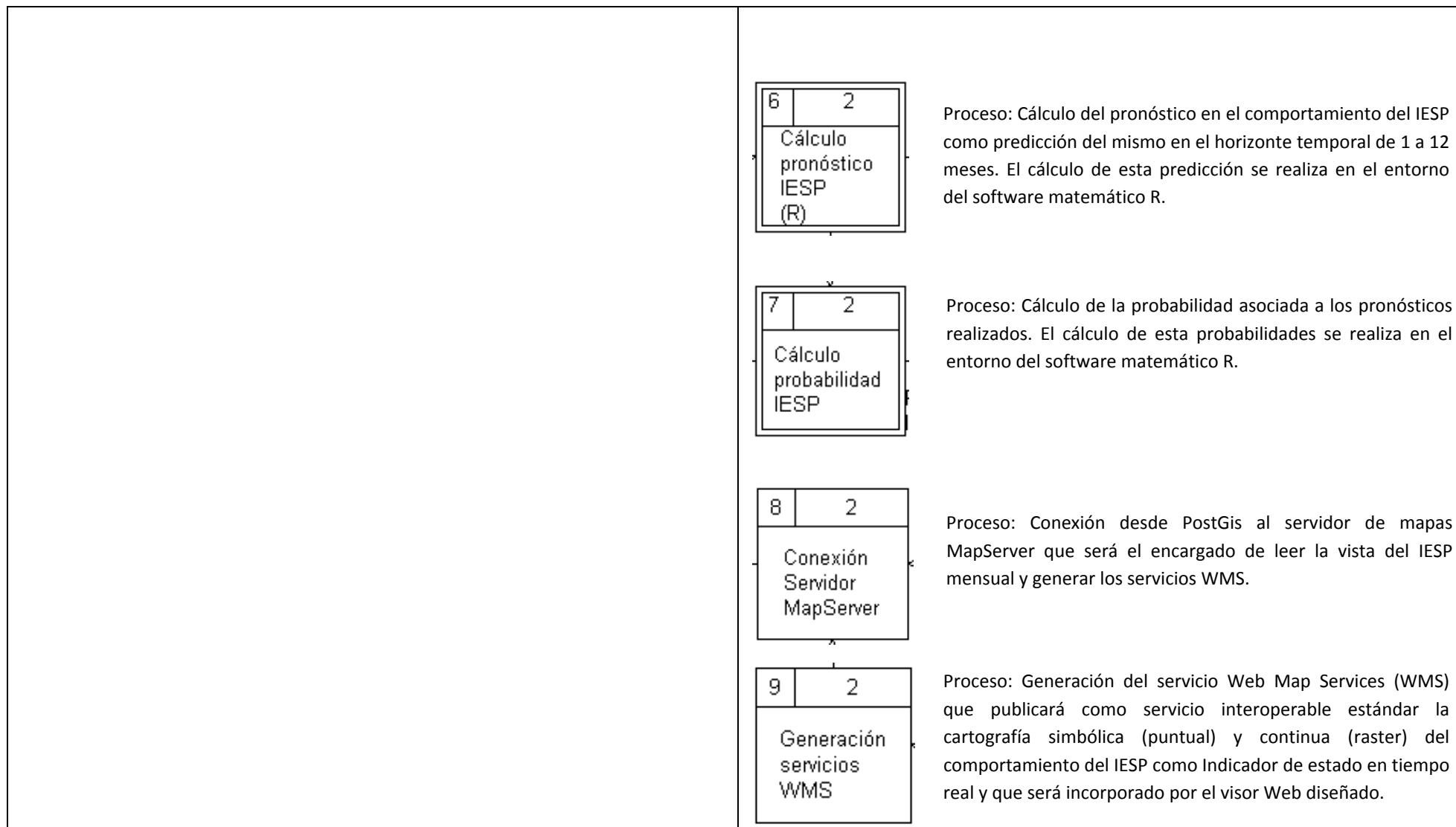
- Conexión con CLIMA.
- Conexión con Subsistema de Indicadores Climático-Ambientales.
- Interpolación de lagunas en los valores de precipitación diaria para obtener valores mensuales.
- Cálculo de IESP mensual (último mes disponible), pronósticos y probabilidades
- Generar servicios interoperables wms
- Generar cartografías de precipitación diaria en Andalucía para su incorporación en el Subsistema CLIMA.
- Incorporar herramientas de software libre.

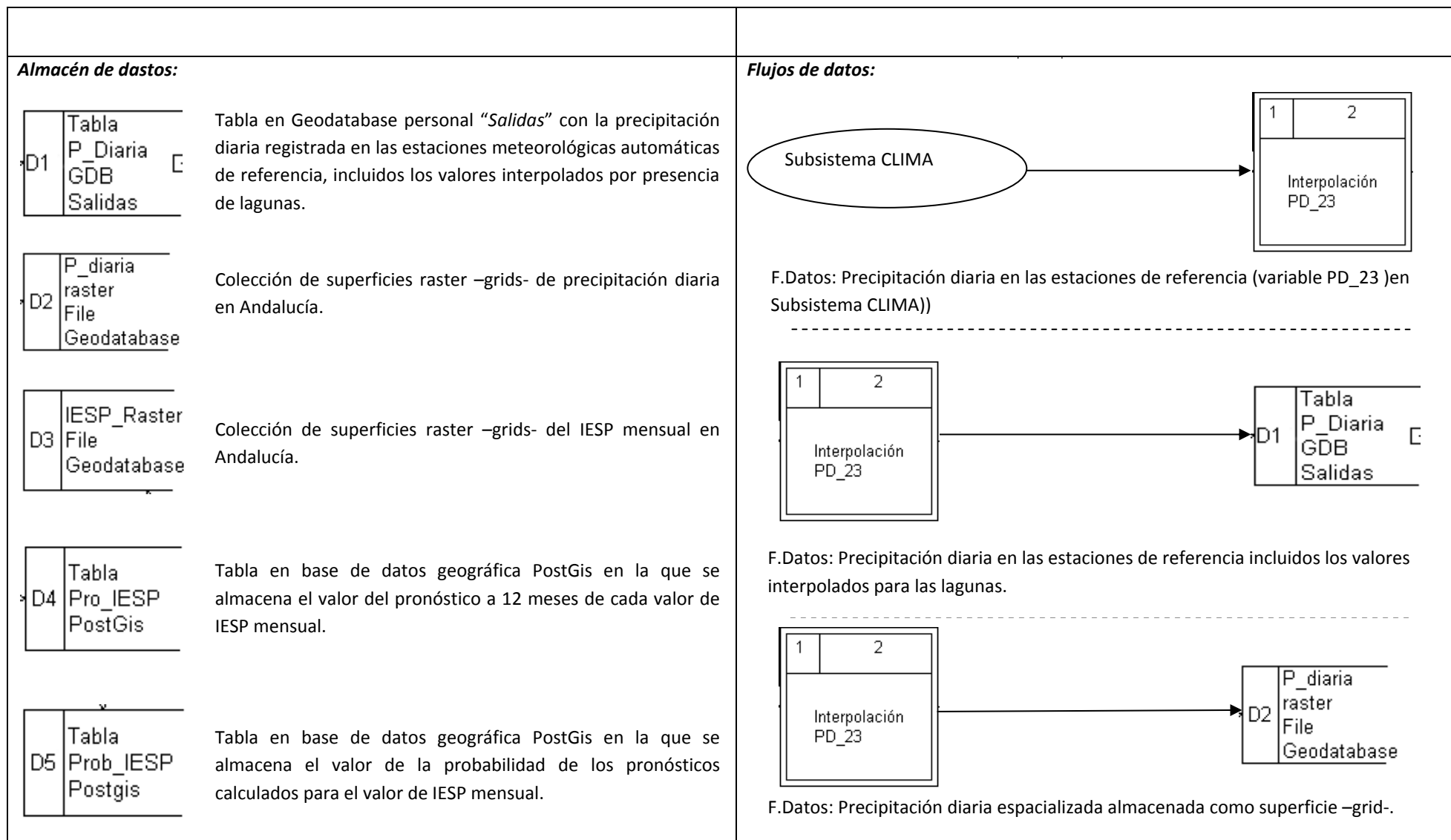
Procesos de datos:

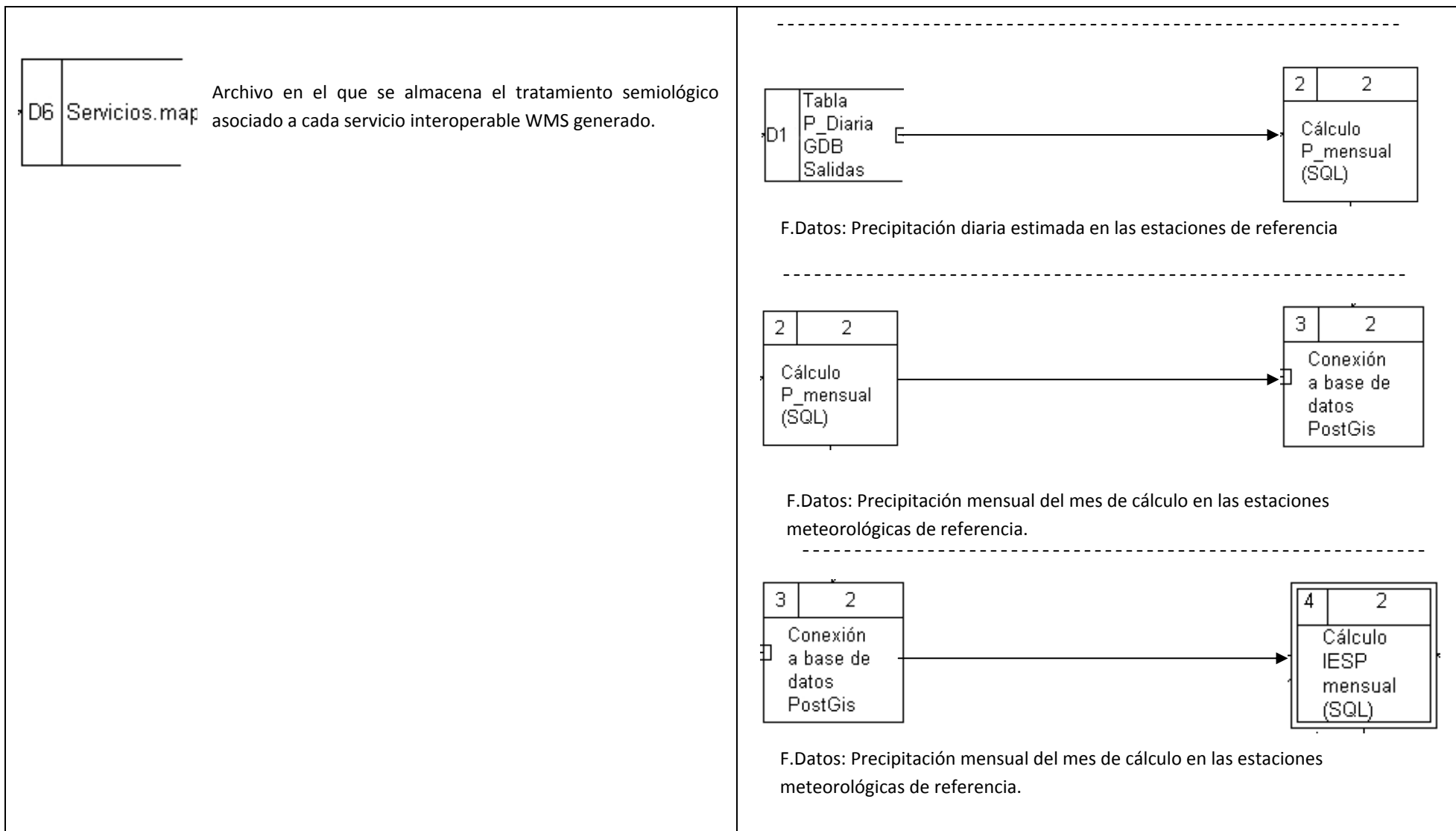


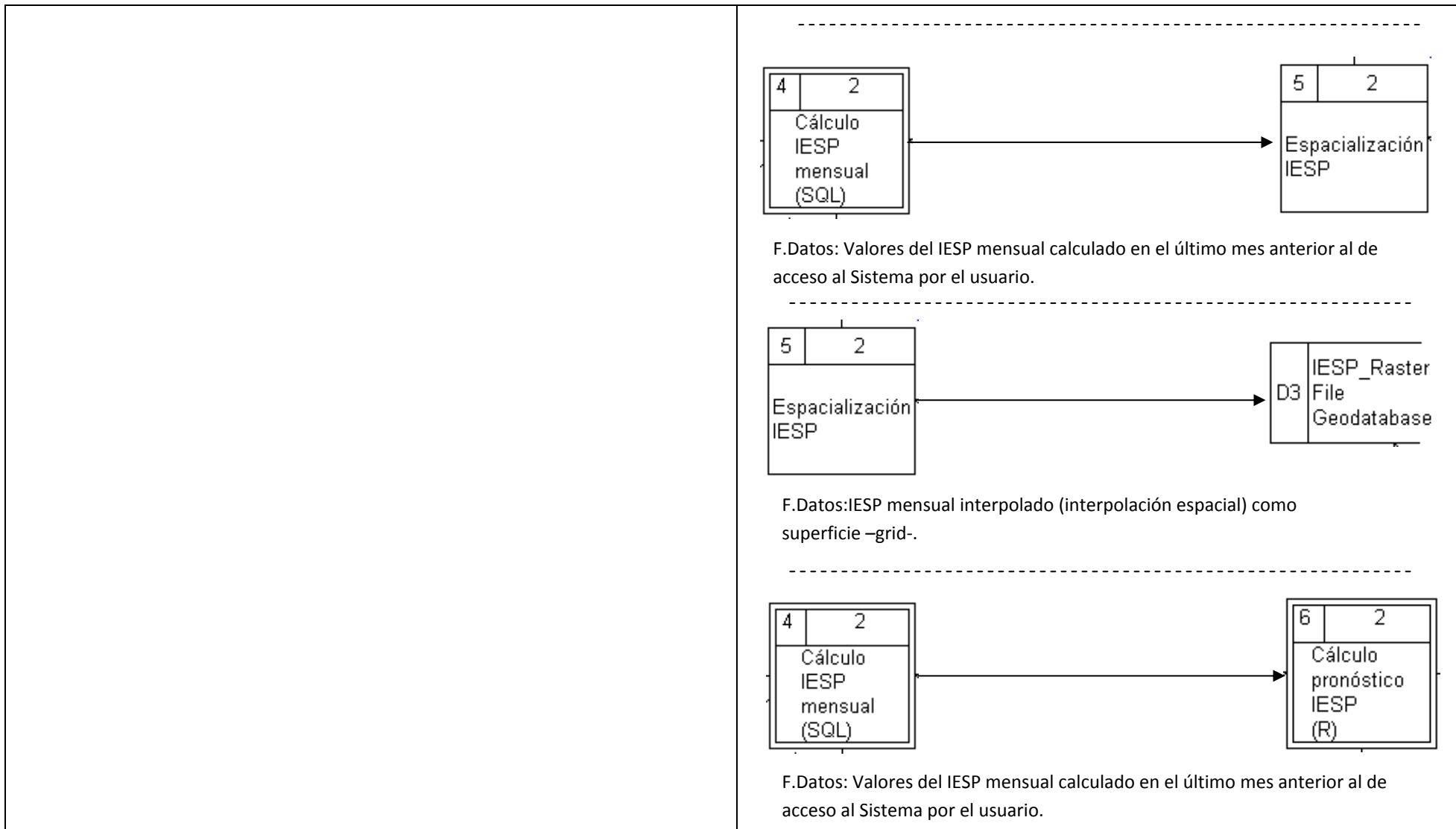
Proceso: Interpolación de los valores de precipitación diaria (variable PD_23 en Subsistema CLIMA) ausentes en la serie del mes de cálculo mediante procesos geoestadísticos asociados.

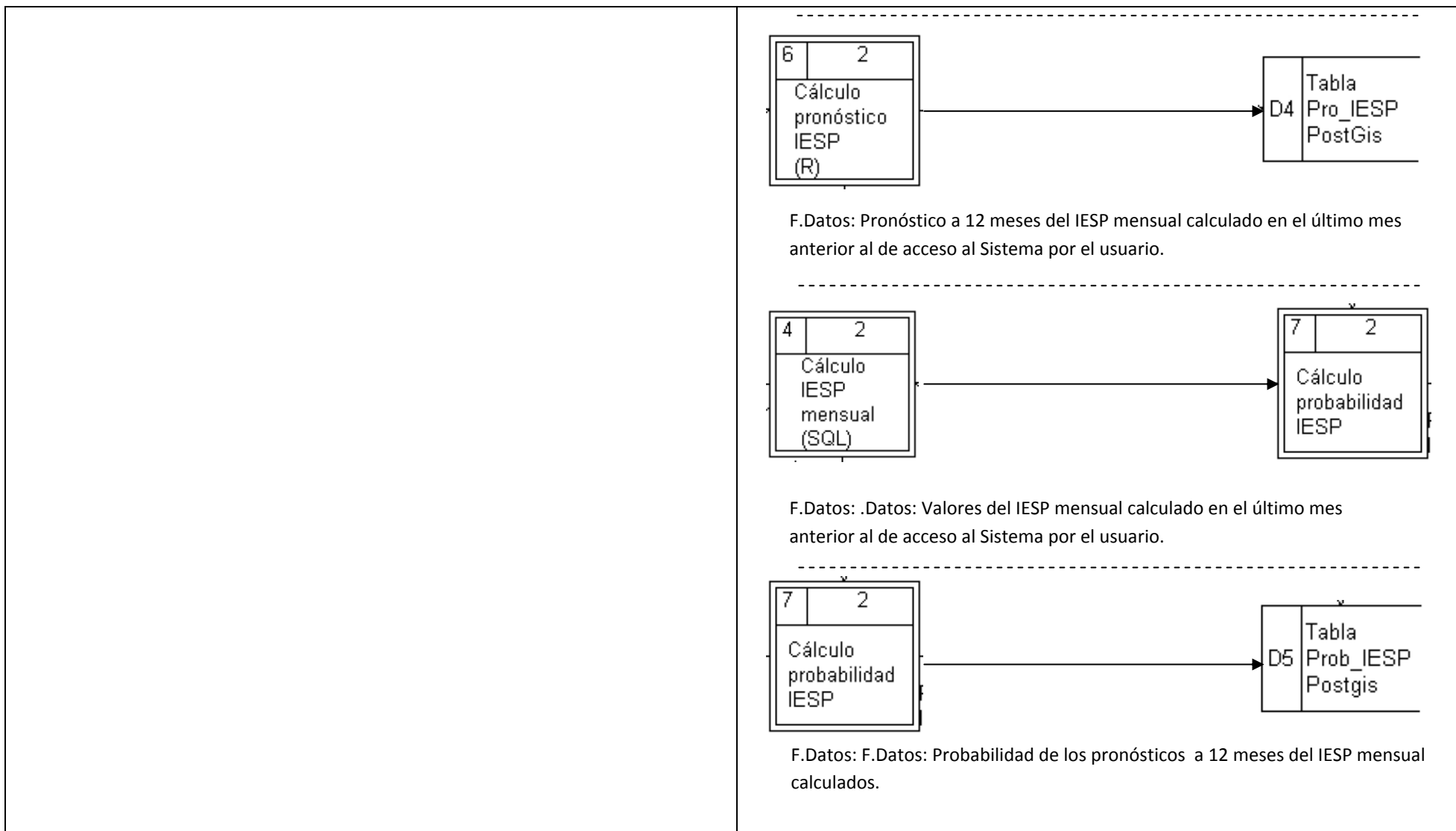


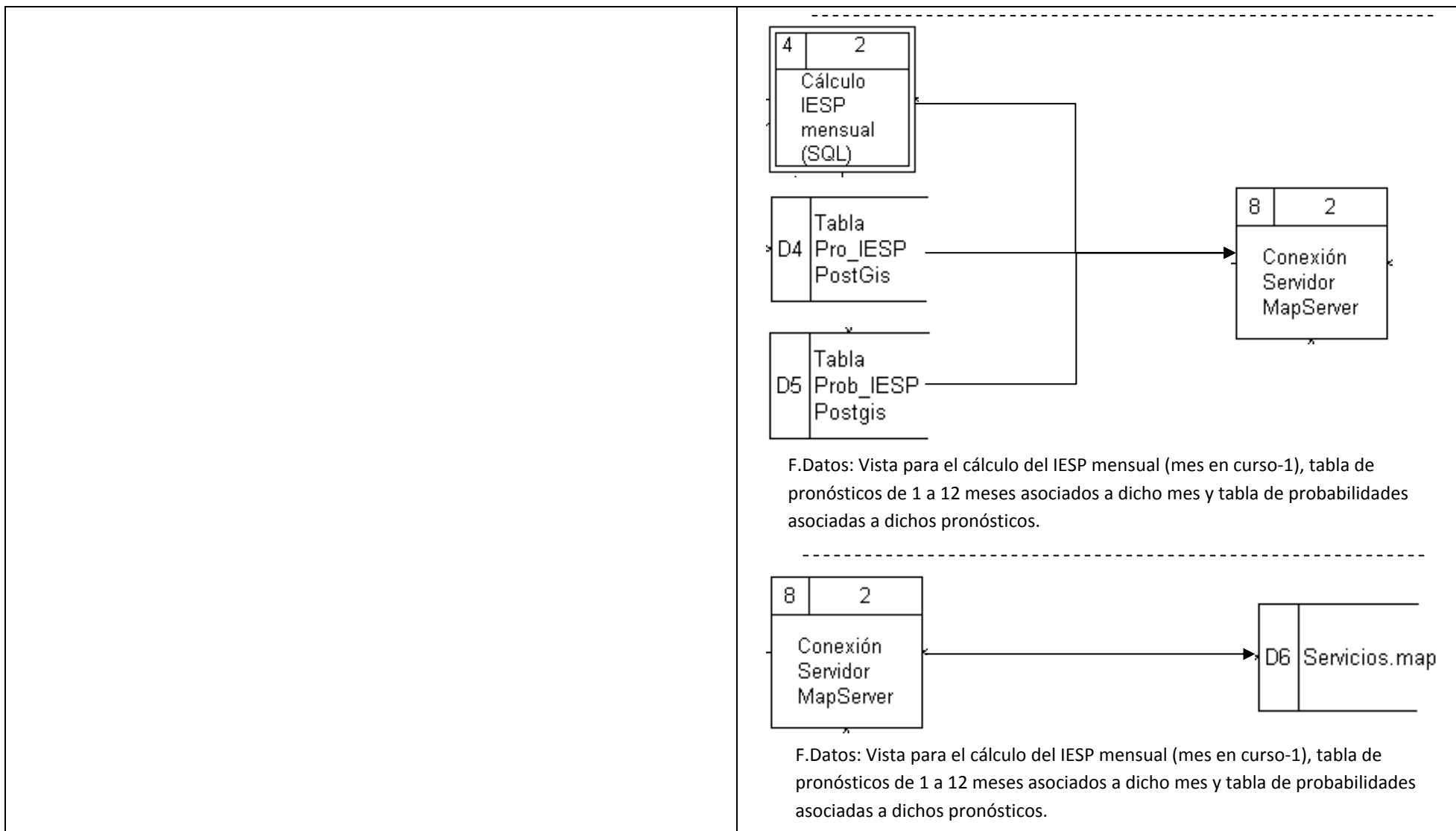


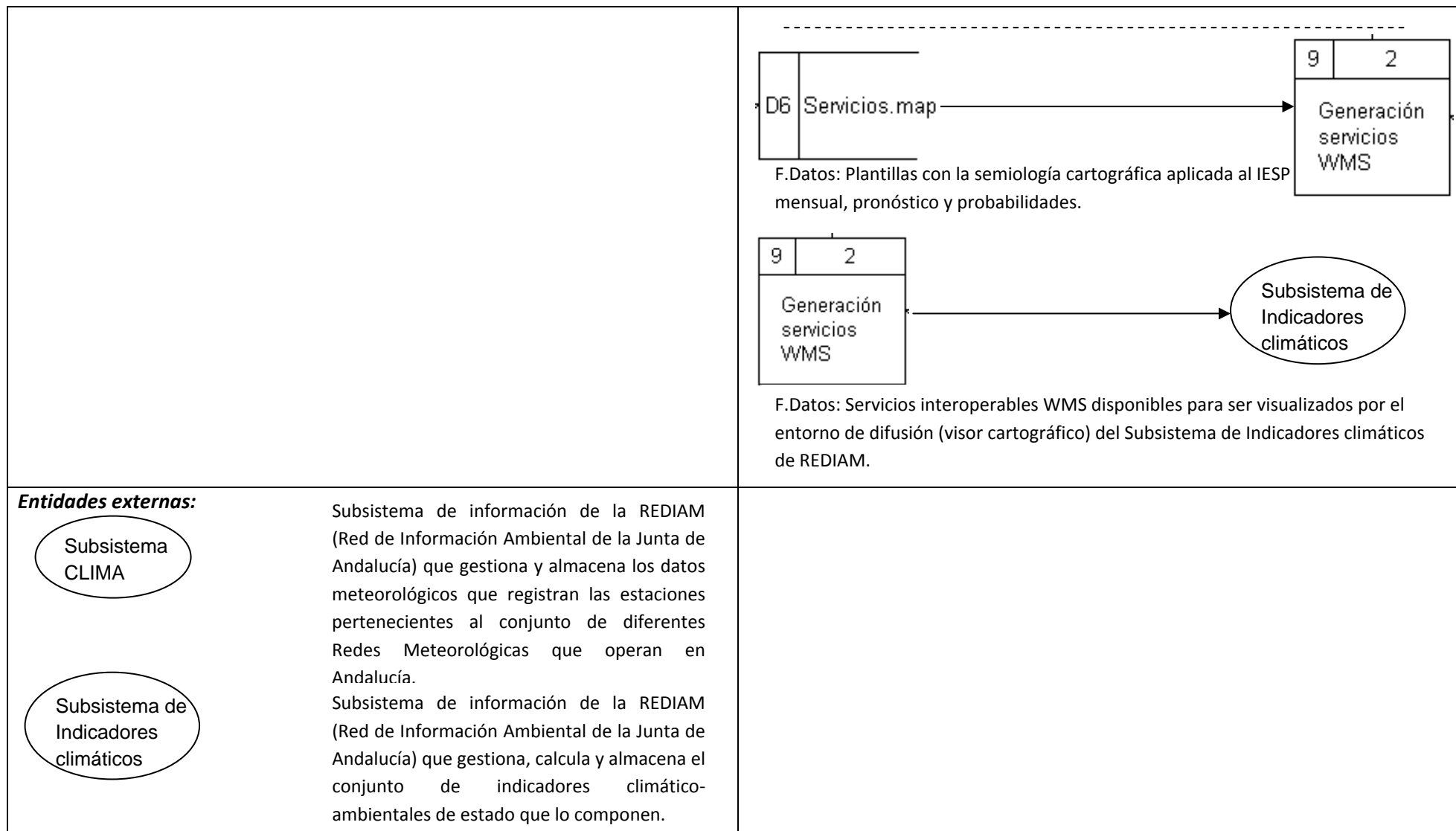












Requisitos de naturaleza geográfica y/o climática:

Entre los requisitos funcionales representados por este diagrama funcional del Sistema que suponen procesos directamente asociados a la naturaleza geográfica de la información climática, cabe destacar los siguientes:

1. Conocimiento de las problemáticas asociadas al registro diario de datos meteorológicos en estaciones automáticas (intervalos de medida, detección de *outliers*, ausencia de valores registrados, ...).
2. Necesidad de diseño de un proceso automático de interpolación de lagunas en los valores de precipitación diaria basado en el comportamiento de la precipitación en el entorno de la estación y, por tanto, con necesidades de incorporar algoritmos de interpolación espacial y de funciones matemáticas (regresión simple y múltiple), que permitan la consecución de una serie temporal continua en los valores de precipitación diarios de los que se derivarán los mensuales necesarios para el Índice.
3. Conocimiento del funcionamiento, comportamiento y requisitos del IESP como índice estandarizado basado en la evaluación de las anomalías pluviométricas acumuladas.
4. Diseño de procesos automatizados de interpolación espacial del IESP con el objetivo de obtener cartografías continuas del mismo.
5. Exploración previa de la estructura empírica del comportamiento del IESP en su serie histórica con el objetivo de extraer en cada mes el pronóstico del mismo en una serie de 1 a 12 meses, así como la probabilidad asociada a estos.
6. Conocimiento del funcionamiento y manejo de geometrías en entornos de SGDB espaciales.
7. Diseño de semiologías gráficas que garanticen una correcta transmisión del comportamiento del Índice a los usuarios del Sistema.
8. Conocimiento del funcionamiento y manejo de la construcción de semiologías para servicios WMS (*Web Map Services*) en el contexto de servidores de mapas
9. Conocimiento del funcionamiento y diseño de visores web como ventana de difusión del indicador de estado.

IV.1.3. Metodología completa propuesta para la interpolación de la precipitación diaria (PD_23) en Andalucía

Los indicadores climáticos son las variables de estado básicas en la gestión y seguimiento del medioambiente siendo su utilización fundamental en los procesos de evaluación y toma de decisiones a todas las escalas de gestión.

Uno de los objetivos fundamentales del presente trabajo consiste en el diseño y la implementación de una serie de indicadores climáticos, generados a partir de las variables del Subsistema de Información de Climatología Ambiental (CLIMA) de la Consejería de Medio Ambiente (CMA) de la Junta de Andalucía. Los recursos hídricos constituyen el eje central de la política ambiental en Andalucía y, por ello, la generación en rutina de indicadores climáticos relacionados con la precipitación en Andalucía es una cuestión de especial importancia en la gestión ambiental de la región. Entre ellos se encuentra el Indicador objeto de estudio de este trabajo de investigación, el IESP.

Sin embargo, las dificultades derivadas de la adquisición de los datos de precipitación necesarios que posibiliten la obtención de estos indicadores de seguimiento, de forma sincrónica y continua en el tiempo, constituyen un problema de difícil solución. Los registros provenientes de las estaciones climáticas automáticas no permiten disponer, hoy por hoy, de una información regular ni en el espacio ni en el tiempo, lo que impide generar estos indicadores con la actualización y continuidad necesarias. El principal reto se centra en la obtención de información regional pluviométrica, a escala diaria, a partir de la cual se calculen los valores mensuales, con el objetivo de generar series continuas sin presencia de lagunas, ya que incluso un bajo porcentaje de las mismas tienen capacidad de distorsión en las series mensuales (Stooksbury y col. (1999)). En el caso de algunos indicadores, como los de sequía, esta continuidad temporal de los registros mensuales de precipitación es una premisa básica para poder efectuar su cálculo, al estar basados en el déficit pluviométrico calculado, como diferencia de precipitación con el mes anterior.

En la figura IV.4 pueden observarse el número medio mensual de registros diarios de precipitación con existencia de lagunas en el conjunto de observatorios de referencia usados para nuestro estudio. El promedio mensual para el conjunto de series es de 6 días. De hecho, de las 297 estaciones automáticas, solamente en 3 de ellas encontramos para el año 2010 el conjunto completo de registros diarios anual. Esta realidad, junto a las características de la precipitación en la región, con valores diarios que en el caso más extremo han llegado a superar los 600 mm (Almuñécar, 1963) y donde valores de 50 a 100 mm diarios no son infrecuentes, generan la principal problemática para la configuración de sistemas de indicadores ambientales de estado en tiempo real.

La disponibilidad y acceso a los registros del mes con un lapso aproximado de una semana, límite que hemos puesto para el cálculo de los indicadores, solamente puede hacerse a partir de conexiones a las bases de datos meteorológicas, en nuestro caso el Subsistema CLIMA y a los valores diarios registrados en las estaciones meteorológicas automáticas. El registro de datos en estaciones automáticas presenta una serie de problemáticas y dificultades muy bien puestas de manifiesto por diversos autores incluidas en los propios documentos de la Organización Meteorológica Mundial (OMM, 2004), entre los que se encuentran la presencia de errores aleatorios, errores sistemáticos o errores aberrantes, así como la ausencia de registros por fallo en el sensor o en el *datalogger*. Todos ellos tendrán influencia en el comportamiento del Indicador, si bien, en esta primera fase de desarrollo del proyecto, son las lagunas existentes en los datos diarios, las que imposibilitan el cálculo inmediato del mismo.

En nuestro caso, además, existen dos filtros en relación con el resto de errores, aquellos que mantiene la propia Agencia Estatal de Meteorología, a cuyos servidores se conecta el Subsistema CLIMA y los del propio Subsistema, orientados exclusivamente a los errores aberrantes.

Es, por tanto, la persistencia en la presencia de lagunas en el conjunto de estaciones, la dificultad más importante que será necesario solventar para poder llevar a cabo el cálculo del IESP en rutina como es pretendido.

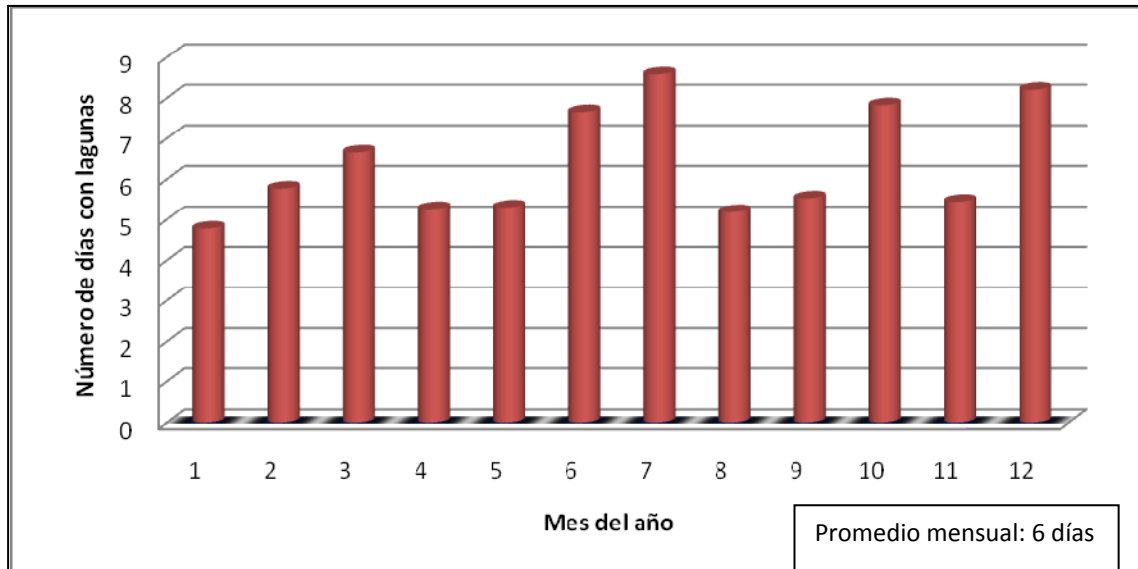


Figura IV.4. Número medio mensual de registros diarios de precipitación con presencia de lagunas

Fuente: elaboración propia

A continuación se detalla la secuencia metodológica seguida para la creación de un modelo de estimación de la precipitación diaria en estaciones con lagunas.

IV.1.3.1. Redes de estaciones y variables utilizadas

El Subsistema CLIMA dispone de datos procedentes de redes de estaciones automáticas meteorológicas con una cobertura espacial suficiente como para garantizar una buena representación de las variables climáticas. Un conjunto de 84 estaciones automáticas, del total de las 297 existentes, han sido utilizadas para aplicar la metodología propuesta de relleno de lagunas a partir de la selección un número medio de 300 valores de precipitación diaria registrados en el periodo 2005-2009 en las 84 estaciones de referencia (Figura IV.5). Esta muestra fue seleccionada de forma aleatoria entre aquellos días sin lagunas y que hubiesen registrado precipitación, es decir, los contabilizados con valores superiores a cero en este período de 5 años.

La selección de las 84 estaciones está justificada en la necesidad, al ser el IESP un índice diacrónico que va acumulando las anomalías acumuladas estandarizadas de la precipitación mensual, de contar con las series más alejadas en el tiempo posible calculadas sobre las series pluviométricas procedentes de las estaciones manuales de la región. Se realizó un estudio (Anexo I) a partir del cual se pudieron obtener series completas mensuales –interpolados los valores ausentes por regresión con la estación mejor correlacionada siempre que estos no superaran el 10% de los valores de la serie- de un total de 84 estaciones. El año de inicio de cálculo del IESP que garantizara una suficiente cobertura espacial de la región fue 1950, si bien algunas estaciones comienzan en años posteriores que no sobrepasan el año 1960 por insuficiencia de registros.

El cálculo del IESP, como ya se apuntó en el Capítulo I, necesita de los valores de las medianas del periodo normal de precipitaciones así como de los estadísticos descriptivos –media y desviación típica- de las propias series de IESP. Lógicamente, tanto las medianas como los estadísticos no pueden ser calculados sobre series de EMA ya que su alcance temporal tiene como límite la década de los años 80 en los que comienzan a estar operativas. Por lo tanto, fue necesario asignar a cada una de las estaciones EMA disponibles -297- estos parámetros a partir de un criterio de cercanía a alguna de las estaciones manuales para las que disponíamos de las series completas de IESP desde 1950. El criterio de distancia fijado fue de 10 Km entendiendo que un fenómeno como la sequía, de carácter extenso en el territorio, no supera este umbral de 10 km en términos de variabilidad espacial. No obstante la media de esas distancias es de solo 4,2 Km. Al mismo tiempo, la selección de esas 84 estaciones está también basada en una selección previa de 100 estaciones EMA sobre las que el servicio de la Consejería de Medio Ambiente realiza un seguimiento y calcula diferentes indicadores ambientales.

Los datos utilizados en este trabajo provienen de las cuatro grandes redes de observación automática integradas en el Subsistema CLIMA provenientes de diferentes instituciones:

- Red de Estaciones Meteorológicas Automáticas de la Agencia Estatal de Meteorología (AEMET).
- Red de Estaciones Automáticas y Remotas de Meteorología de la Consejería de Medio Ambiente de la Junta de Andalucía (CMA).
- Red de Información Agroclimática de la Consejería de Agricultura y Pesca de la Junta de Andalucía (IFAPA).
- Red de Alerta e Información Fitosanitaria de la Consejería de Agricultura y Pesca de la Junta de Andalucía (CAP).

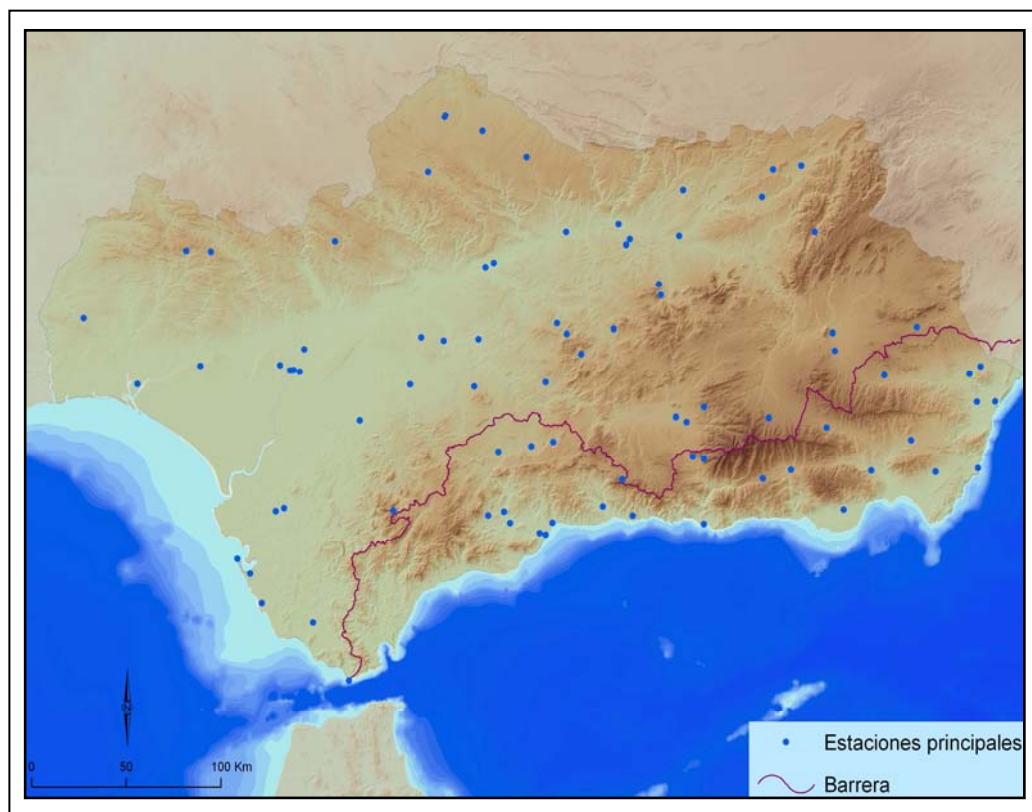


Figura IV.5. Estaciones automáticas seleccionadas

Fuente: elaboración propia

IV.1.3.2. Metodología propuesta y aplicada

Con el fin de obtener datos diarios de precipitación en las 84 estaciones seleccionadas que permitan la generación de indicadores en rutina, se ha diseñado una metodología cuyo objetivo no es la espacialización de la precipitación diaria en Andalucía², sino garantizar la estimación estadística de los valores de precipitación diaria que asegure la existencia de información a escala mensual en los observatorios seleccionados. El procedimiento consta de diferentes etapas y está diseñado para conseguir series de registros diarios continuos, interpolando las lagunas para cada estación y fecha, mediante un proceso que optimiza la calidad de la estimación en cada punto de forma automatizada.

La interpolación de las lagunas se llevó a cabo combinando las técnicas de interpolación espacial (interpoladores) y de regresión lineal utilizando las herramientas contenidas en los programas ArcGis y SPSS. La implementación del proceso, los análisis desarrollados, y los resultados obtenidos en forma de nuevos datos, han sido estructurados a partir del diseño de un modelo conceptual que posteriormente se tradujo en la configuración de una base de datos geográfica (*Geodatabase*).

² Los métodos de espacialización de la estructura de la precipitación diaria en una región proceden básicamente de las fuentes de radar, los datos procedentes de los satélites de última generación –Meteosat 5- o de metodologías de *downscaling* dinámico –basados en la obtención de modelos numéricos de escalas limitadas –LAM- o de *downscaling* estadístico –basados en modelos de cadenas de markov, modelos de regresión lineal, correlación canónica, redes neuronales o predicción por análogos-. Estos últimos están basados en la correlación empírica existente entre los valores medidos de una serie de muestra y la estructura vertical de la atmósfera en esos días medida en los parámetros de humedad absoluta, humedad relativa, temperatura, presión, geopotencial y componentes u y v del viento y determinación del tipo de tiempo de cada día a partir de patrones preestablecidos. La dificultad de automatización de estos métodos –especialmente los basados en el *downscaling* estadístico (Abaurrea, 2005)-, las elevadas necesidades de computación –es el caso del *downscaling* dinámico- junto a los importantes requisitos de datos de entrada necesarios, así como el horizonte temporal de este proyecto, 16 meses, obligaron a un cambio de perspectiva en el enfoque de la cuestión.

La hipótesis de partida es que la precipitación diaria medida en un observatorio en un día cualquiera dado (PD_23: precipitación de 0 a 24 horas) tiene relación con la precipitación observada en el entorno de dicho observatorio. El aplicar esta hipótesis debe reflejarse en los correspondientes diagramas de dispersión y en los valores establecidos por unos coeficientes marcados de correlación de *Pearson* superiores a 0,8. Ello no es óbice para admitir que la afirmación formulada puede no verificarse en algunos de los valores de los volúmenes de precipitación diarios alcanzados en las estaciones de partida. La naturaleza conocida de los procesos que se sitúan en la raíz de las precipitaciones de carácter local, que sin duda están presentes en las estaciones de nuestros entornos climáticos, nos obliga a testar el peso de dichas situaciones en el conjunto de la series de datos manejadas, con el objetivo de trasladar la hipótesis de partida al conjunto de los datos o bien solamente a un subconjunto de los mismos.

En la primera etapa el objetivo consiste en disponer, para cada estación, de dos series de datos, una con los datos de precipitación diaria observados, registrados en un número medio de 300 días seleccionados en el conjunto de estaciones de referencia, y otra con los valores estimados por interpolación espacial para las mismas estaciones y días. Esta última serie se obtiene mediante un procedimiento que consiste en tener en cuenta el comportamiento pluviométrico del entorno, lo que hemos llamado ***precipitación regional***, obtenida por procedimientos de interpolación espacial, que logra optimizar así el proceso desde un punto de vista climatológico. Es importante destacar, la ventaja de este cálculo en relación a la mera regresión lineal simple o múltiple con observatorios cercanos, en primer lugar por la influencia de la distancia máxima aceptable (Hubbard (1994), y en segundo lugar, por la alta probabilidad, dado el comportamiento de las estaciones automáticas, de encontrar lagunas en las estaciones cercanas y por tanto en los valores de la variable independiente.

La siguiente fase utiliza las dos series para obtener la recta de regresión simple que las relaciona y que permite estimar cualquier valor ausente en las series de registros.

Los coeficientes de regresión se incorporan a la *geodatabase* del estudio para ser utilizados en el proceso de forma automática, como parámetros constantes para cada observatorio. Finalmente, se ha realizado un control de calidad de las técnicas empleadas mediante una validación cruzada –*cross validation*–, estimando diferentes medidas de los posibles errores cometidos en el proceso interpolación. A continuación se detallan cada una de las etapas de este estudio.

IV.1.3.3. Cálculo de la precipitación regional asociada a cada estación de partida

Para llevar a cabo esta fase se ha interpolado espacialmente según método anteriormente descrito, para cada una de las 84 estaciones, la precipitación de los 300 días seleccionados en el conjunto de Andalucía, en el supuesto de que no existiera el dato real en la estación, con el objetivo de evitar que dicho valor tenga peso en la interpolación regional efectuada. Este proceso se repite por tanto utilizando todos los datos disponibles del conjunto de 297 estaciones automáticas activas de la región, excluyendo en cada realización una estación diferente (“estación problema”). Todo el proceso se ejecuta mediante un modelo iterativo creado con la aplicación “*Model Builder*” de ArcGis 9.3 (Figura IV.6 y Anexo 2). El modelo se ejecuta 84 veces, una vez por cada estación, y en cada una de ellas consta a su vez de 300 ejecuciones, una por cada dato diario, resultando un total de 3650 realizaciones.

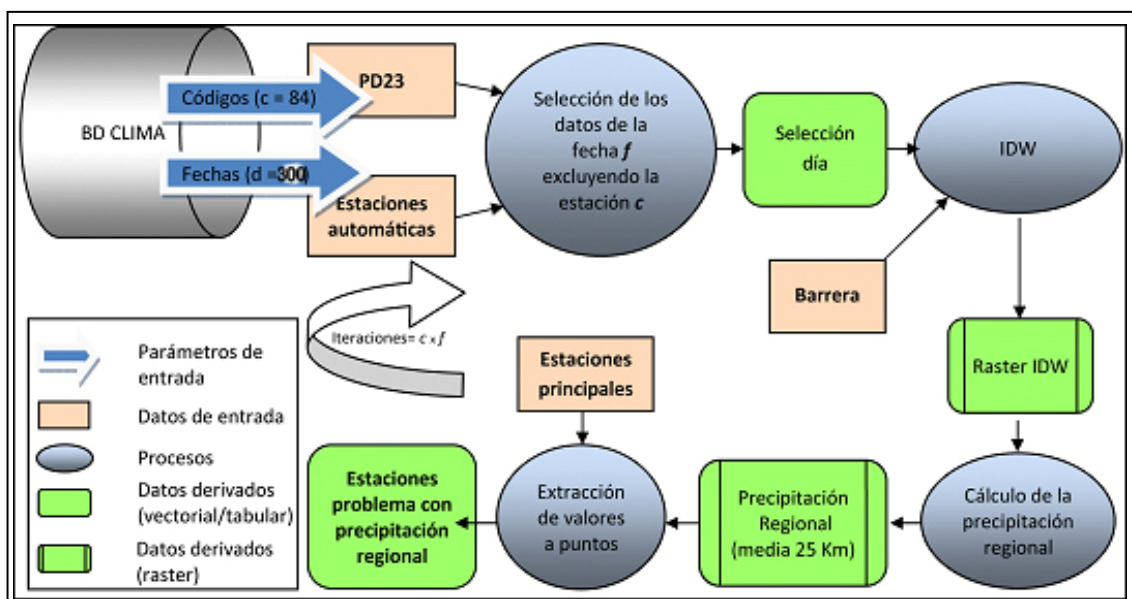


Figura IV.6. Esquema del modelo para la obtención de la precipitación regional en las estaciones problema

Fuente: elaboración propia

El método de interpolación utilizado ha sido el inverso de la distancia elevada al cuadrado (*IDW*)³, con barrera, utilizando un radio de búsqueda variable con un alcance máximo de 50 kilómetros y 12 vecinos. La barrera empleada es la divisoria entre las cuencas hidrográficas del Guadalquivir y del Mediterráneo cuya línea se obtiene de la capa de cuencas hidrográficas del Mapa Topográfico de Andalucía a escala 1:100.000 (Sistema Cartográfico de Andalucía, 2009). También se testó el interpolador por *Splines*, no encontrándose diferencias significativas al evaluar las diferencias globales en el conjunto de las situaciones diarias, jugando en contra de su utilización la limitación de la extensión espacial del interpolador, lo que dejaba a algunas estaciones fuera de las estimaciones. Para cada estación y fecha se generó una capa de “precipitación regional” aplicando en cada píxel la media pluviométrica de un entorno de 25 kilómetros de radio mediante la herramienta *Focal Statistics*. Un ejemplo de cada una de estas salidas puede verse en la figura IV.7 para una fecha determinada. Finalmente se extrajo el valor del píxel correspondiente a la estación problema para cada uno de los 300 días.

³ Basado en el método de Shepard (1968), el método Inverse Distance Weighted es una función de interpolación espacial de la forma:

$$u(\mathbf{x}) = \frac{\sum_{i=0}^N w_i(\mathbf{x})u_i}{\sum_{j=0}^N w_j(\mathbf{x})},$$

donde el valor interpolado u en un punto dado x se basa en muestras $u_i = u(x_i)$ para $i = 0, 1, \dots, N$.

y donde:

$$w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^p}$$

es la función simple ponderada del inverso de la distancia donde x es un punto arbitrario del espacio, x_i es un punto de valores conocidos y d es la distancia euclídea entre el punto con valor conocido x y el punto desconocido x_i , N es el número total de puntos conocidos usados en la interpolación y p es un número real positivo que actúa como potencia de la relación de distancia.

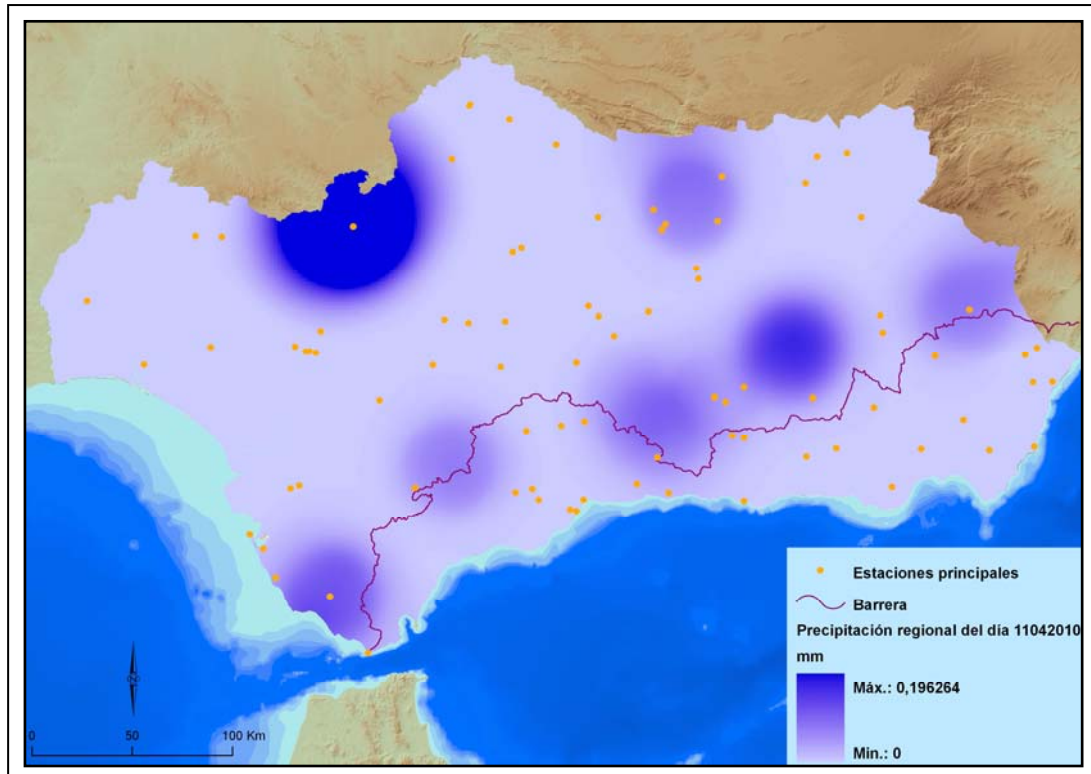


Figura IV.7. Ejemplo de *raster* de “precipitación regional” para el 11/04/2010

Fuente: elaboración propia

Es decir, se obtuvieron mediante este modelo, y para cada estación evaluada, 300 valores de precipitación diaria como media espacial registrada en el entorno de cada estación, la *precipitación regional*, considerando como desconocida la precipitación real registrada en cada una de las estaciones de estudio, al haber sido éstas excluidas del proceso de interpolación espacial llevado a cabo.

IV.1.3.4. Modelos de regresión lineal

Una vez obtenidas las series de valores estimadas de precipitación en cada estación a partir de las capas de precipitación regional para todas las fechas consideradas, se calculan los coeficientes de correlación entre estas series y las de valores reales mediante el programa de tratamiento estadístico SPSS. El promedio del coeficiente de correlación de *Pearson* en las 84 estaciones es de 0,85 (Figura IV.8), corroborándose de esta manera la hipótesis de partida en la mayoría de los casos.

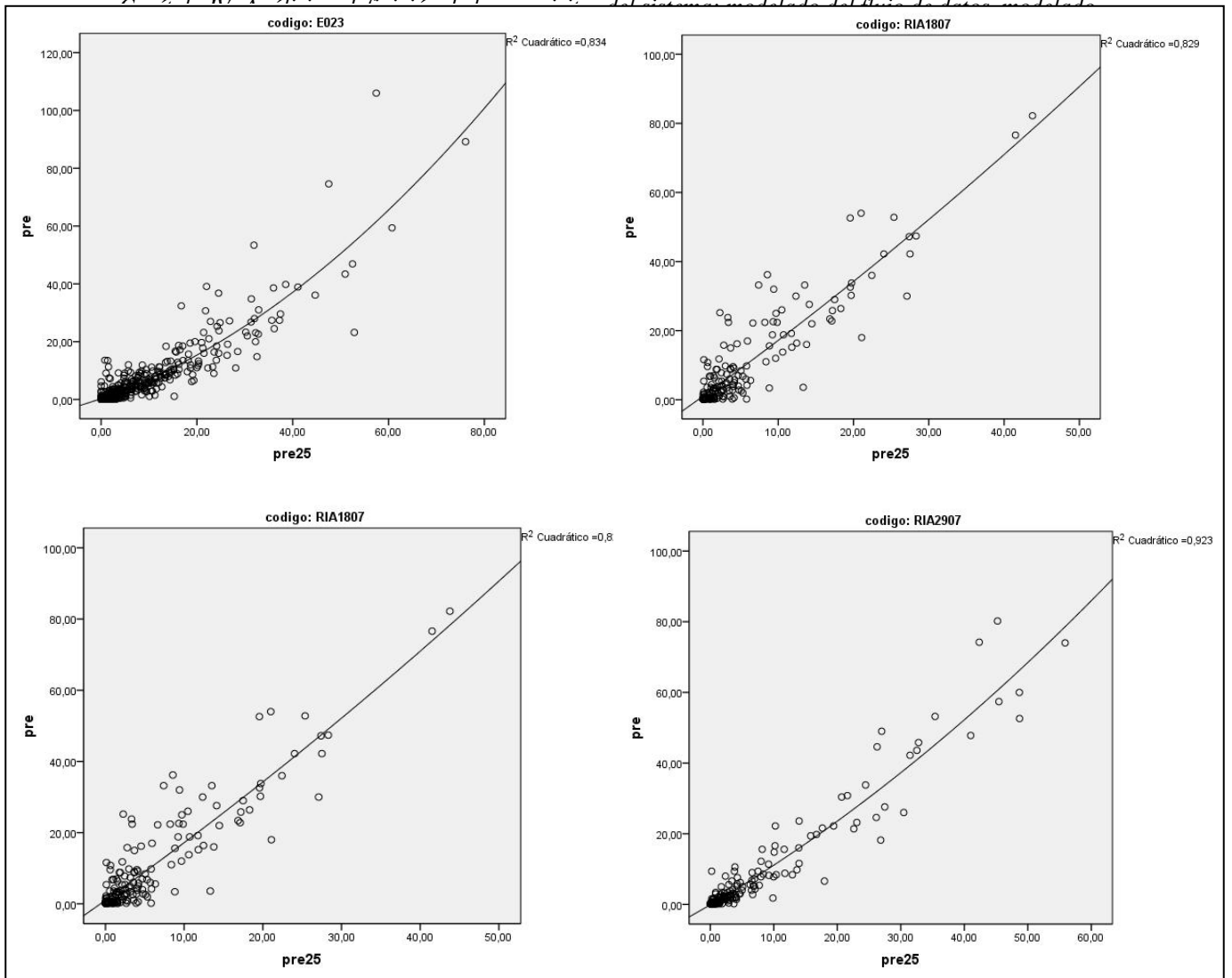


Figura IV.8. Ejemplos de gráficos de dispersión entre precipitación observada y precipitación del entorno para estaciones de las 4 redes de observación

Fuente: elaboración propia

En un conjunto importante de estaciones (53) se obtienen correlaciones superiores a este valor y sólo descienden por debajo de 0,7 en seis de ellas. Las que presentaban estos valores inferiores en los coeficientes de correlación, se corresponden bien con estaciones en las que el número de procesos de precipitación local es más alto (i.e surco intrabético), o bien con estaciones con un número insuficiente de vecinos que garantice el funcionamiento del interpolador. Ambos tipos de estaciones fueron eliminadas del proceso final y, por lo tanto, del conjunto final de estaciones final seleccionadas para el cálculo del IESP, que fueron finalmente 78. A continuación se llevó a cabo un análisis de regresión simple, tomando la precipitación observada real en cada estación como variable dependiente, y la precipitación estimada obtenida a partir de la media del entorno en 25 kilómetros como variable independiente. Los

coeficientes no estandarizados de los modelos en cada observatorio fueron asociados a la capa geométrica de las estaciones, a fin de que pudieran ser utilizados en el modelo de geoproceso posterior a partir de una cardinalidad en la base de datos geográfica de 1 a 1 (Figura IV.9). El método por tanto supone encontrar 78 modelos de relación en la precipitación diaria, uno por cada estación. **No se trata por tanto de un modelo de regresión para la espacialización de esta variable en Andalucía, sino del ajuste de un modelo individual para cada estación meteorológica que sea un interpolador válido de lagunas en sus registros.**

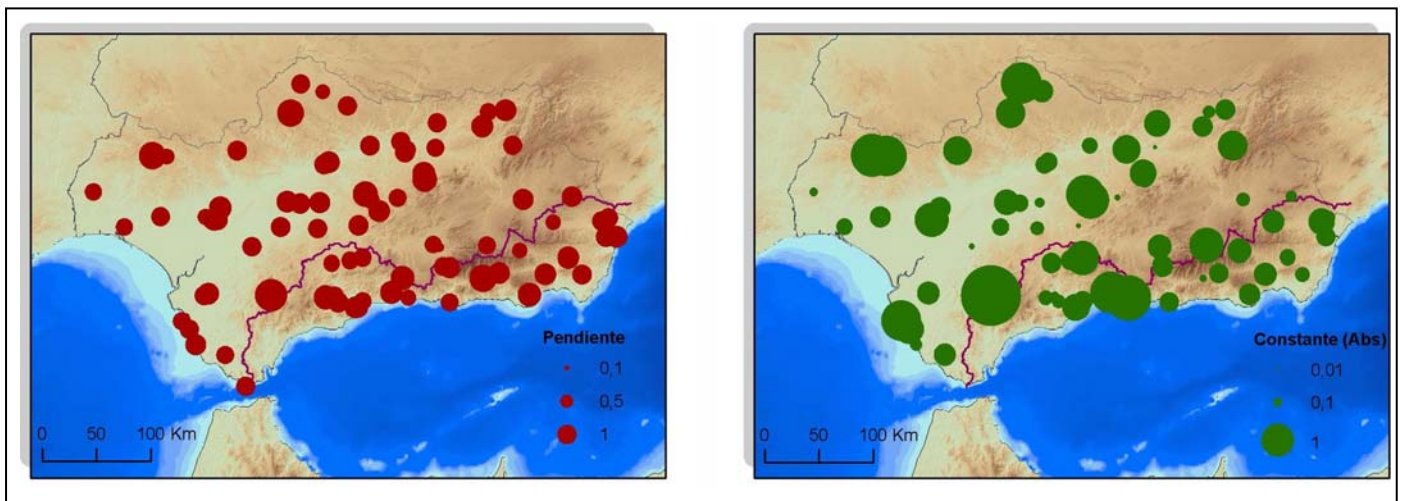


Figura IV.9. Modelo para el cálculo diario de la precipitación estimada

Fuente: elaboración propia

IV.1.3.5. Modelo para el cálculo diario de la precipitación estimada

Una vez obtenidos los coeficientes de regresión asociados a cada observatorio o entidad espacial, se procedió al diseño del modelo de geoproceso (Figura IV.10) que implementa los parámetros estadísticos con el fin de estimar el valor de la precipitación en aquellas estaciones que presentan lagunas en el día del análisis.

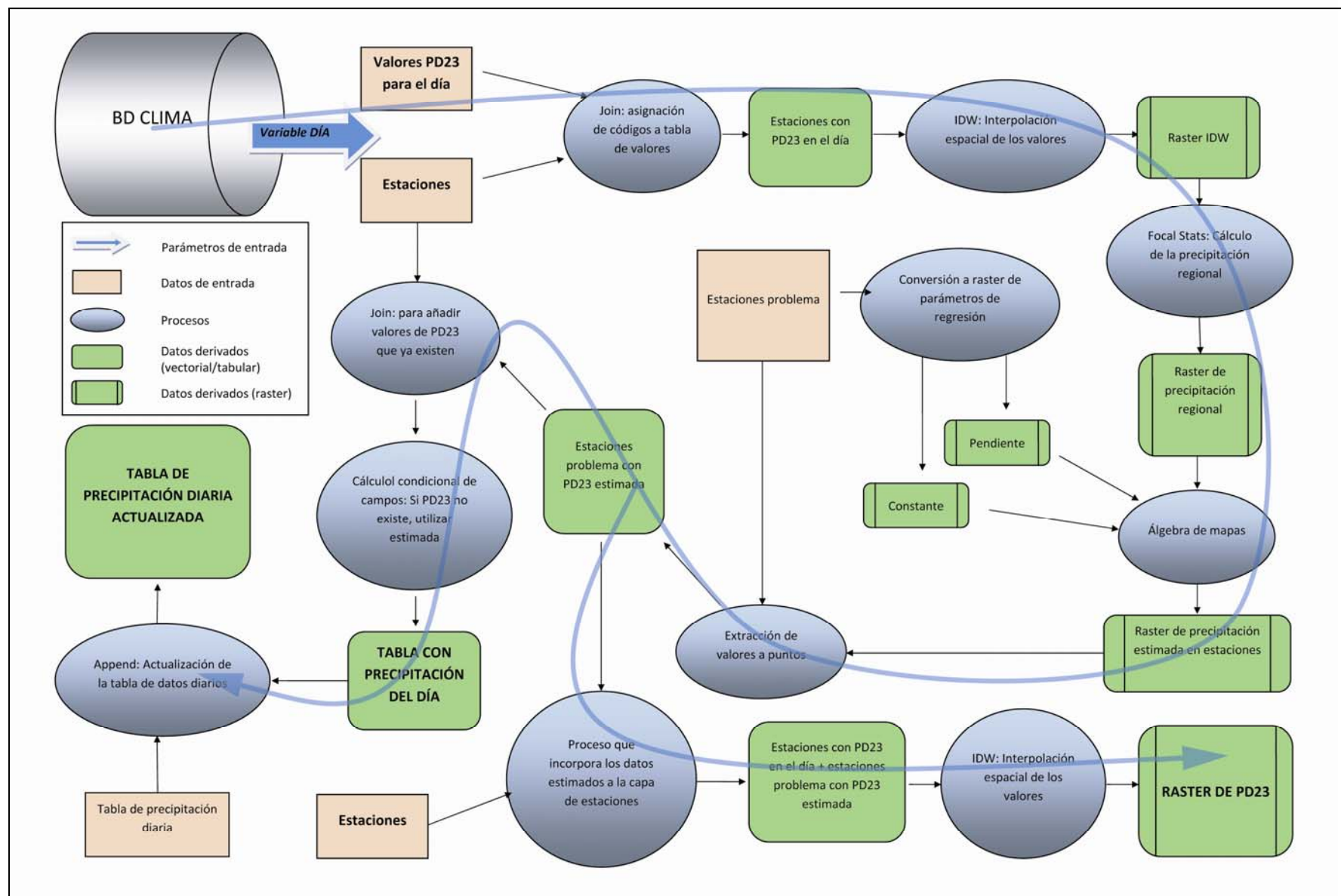


Figura IV.10. Esquema del geoproceso para la interpolación de PD_23

Fuente: elaboración propia

Los datos de entrada y los nuevos datos generados por los diferentes procesos hasta ahora descritos se almacenan en tres *geodatabases*: *Geodatabase Interpolación_día*, *Geodatabase Temporales*, *Geodatabase Salidas* (Figura IV.11). En la primera de ellas se encuentra la información de partida tanto climática –la precipitación diaria PD_23 procedente del Subsistema CLIMA-, como de información geográfica; en la segunda se realizan y guardan los procesos descritos hasta ahora y, en la última de ellas, se llevan a cabo los procesos de estimación por regresión así como los resultados –valores de precipitación diaria- obtenidos. Esta *geodatabase* final sirve, por tanto, para ejecutar la última fase del proceso que permite interpolar las lagunas a escala diaria, para cualquiera de las 78 estaciones y para cualquier fecha.

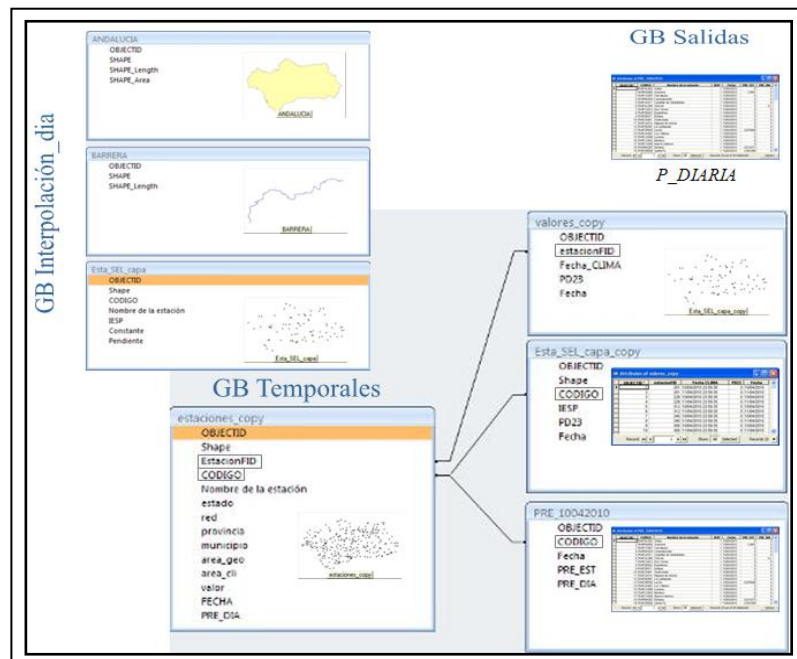


Figura IV.11. Esquema del modelo de almacenamiento de datos

Fuente: elaboración propia

Esta etapa final consta a su vez de diferentes etapas encadenadas. En primer lugar se crean dos *rasters* con la información de los coeficientes de regresión de cada estación en su celdilla correspondiente. A continuación se aplican las ecuaciones de regresión para las fechas con lagunas, empleando la herramienta de álgebra de mapas, se actualiza la tabla de datos diarios con estos valores estimados. Finalmente, se procede a la interpolación espacial de la precipitación diaria con el método del

Inverso de la Distancia Ponderado (*IDW*), utilizando tanto las estaciones con registros originales como aquellas en las que los valores se han estimado, a una resolución de 2000 m (Figuras IV.12 y IV.13).

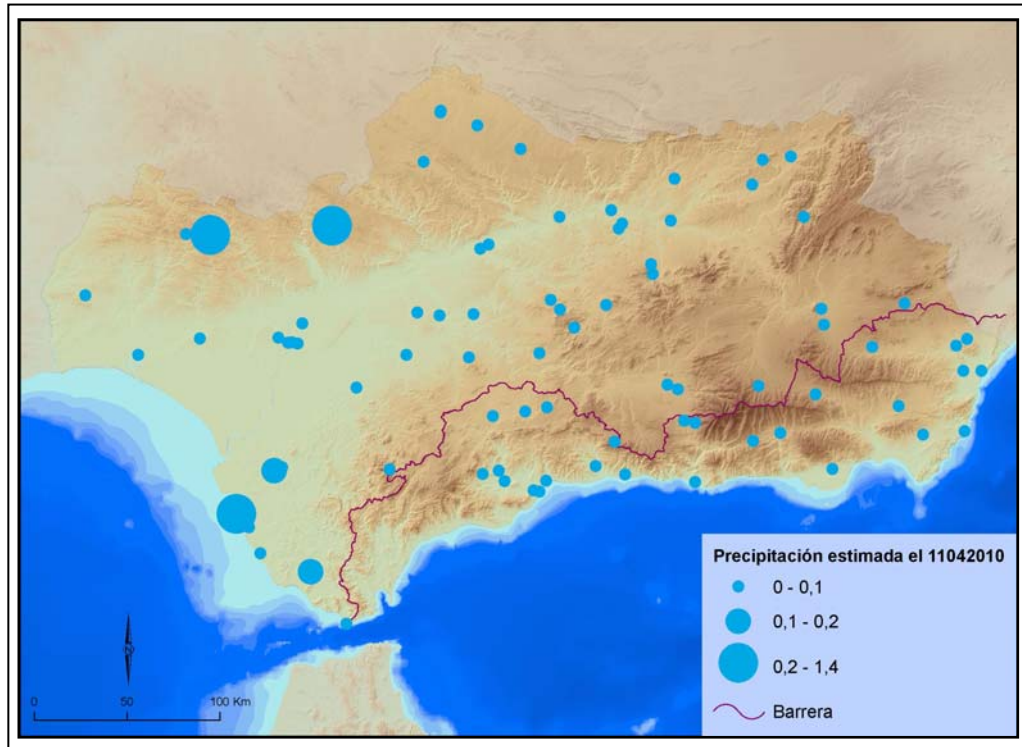


Figura IV.12. Ejemplo de precipitación estimada en las estaciones para el 11/04/2010

Fuente: elaboración propia

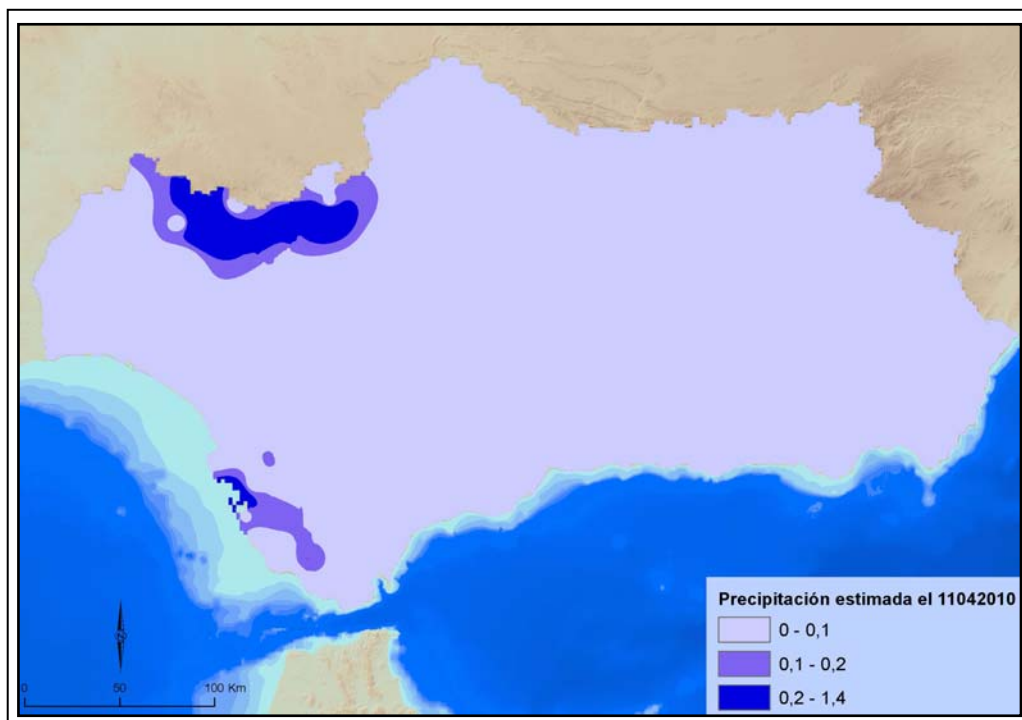


Figura IV.13. Ejemplo de *raster* de precipitación especializada para el 11/04/2010

Fuente: elaboración propia

IV.1.3.6. Validación y estimación del error de interpolación

La última etapa de este estudio aborda el control de calidad de la metodología empleada mediante dos procedimientos contrastados. En ambos se han calculado tres estadísticos de estimación del error: Error Medio (*EM*), Error Absoluto Medio (*EAM*) y Raíz cuadrada del Error Medio Cuadrático (*REMC*), cuyas expresiones matemáticas son las siguientes:

$$EM = \frac{1}{n} \sum_{k=1}^n O_k - E_k$$
$$EAM = \frac{1}{n} \sum_{k=1}^n |O_k - E_k|$$
$$REMC = \sqrt{\frac{1}{n} \sum_{k=1}^n (O_k - E_k)^2}$$

Donde, O_k y E_k son, respectivamente, el valor observado y valor estimado por el modelo en el momento k .

El primero de los procedimientos consiste en una validación cruzada entre los valores observados y los estimados según la metodología descrita en la que, como ya hemos comentado, no participa cada estación en el cálculo mediante regresión de los valores de su serie. Esta validación cruzada se realizó sobre una muestra de control de 100 días de media en el conjunto de estaciones. En este caso se esperan obtener los mejores resultados ya que el proceso está muy ajustado al entorno pluviométrico de cada observatorio. Los errores calculados mediante este modelo de regresión se comparan con los que resultan de la aplicación de otro método de estimación por la simple interpolación de los valores existentes en cada fecha con el método del inverso de la distancia (*IDW*), utilizado en los procesos de la propia Consejería de Medio Ambiente. En la mayor parte de las estaciones (60% de las mismas) el método de regresión es el que muestra los valores más bajos en todos los estadísticos de error. El restante 40 % (38 estaciones) muestra menores errores

cuadráticos utilizando el interpolador *IDW* o la propia precipitación regional. Estos casos coinciden con estaciones que se encuentran a poca distancia unas de otras y en entornos geográficos donde los mecanismos de la precipitación son más homogéneos. Por ello, el modelo final de relleno de lagunas diarias consta de dos métodos diferenciados para cada uno de los dos grupos de estaciones en virtud de su adecuación a cada uno de los dos interpoladores propuestos.

La tabla que se presenta a continuación refleja los tres estadísticos calculados para cada uno de los métodos testados en las 46 estaciones en las que el modelo de regresión ha arrojado menores valores en la prueba de validación cruzada:

Tabla IV.1. Errores medios de las estaciones en las que el modelo propuesto ha arrojado menores valores comparando los dos procedimientos de estimación.

Método de estimación	Media de <i>EM</i> (mm)	Media de <i>EAM</i> (mm)	Media de <i>REMC</i> (mm)
REGRESIÓN	0,06	2,69	6,40
<i>IDW</i>	0,75	2,89	7,15

De estos estadísticos el valor medio de *EM* mide el sesgo o la desviación de la media que han podido incorporar sobre o subestimaciones sistemáticas. La precipitación es especialmente susceptible a las consecuencias del sesgo, ya que al acumular sus valores en las escalas temporales superiores mensual y/o anual, estos errores sesgados van a ir también acumulándose en las referidas escalas. Es por ello que un buen indicador de la bondad del ajuste sea que el valor medio de *EM* sea muy cercano a cero. En la tabla se observa cómo el modelo de regresión propuesto revela en este estadístico valores muy cercanos a 0, mientras que por el contrario el interpolador *IDW* se aleja sensiblemente de este valor.

El *EAM* mide la magnitud del error medio sin tener en cuenta el signo, mientras que el *REMC* mide lo mismo, si bien dando más peso a los errores más grandes. En ambos casos, los valores son muy similares, mejorando aproximadamente el método propuesto en un 10 % al interpolador global *IDW* en términos de *REMC*. Estos resultados son comparables a otros estudios, realizados en Andalucía, en los que se han empleado interpolaciones basadas en *kriging* y en la construcción de semivariogramas empíricos para cada uno de los doce meses del año (Vanderlinden, 2001). Estos trabajos utilizan un conjunto de 164 estaciones meteorológicas termopluviométricas manuales de Andalucía –un 50% más de las disponibles en nuestro estudio- y muestran valores de *REMC* superiores a 4 mm (medio) con máximos de 6,7 mm (meses lluviosos) en el periodo 1971-1998, obtenidos también por validación cruzada, lo cual refuerza la metodología aquí propuesta y simplifica, además, el diseño e implementación automática de los procesos. Se refuerza por tanto la metodología diseñada ya que la disponibilidad de un mayor número de estaciones en el estudio aumenta la densidad espacial de la muestra y permite, por tanto, un mejor funcionamiento de los algoritmos de interpolación espacial.

Tomando los datos globalmente la mejora del método propuesto respecto al interpolador global *IDW* utilizado en los procedimientos de la Consejería es de aproximadamente un 10%.

Si adicionalmente observamos la evolución del porcentaje de mejora del método propuesto respecto al interpolador global *IDW* ($100 - ((EMC_{regresión} * 100) / EMC_{IDW})$) representado por intervalos de precipitación –global, > 1mm, > 3mm, > 5mm, > 10mm, > 15mm, > 20mm), como se refleja en la figura IV.14, puede observarse que la mejora se mantiene estable incrementándose hasta un 15% en los intervalos más altos, lo cual significa que el comportamiento es estable e independiente de los totales pluviométricos alcanzados.

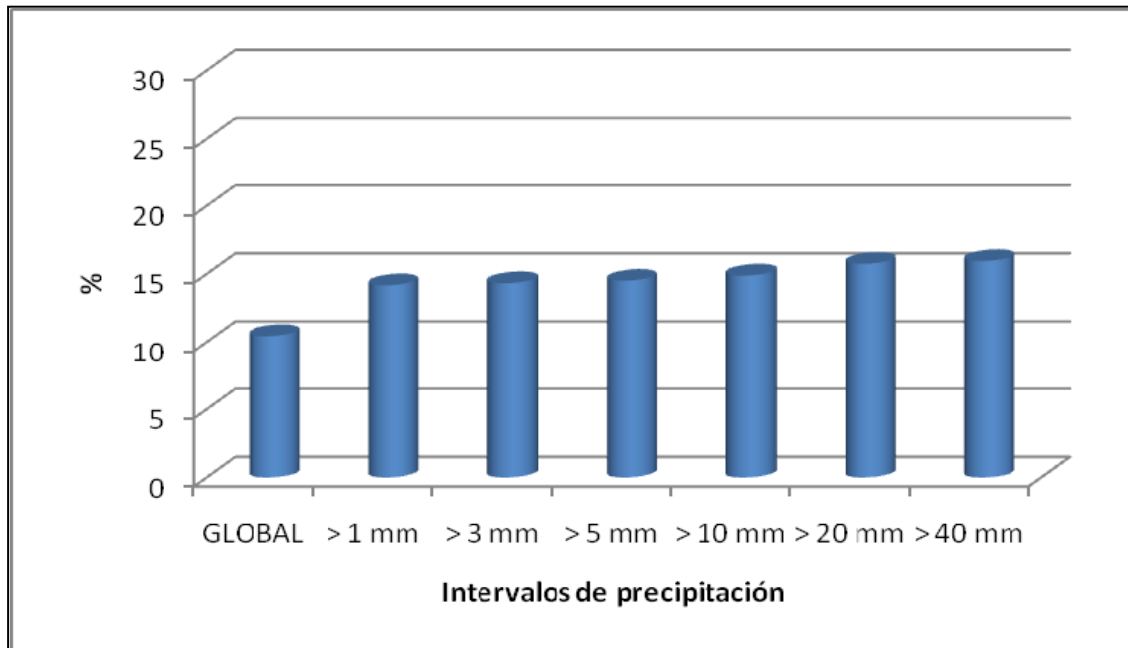


Figura IV.14. Evolución del porcentaje de mejora del método propuesto respecto al interpolador global IDW por intervalos de precipitación

Fuente: elaboración propia

Si bien se observan niveles modestos de mejora global, sí es importante destacar las mejoras del método propuesto en aquellas estaciones en las que la pendiente de la regresión es distinta de 1 y que representan aquellas estaciones en las que existen condiciones locales en la configuración y funcionamiento de los procesos que dan lugar a la precipitación que producen una infraestimación o sobrestimación sistemática del interpolador global *IDW* de los valores estimados en estas respecto a los registros medidos.

Aproximadamente el 32% de las estaciones en las que el método propuesto presenta menores errores que el interpolador global *IDW* responden a esta premisa. Tal y como se puede observar en las figuras IV.15 y IV.16, estos porcentajes progresivos de mejora son sensiblemente superiores, partiendo de un 20% de mejora global – $REMC = 5,69$ - y llegando a valores superiores al 40% de mejora en los intervalos de días con precipitaciones más voluminosas. Es precisamente este grupo de estaciones las que justifican la bondad del método propuesto ya que, en las mismas, los interpoladores globales como el *IDW* presentan errores sistemáticos debido a los mecanismos locales cuyo funcionamiento no queda registrado al utilizar las estaciones del entorno. En cambio, es la relación funcional entre el comportamiento

del entorno y los registros en esas estaciones –tal y como se aprecia en los diagramas de dispersión y en los coeficientes de determinación (r) superiores a 0,8 alcanzados, los que van a permitir una mejora sustancial en la estimación de los valores de precipitación diarios interpolados.

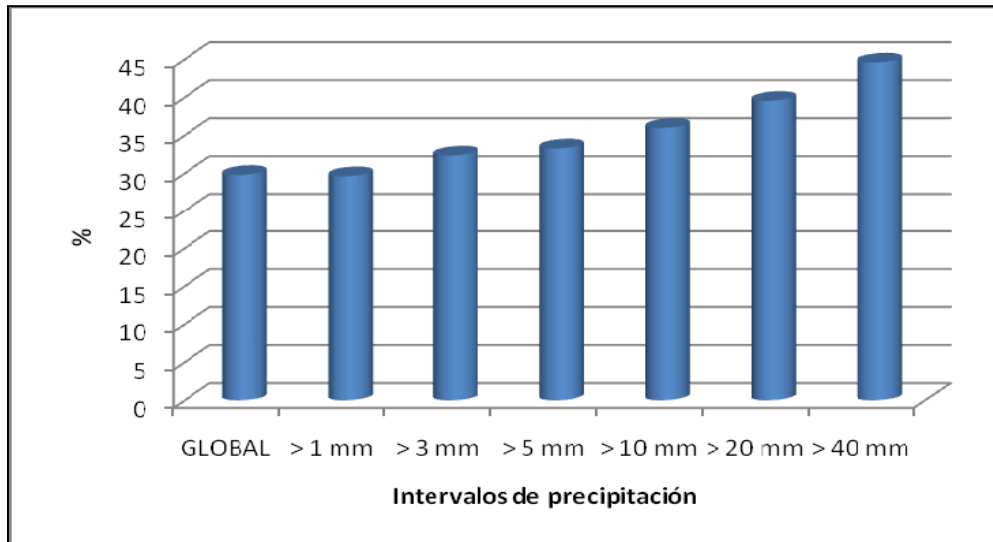


Figura IV.15. Evolución del porcentaje de mejora del método propuesto respecto al interpolador global IDW por intervalos de precipitación en las estaciones con $b < 1$ Fuente: elaboración propia

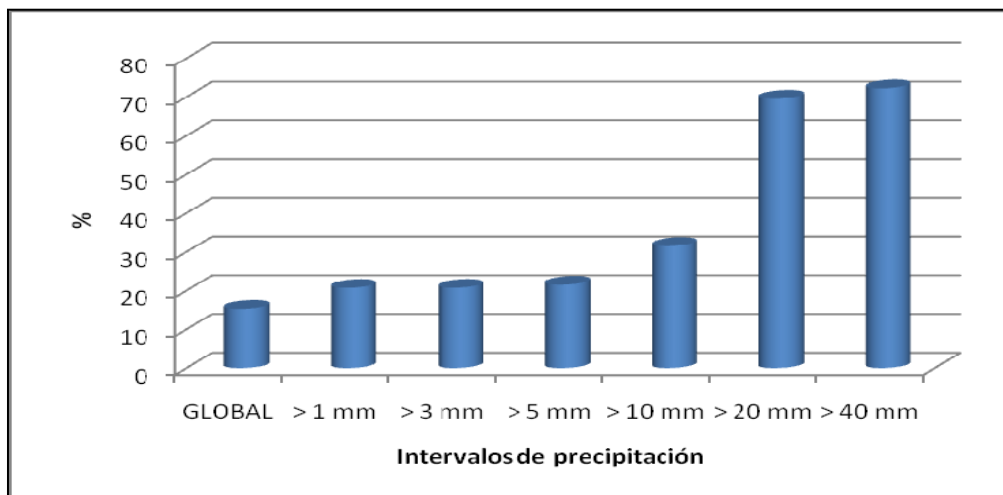


Figura IV.16. Evolución del porcentaje de mejora del método propuesto respecto al interpolador global IDW por intervalos de precipitación en la estación RIA2908 (Pizarra, Málaga) Fuente: elaboración propia

Por último es importante destacar que el error medido se suaviza al incorporarse los valores diarios estimados en el cómputo global mensual, en el que hay que recordar

que la media de días con lagunas en las estaciones de referencia no supera los 6 días y además reforzado por la práctica ausencia de sesgo con *EM* prácticamente igual a 0. No obstante, en cualquier día con el registro ausente, pueden y de hecho sucede, producirse lluvias intensas que hacen necesaria la adopción de estos métodos y especialmente en aquellas estaciones que, por sus condiciones locales, responden de forma poco ajustada a los valores calculados mediante interpoladores globales como es el *IDW*.

En este último sentido será de obligado abordaje la experimentación con la aplicación de interpoladores locales que permitan ajustar el grado de variabilidad espacial en las muestras que suponen cada día de precipitación registrada en Andalucía. Testar las tendencias globales y las posibles correcciones y efectos locales estamos convencidos que aportará mejoras en la fiabilidad de las estimaciones diarias de la precipitación.

IV.1.3.7. Conclusiones

La metodología propuesta permite solucionar uno de los problemas fundamentales que se plantean con la utilización de los datos procedentes de estaciones meteorológicas automáticas, la falta de continuidad en los registros diarios de precipitación que impiden elaborar indicadores mensuales. Los resultados obtenidos mejoran globalmente el interpolador usado en los procedimientos de la Consejería de Medio Ambiente de la Junta de Andalucía, y presentan valores muy cercanos de bondad a los de otros estudios que aplican técnicas más sofisticadas, complejas y de menor posibilidad de incorporarse como proceso automático.

Otra de las aportaciones del método propuesto es su relativa simplicidad ya que presenta como único parámetro de entrada la fecha para la que se desea ejecutar el proceso. Una vez introducida ésta, el modelo se ocupa de leer los datos descargados del Subsistema CLIMA, realizar las interpolaciones espaciales, aplicar los coeficientes de regresión, obtener los valores estimados y actualizar la tabla de datos

diarios añadiéndole los nuevos. Esta capacidad de ser implementado como un modelo procedimental en rutina permite una integración fácil en las tareas de los usuarios de la Consejería.

Por último, su aplicación es válida para toda la región y para cualquier evento pluviométrico, lo que extiende su ámbito implantación a un gran número de usuarios y tareas. Además, permite el cálculo del IESP y de otros indicadores climáticos, a escala mensual, para el seguimiento de los recursos hídricos dentro del subsistema CLIMA en Andalucía.

Los buenos resultados obtenidos animan a probar la metodología en nuevas estaciones, con un comportamiento pluviométrico complejo y en periodos diferentes a los utilizados. Asimismo abren nuevas posibilidades metodológicas que van a ser abordadas en un futuro próximo completando y perfeccionando este trabajo. Especialmente destacables son las verificaciones del comportamiento de otros interpoladores deterministas, especialmente aquellos de ajuste polinómico local (*Local Polynomial*). Este interpolador ajusta funciones polinómicas de diferente orden de forma local a partir del número de vecinos seleccionado, asumiendo que no existe una función única para todos los puntos de estimación, por lo que permite un ajuste selectivo en función de la variabilidad espacial local intrínseca de la variable a interpolar (Figura IV.17).

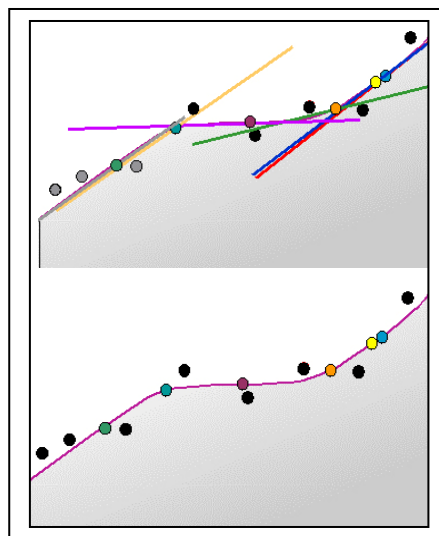
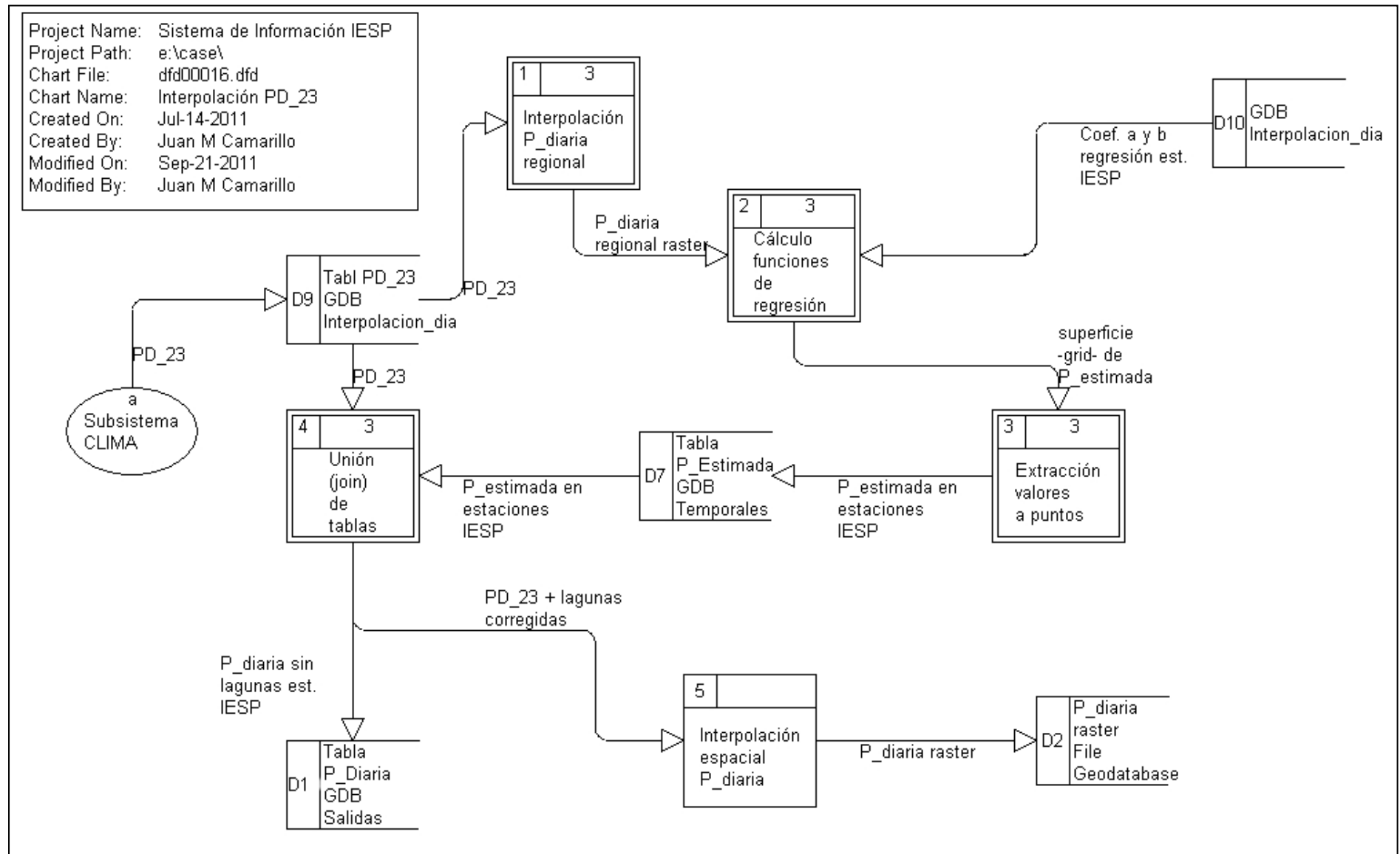


Figura IV.17. Ejemplo del funcionamiento del interpolador *Local Polynomial* (función de primer orden –recta- y superficie generada)

Fuente: Manual ArcGis 9.3

IV.1.4. Diagrama del proceso “Interpolación de PD_23 (Figura IV.18)



DFD nº 00016 Diagrama de Flujo de Datos del Proceso para la interpolación de la precipitación diaria PD_23. Nivel 2

Objetivos: Representación de los procesos y requisitos funcionales necesarios para el proceso de interpolación de la precipitación diaria (PD_23) registrada en el Subsistema CLIMA de la REDIAM (Red de Información Ambiental de Andalucía). El proceso debe:

- Interpolación de las lagunas existentes en dichos registros para cada día de cada mes en las 78 estaciones de referencia para el cálculo del IESP.
- Generar la tabla P_diaria dentro de la Geodatabase personal (GDB) P_estimada.
- Generar las superficies -raster- de la precipitación diaria como objetivo secundario solicitado por la Consejería de Medio Ambiente de la Junta de Andalucía.

Datos de partida: Los únicos datos de partida son los valores de precipitación diaria registrados en un conjunto de 297 estaciones meteorológicas automáticas presentes en las distintas redes meteorológicas que operan en Andalucía, entre las que se encuentran las 78 estaciones automáticas seleccionadas para el cálculo del IESP.

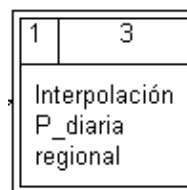
Asimismo también se incorporan los coeficientes a y b de las funciones de regresión simple calculados previamente para cada estación de las 84 de referencia entre la precipitación diaria regional -variable independiente- y la precipitación diaria registrada en la estación sobre una serie histórica diaria de 5 años.

El cálculo de las lagunas existentes se circunscribe a las 78 estaciones automáticas de referencia para el IESP ya que el cálculo del mismo exige continuidad temporal en las series.

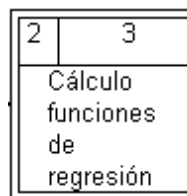
Requisitos funcionales:

- Conexión con CLIMA.
- Garantía de poder realizar el cálculo por proximidad de estaciones en caso de ausencia de registros en las estaciones independientes.
- Interpolación de la precipitación diaria regional de los días de registro de cada mes mediante procesos de espacialización -IDW- y de funciones de ventana espaciales -filtros espaciales-.
- Aplicación de álgebra de mapas sobre estructuras de datos raster para la aplicación de las funciones de regresión sobre cada estación de referencia (coeficientes a y b calculados previamente y variable independiente x - precipitación diaria regional-).
- Aplicación de procesos de estadística focal -focal statistics- para la extracción de los valores estimados a puntos y tabla de atributos de estos.
- Aplicación de procedimientos de interpolación espacial para la obtención de las superficies diarias de precipitación en Andalucía.

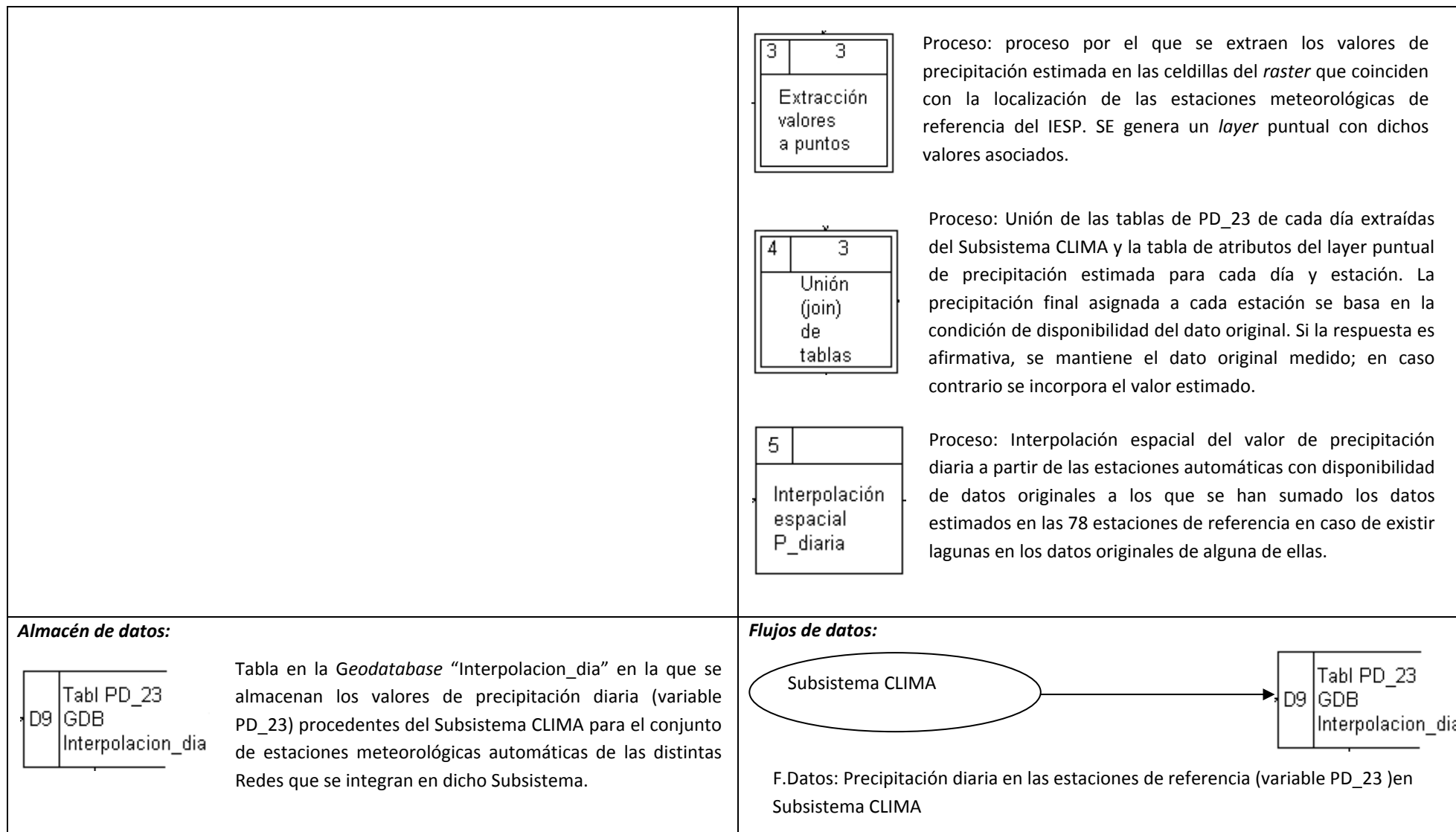
Procesos de datos:

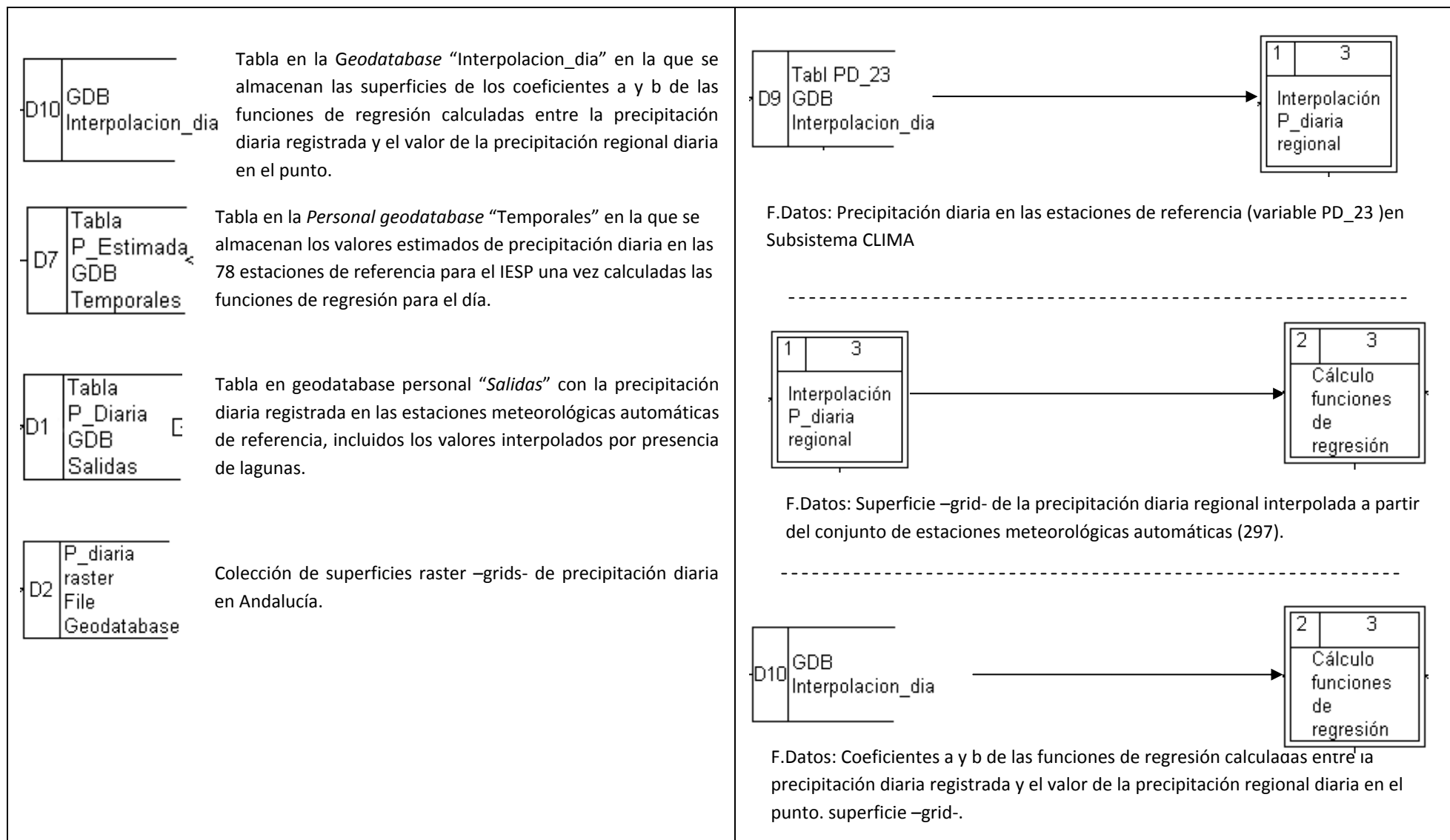


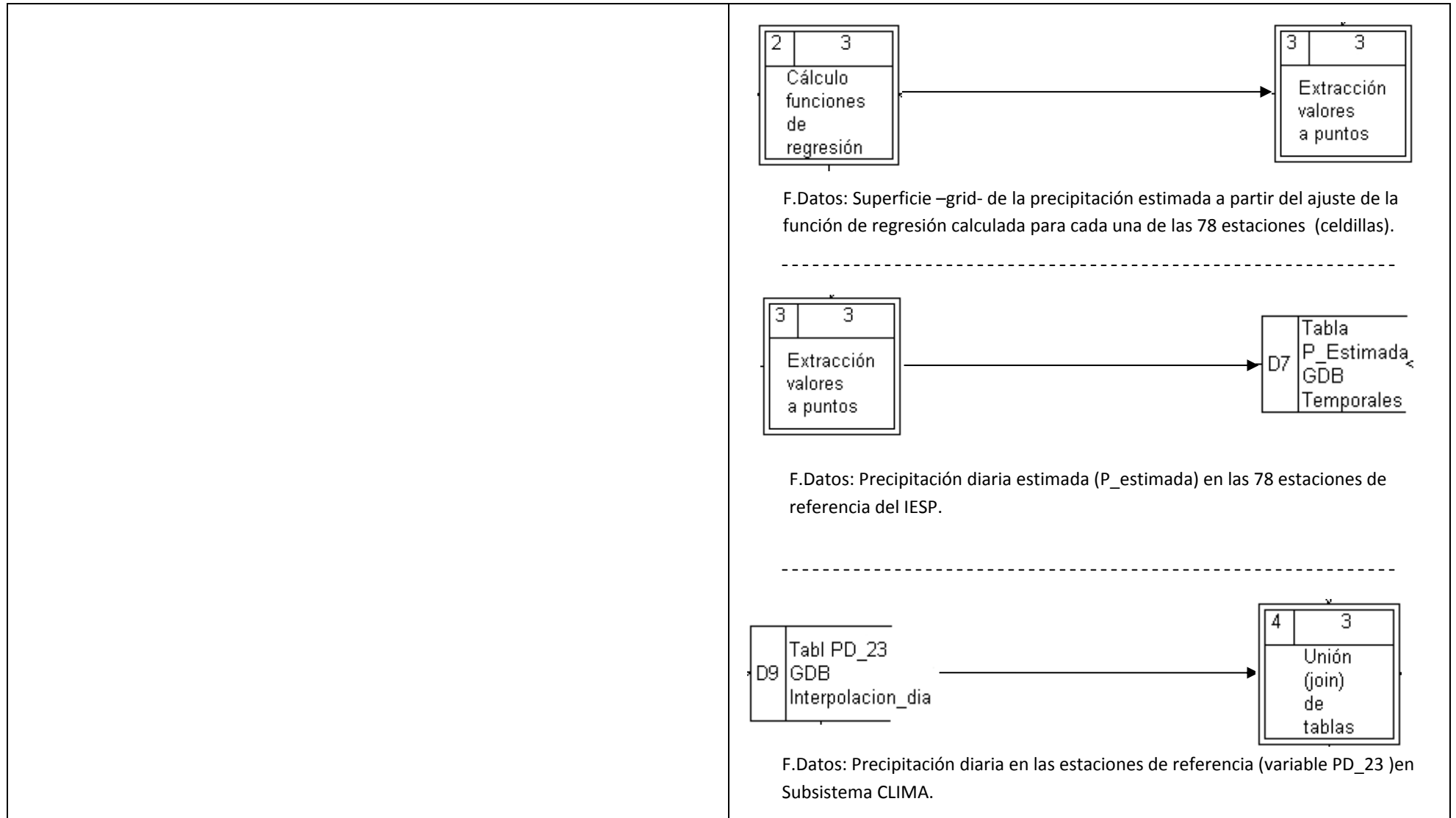
Proceso: Interpolación espacial de la precipitación diaria regional a partir de los registros de PD_23 mediante aplicación del algoritmo *Inverse Distance Weighted (IDW)* con potencia 3 y aplicación posterior de la función de ventana basada en la media de los valores en un radio de 25 Kms. Calculado para cada día del mes.

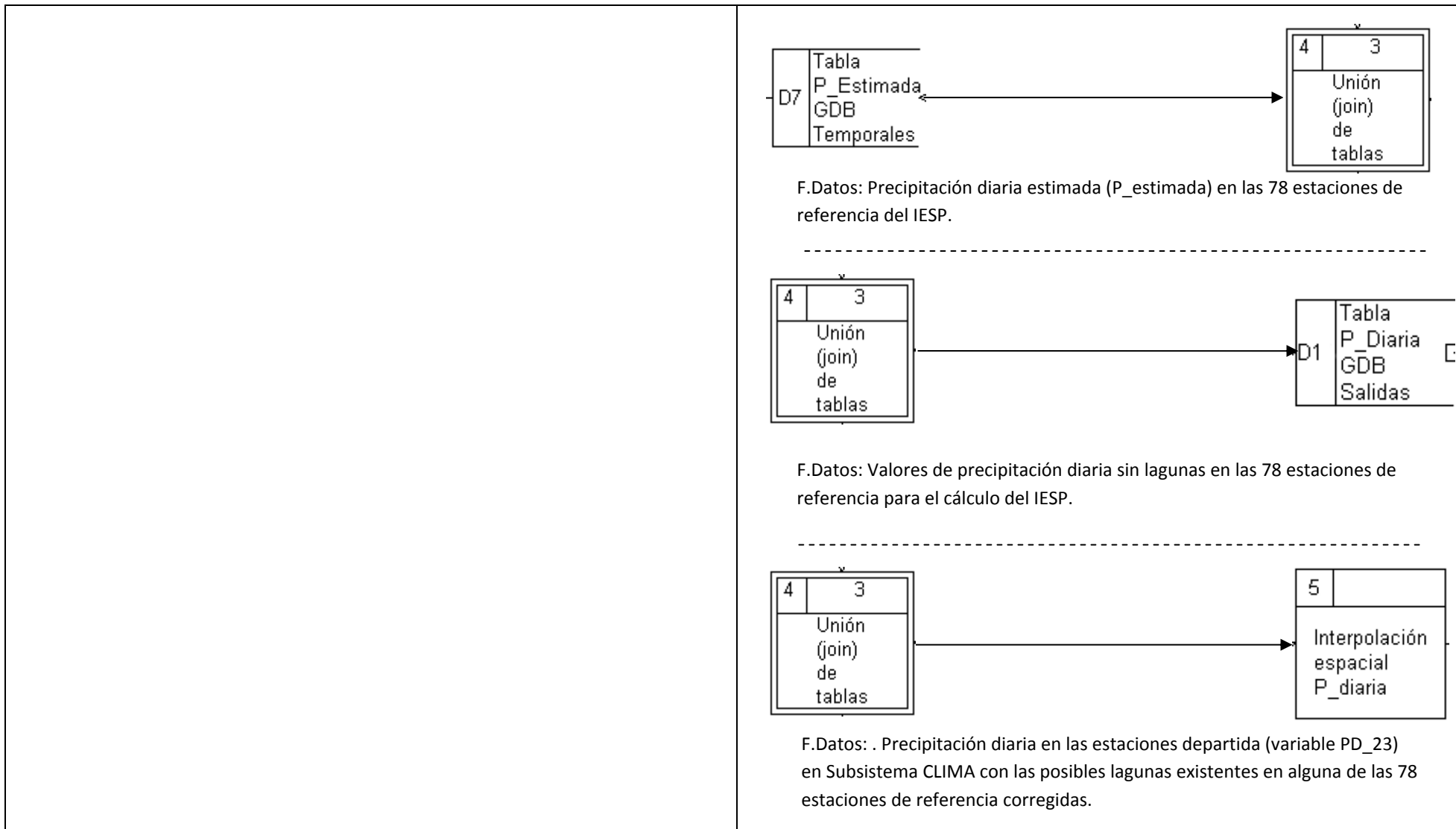


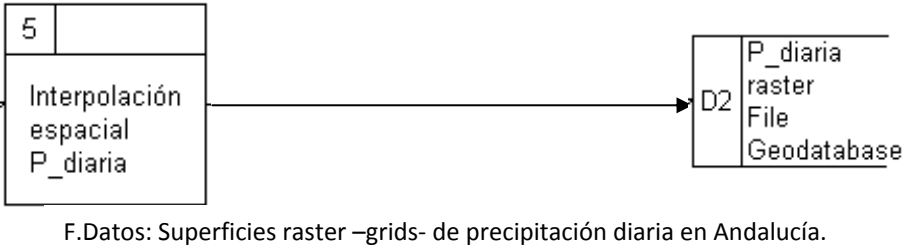
Proceso: Cálculo de las funciones de regresión para cada día de registro del mes de cálculo.



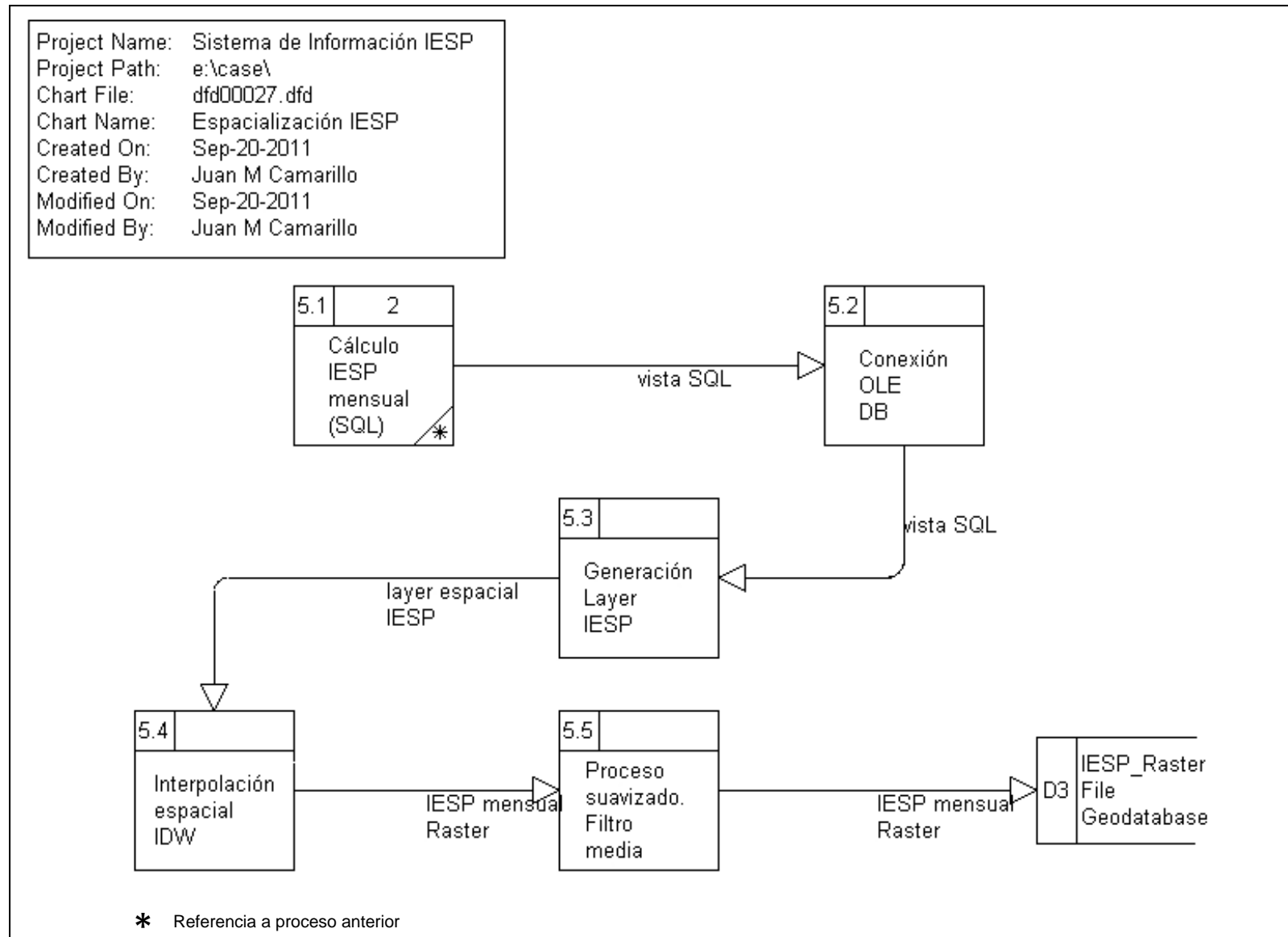






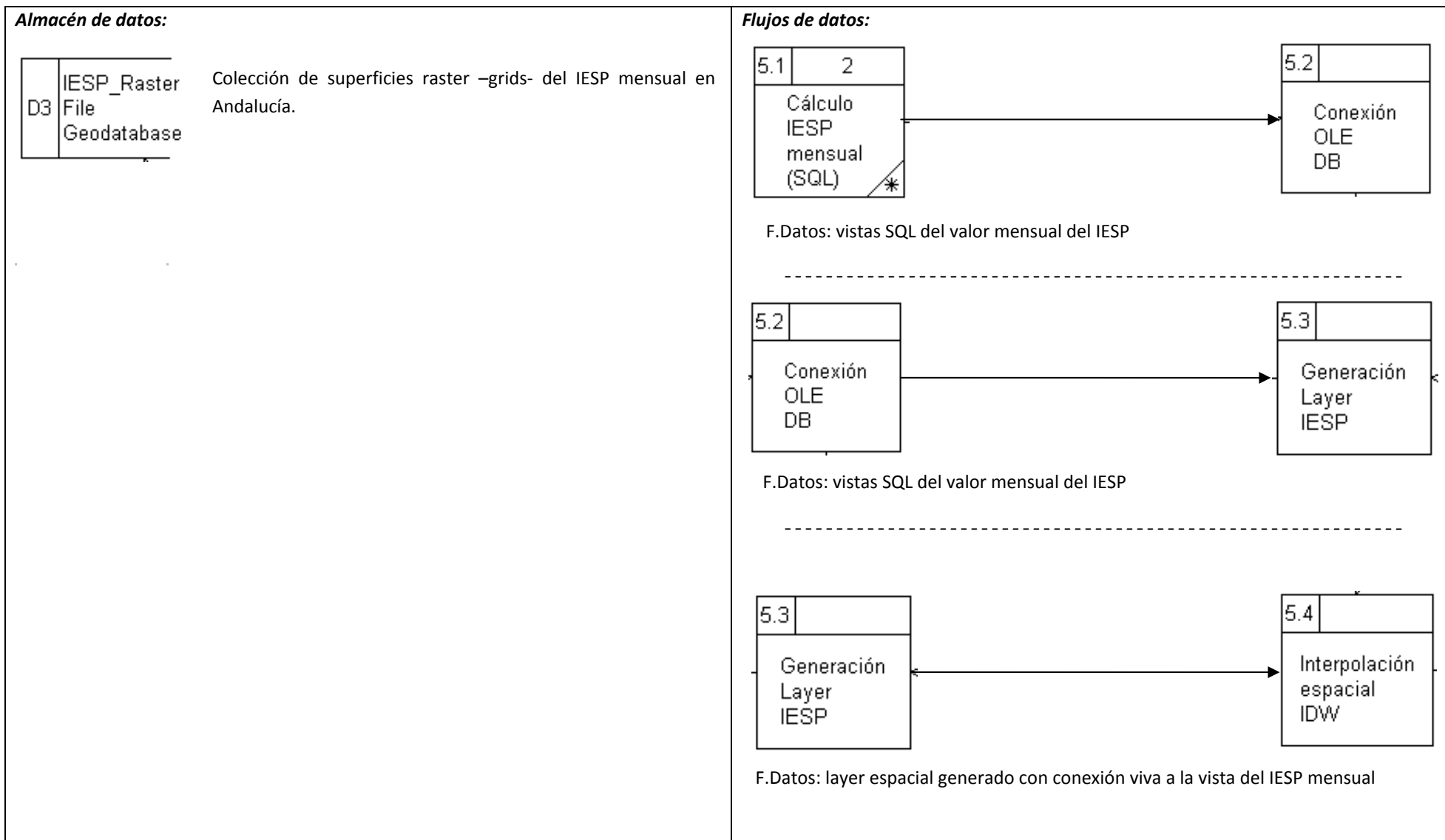
	 <p>F.Datos: Superficies raster -grids- de precipitación diaria en Andalucía.</p>
<p>Entidades externas:</p> <div style="border: 1px solid black; border-radius: 50%; width: 60px; height: 60px; display: flex; align-items: center; justify-content: center; margin: 10px auto;"> <p style="margin: 0;">Subsistema CLIMA</p> </div> <p>Subsistema de información de la REDIAM (Red de Información Ambiental de la Junta de Andalucía) que gestiona y almacena los datos meteorológicos que registran las estaciones pertenecientes al conjunto de diferentes Redes Meteorológicas que operan en Andalucía.</p>	
<p>Requisitos de naturaleza geográfica y/o climática:</p> <p>Entre los requisitos funcionales representados por este diagrama funcional del Sistema que suponen procesos directamente asociados a la naturaleza geográfica de la información climática, cabe destacar los siguientes:</p> <ol style="list-style-type: none"> 1. Conocimiento de las problemáticas asociadas al registro diario de datos meteorológicos en estaciones automáticas (intervalos de medida, detección de <i>outlyers</i>, ausencia de valores registrados, ...). 2. Necesidad de diseño de un proceso automático de interpolación de lagunas en los valores de precipitación diaria basado en el comportamiento de la precipitación en el entorno de la estación y, por tanto, con necesidades de incorporar algoritmos de interpolación espacial y de funciones matemáticas (regresión simple y múltiple), que permitan la consecución de una serie temporal continua en los valores de precipitación diarios de los que se derivarán los mensuales necesarios para el Índice. 3. Diseño de un proceso automático de interpolación de la precipitación diaria. 4. Necesidad de manejar información en formato <i>raster</i> y de herramientas de análisis espacial <i>raster -spatial analyst-</i>. 	

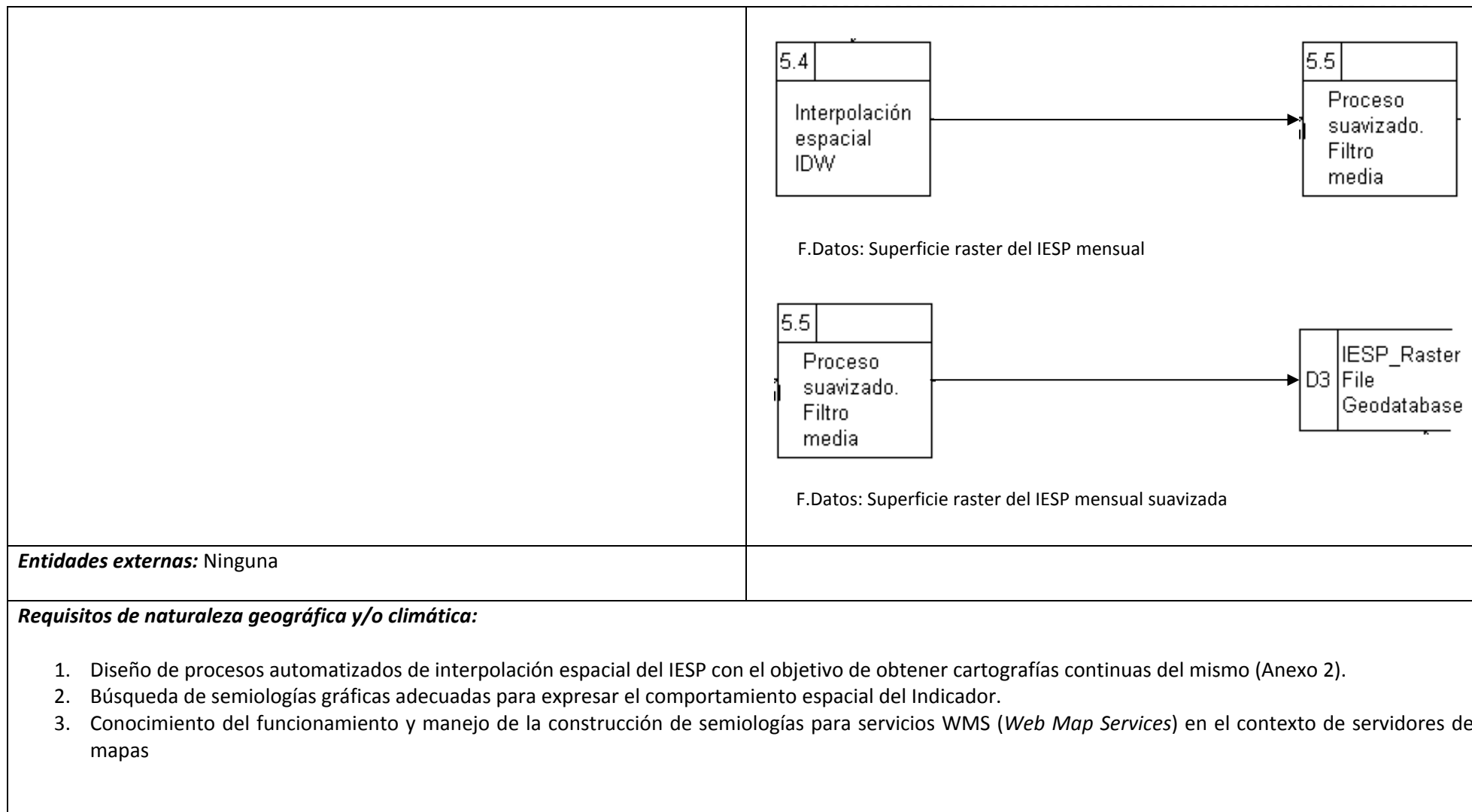
IV.1.5. Diagrama del proceso “Espacialización IESP” (Figura IV.19)



<u>DFD nº 00027 Diagrama de Flujo de Datos del Proceso para el cálculo de los pronósticos a 12 meses del IESP. Nivel 2</u>											
<p>Objetivos: El proceso descrito a continuación tiene como objetivo integrar en el sistema el procedimiento de espacialización de los valores del IESP mensual con el propósito de generar una cartografía continua del fenómeno.</p>	<p>Datos de partida: La vista (SQL) diseñada en PostGis para el cálculo del valor mensual del IESP en cada uno de los meses de cálculo.</p>										
<p>Requisitos funcionales:</p> <ul style="list-style-type: none"> - Conexión entre servidores (conexiones OLE DB) para integrar en el software GIS utilizado –ArcGis-, encargado de realizar los procedimientos de análisis espacial necesarios, los valores mensuales del IESP asociados a sus correspondientes estaciones –layer espacial-. - Disponibilidad de algoritmos de interpolación y de filtros –media- necesarios para la generación de las superficies <i>raster</i> deseadas. 	<p>Procesos de datos:</p> <div style="display: flex; justify-content: space-between; align-items: flex-start;"> <div style="border: 1px solid black; padding: 5px; margin-bottom: 20px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20px; text-align: center;">5.1</td> <td style="width: 20px; text-align: center;">2</td> </tr> <tr> <td colspan="2" style="text-align: center;">Cálculo IESP mensual (SQL)</td> </tr> <tr> <td colspan="2" style="text-align: right;">*</td> </tr> </table> </div> <div style="flex-grow: 1;"> <p>Proceso: generación de la vista de los valores del IESP mensual en el mes de cálculo del indicador a partir de las consultas SQL definidas en el entorno PostGis.</p> </div> </div> <div style="display: flex; justify-content: space-between; align-items: flex-start;"> <div style="border: 1px solid black; padding: 5px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20px; text-align: center;">5.2</td> <td></td> </tr> <tr> <td colspan="2" style="text-align: center;">Conexión OLE DB</td> </tr> </table> </div> <div style="flex-grow: 1;"> <p>Proceso: proceso de debe establecer la conexión entre el servidor PostGis y la geodatabase personal en la que se almacena la feature class de estaciones meteorológicas para la construcción del layer necesario para la ejecución del geoproceso de interpolación (Anexo 2).</p> </div> </div>	5.1	2	Cálculo IESP mensual (SQL)		*		5.2		Conexión OLE DB	
5.1	2										
Cálculo IESP mensual (SQL)											
*											
5.2											
Conexión OLE DB											

	<table border="1"> <tr> <td data-bbox="1189 272 1245 323">5.3</td> <td data-bbox="1245 272 1400 323"></td> <td data-bbox="1424 292 2110 357">Proceso: proceso por el que construye el <i>layer</i> (“capa” con la conexión viva a la vista IESP mensual de <i>PostGis</i>)</td> </tr> <tr> <td colspan="2" data-bbox="1189 323 1400 480">Generación Layer IESP</td> <td></td> </tr> </table> <table border="1"> <tr> <td data-bbox="1189 608 1245 659">5.4</td> <td data-bbox="1245 608 1400 659"></td> <td data-bbox="1424 627 2110 756">Proceso: proceso por el que se generan las cartografías continuas orientadas a la difusión del comportamiento del IESP en la Comunidad Autónoma de Andalucía. Interpolador utilizado: <i>IDW</i></td> </tr> <tr> <td colspan="2" data-bbox="1189 659 1400 815">Interpolación espacial IDW</td> <td></td> </tr> </table> <table border="1"> <tr> <td data-bbox="1189 943 1245 994">5.5</td> <td data-bbox="1245 943 1400 994"></td> <td data-bbox="1424 930 2110 1187">Proceso: proceso por el que las superficies raster son suavizadas para evitar el efecto “ojo de buey” característico de este interpolador. El objetivo es fundamentalmente mejorar la distribución visual del valor del IESP, cuya variabilidad espacial tiende, al ser un fenómeno de alcance regional, a presentar valores no muy extremos. La función utilizada a sido un filtro de paso bajo como la media</td> </tr> <tr> <td colspan="2" data-bbox="1189 994 1400 1150">Proceso suavizado. Filtro media</td> <td></td> </tr> </table>	5.3		Proceso: proceso por el que construye el <i>layer</i> (“capa” con la conexión viva a la vista IESP mensual de <i>PostGis</i>)	Generación Layer IESP			5.4		Proceso: proceso por el que se generan las cartografías continuas orientadas a la difusión del comportamiento del IESP en la Comunidad Autónoma de Andalucía. Interpolador utilizado: <i>IDW</i>	Interpolación espacial IDW			5.5		Proceso: proceso por el que las superficies raster son suavizadas para evitar el efecto “ojo de buey” característico de este interpolador. El objetivo es fundamentalmente mejorar la distribución visual del valor del IESP, cuya variabilidad espacial tiende, al ser un fenómeno de alcance regional, a presentar valores no muy extremos. La función utilizada a sido un filtro de paso bajo como la media	Proceso suavizado. Filtro media		
5.3		Proceso: proceso por el que construye el <i>layer</i> (“capa” con la conexión viva a la vista IESP mensual de <i>PostGis</i>)																	
Generación Layer IESP																			
5.4		Proceso: proceso por el que se generan las cartografías continuas orientadas a la difusión del comportamiento del IESP en la Comunidad Autónoma de Andalucía. Interpolador utilizado: <i>IDW</i>																	
Interpolación espacial IDW																			
5.5		Proceso: proceso por el que las superficies raster son suavizadas para evitar el efecto “ojo de buey” característico de este interpolador. El objetivo es fundamentalmente mejorar la distribución visual del valor del IESP, cuya variabilidad espacial tiende, al ser un fenómeno de alcance regional, a presentar valores no muy extremos. La función utilizada a sido un filtro de paso bajo como la media																	
Proceso suavizado. Filtro media																			





IV.1.6. Metodología de cálculo del pronóstico de IESP mensual en 1,2,...,12 meses y probabilidades asociadas a dichos pronósticos

El diseño de un sistema para alertas por sequías pluviométricas debe obligatoriamente medir desde una perspectiva estadística las probabilidades de acontecimientos futuros, en horizontes temporales amplios, de situaciones secas. Este es el objetivo de la metodología propuesta, diseñada por el equipo del Dr. Carrizosa con la asistencia del propio autor de la tesis y de los investigadores involucrados en el marco del proyecto de investigación financiado por la Consejería de Medio Ambiente de la Junta de Andalucía.

La estandarización que supone la escala del IESP y que, por tanto, permite la comparación entre las distintas series correspondientes a cada una de las estaciones de referencia utilizadas, presenta distintos patrones de distribución que quedan reflejados en la figura IV.20.

En ellos podemos encontrar un primer grupo importante de estaciones que presenta distribuciones normales si bien con un componente leptocúrtico muy acusado, tal y como reflejan un índice de curtosis superior a 0. No obstante, la característica más común es el ajuste a diferentes modelos de distribución –testado mediante *Kolmogorov-Smirnov*², entre los que se repiten algunas funciones de densidad.

Un ejemplo de diferentes ajustes de las series de IESP puede observarse en la figura IV.20

En la figura IV.20a se muestra un ajuste a la función de densidad de la curva normal con un grado de leptocurtosis (coeficiente de curtosis=0,923). La figura IV.20b muestra el ajuste a una función log normal de 3 parámetros. La figura IV.20c ajusta a una función de *Weibull* de 3 parámetros, mientras que la figura IV.20d lo hace a una función *Gamma* incompleta de 3 parámetros y, finalmente, la figura IV.20e ajusta a una función de *Laplace*.

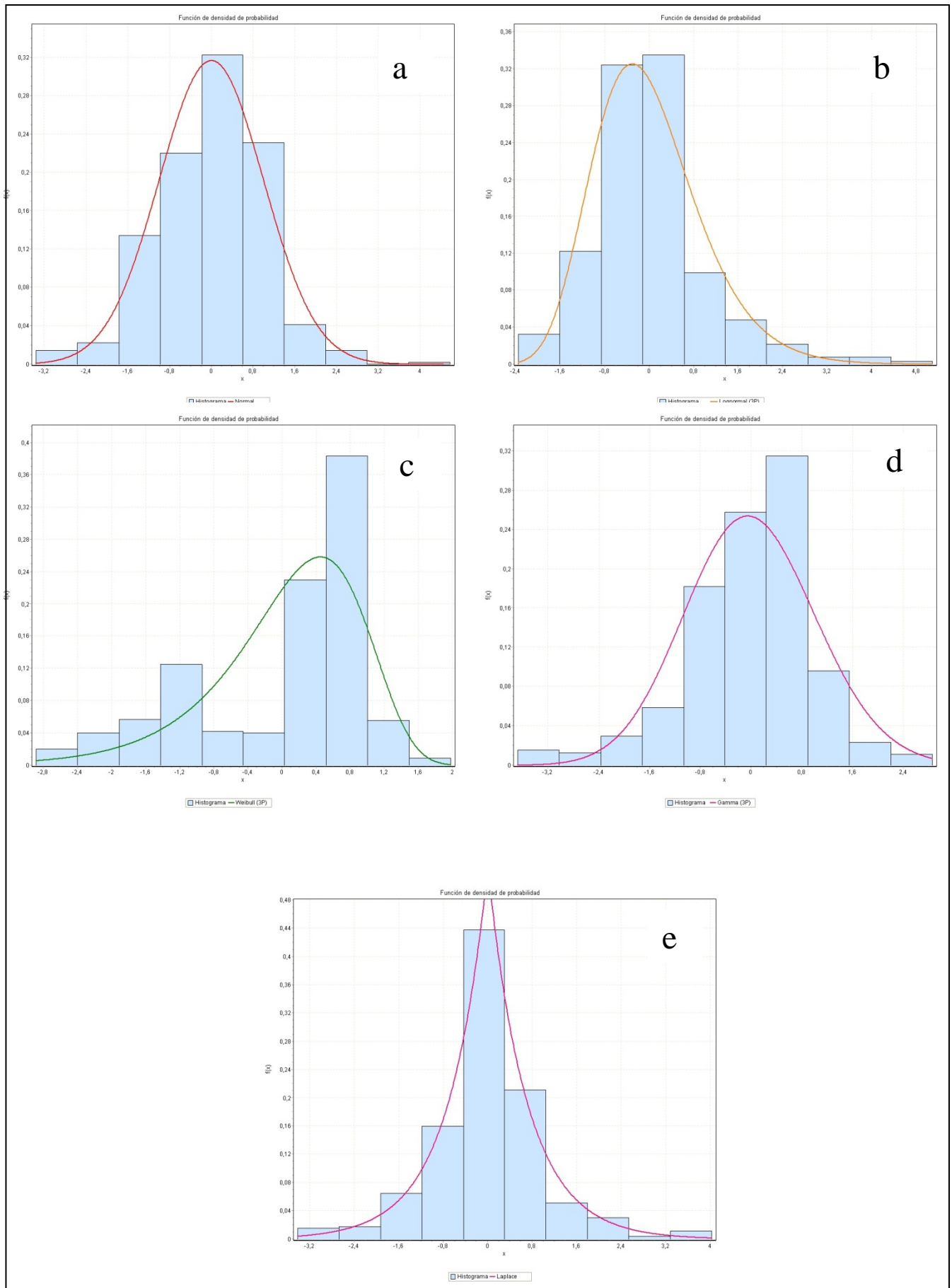


Figura IV.20. Ejemplos de ajustes de series de IESP a funciones de densidad

Fuente: elaboración propia

La diversidad de ajustes que pueden observarse, casi uno para cada serie de IESP y la dificultad de integrar esta variabilidad en los procesos de rutina diseñados, exigían la búsqueda y aplicación de un método global para el conjunto de las series del IESP que permitiera, con un grado de fiabilidad estadística demostrado en la literatura y adaptado a nuestras series, la predicción a 12 meses en el comportamiento del Indicador y la asignación de valores de probabilidad a dichos pronósticos. El diseño del modelo estadístico propuesto es abordado en el epígrafe que sigue a continuación.

IV.1.6.1. Diseño del modelo estadístico

Para el cálculo de dichos pronósticos se ha diseñado un modelo estadístico de predicción dinámica sobre la sequía en Andalucía y su persistencia con un horizonte temporal de $r=12$ meses. Se ha realizado un análisis independiente en cada una de las estaciones, quedando pendiente para estudios futuros explotar la estructura geoespacial de los datos para obtener predicciones aún más precisas.

El método se basa en el empleo de un algoritmo *kNN* (del inglés, *k nearest neighbors* (Fix y Hodges, 1951)). La elección de esta metodología está motivada por el hecho de que ésta está sobradamente contrastada en la literatura estadística, (véase, por ejemplo, Bannayan y Hoogenboom, 2008; Brabec y Meister, 2001; Ouarda et al, 1999), y resulta muy atractiva tanto por su simplicidad como por su capacidad para predecir correctamente fenómenos complejos. Intuitivamente, el algoritmo *kNN* busca aquellas *k* situaciones pasadas más similares a la actual, y predice el futuro promediando lo que ocurrió a continuación en dichas situaciones pasadas similares.

Se trata de un método de clasificación supervisada no paramétrico especialmente empleado en fenómenos de los que a priori no se conoce su modelo de distribución o su ajuste presenta dificultades como es nuestro caso y ha quedado explicado anteriormente. Se trata por tanto de trabajar en la muestra empírica y generar proyecciones y predicciones basadas en la similitud entre las cadenas de vecinos más

próximos presentes en dicha muestra respecto a la cadena actual objeto de predicción.

Es decir, es un método de clasificación supervisada basado en el aprendizaje a partir del establecimiento de un conjunto de entrenamiento o prototipos –fase de aprendizaje–, cuyos resultados van a aplicarse en la fase de predicción– que servirá para estimar la función de densidad $F(x/C_j)$ de las predictoras x en cada clase o cadaena C_j .

Es también un método no paramétrico que estima la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento x pertenezca a la clase C_j a partir de la información proporcionada por el conjunto de prototipos. En este proceso de aprendizaje, por tanto, no se hace ninguna suposición acerca de la distribución de las clases predictoras, lo cual, en nuestro caso, es de gran ayuda, ya que podemos asociar predicciones y probabilidades a cada una de nuestras series de referencia, una por cada estación meteorológica utilizada a partir del aprendizaje empírico sobre cada una de ellas independientemente del modelo de distribución al que pudieran ajustarse.

IV.I.6.2. Asunciones del método *knn*

Las asunciones principales que deben tenerse en cuenta a la hora de trabajar con el algoritmo propuesto son las siguientes:

1. Los puntos o valores tomados en consideración se localizan en un espacio euclideo n -dimensional (E^n), por lo que puede aplicarse en ellos el concepto de distancia euclidea normalmente con el fin de calcular la cercanía o lejanía de las cadenas vecinas.
2. Los datos considerados en la fase de aprendizaje consisten en una serie de vectores y de etiquetas de clase asociadas a cada uno de ellos. En nuestro caso estas clases serán + (valores positivos del IESP) y – (valores negativos del IESP).

3. Se trabaja con un número k de vecinos. Este número decide cuántos vecinos, definidos a partir de la distancia métrica, influyen en la clasificación de las cadenas de vectores. Si $k=1$ entonces el algoritmo se conoce simplemente como algoritmo del vecino más cercano.

IV.1.6.3. Fase de aprendizaje

Como hemos dicho, el método se basa en una fase de aprendizaje y en una fase de predicción

De cada una de las estaciones se tienen los datos históricos del indicador de sequía IESP, que se usa para predecir periodos de sequía (indicador IESP negativo) o sin ella (indicador IESP positivo).

En la fase de aprendizaje se han determinado tres parámetros que se usarán en el modelo de predicciones:

1. v (ancho de la ventana): se trata del ancho de la cadena –número de meses que la componen- óptimo para establecer las comparaciones entre los k vecinos más similares. Son virtualmente la d -dimensiones presentes en la estimación.
2. k (número de vecinos): número de cadenas óptimo para establecer las similitudes con la cadena objeto de predicción.
3. w (peso del mes). Valor de ponderación en función del último mes que compone la cadena objeto de predicción.

Los valores de estos tres parámetros han sido obtenidos usando un procedimiento de validación descrito más adelante, y deben ser recalculados al menos una vez al año. La fase de predicción se ejecutará, con los valores de los parámetros (v,k,w)

disponibles, mensualmente, ajustando las predicciones sobre sequía en los $r=12$ meses siguientes.

Los parámetros (v,k,w) , se fijan en cada estación de modo que se optimice una medida del error global del proceso de estimación en dicha estación. Para ello, se divide la base de datos en dos, que llamaremos *muestra de aprendizaje* y *muestra de validación*, conteniendo aproximadamente el 66% de los datos más antiguos y el 33% de los datos más recientes, respectivamente. Para cada elección de parámetros (v,k,w) , se evalúa sobre la muestra de validación la regla obtenida si consideramos la muestra de aprendizaje como base para construir las secuencia vecinas, y se calculan las tasas de acierto

- $n_1(S)$ = número de secuencias en las que se predice correctamente que el próximo periodo es de sequía
- $n'_1(S)$ = número de secuencias en las que se predice correctamente que no ocurre que el próximo periodo es de sequía
- $n_2(S)$ = número de secuencias en las que se predice correctamente que los dos próximos periodos son de sequía
- $n'_2(S)$ = número de secuencias en las que se predice correctamente que no ocurre que los dos próximos periodos son de sequía
- $n_3(S)$ = número de secuencias en las que se predice correctamente que los tres próximos periodos son de sequía
- $n'_3(S)$ = número de secuencias en las que se predice correctamente que no ocurre que los tres próximos periodos son de sequía
- ...
- $n_r(S)$ = número de secuencias en las que se predice correctamente que los r próximos periodos son de sequía

- $n^r(S)$ = número de secuencias en las que se predice correctamente que no ocurre que los r próximos periodos son de sequía

Las magnitudes n^r son objeto de una corrección durante el proceso de determinación de las mismas. Dicha corrección está encaminada a conseguir una elección de parámetros que potencie la detección de la finalización de un periodo de sequía, por considerarse éste un hecho de especial relevancia. Así, cuando en una secuencia dada se predice correctamente que no ocurre que los t próximos periodos sean de sequía, partiendo de una situación en la cual los $t-1$ próximos periodos lo son, $n^t(S)$ se incrementa en una cantidad adicional $p = 1$.

Finalmente se escogen los valores (v,k,w) para los que se hace máximo un indicador $f(v,k,w)$ de fiabilidad del método:

$$f(v,k,w) = n^1(S) + n^2(S) + b * n^3(S) + n^4(S) + b^2 * n^5(S) + \dots + n^r(S) + b^{(r-1)} * n^r(S)$$

El parámetro b se ha elegido 0.9. Para encontrar los valores óptimos de los parámetros (v,k,w) , se ha realizado una *búsqueda en rejilla* (véase Pintér, 1996): se ha evaluado $f(v,k,w)$ para una rejilla de valores (v,k,w) : v y k han variado en el conjunto $\{1,2,..,10\}$, y w en el conjunto $\{1/10,2/10,..,10/10\}$, seleccionándose la terna que haya dado mayor valor.

IV.1.6.4. Fase de predicción

El modelo realiza predicciones sobre los siguientes parámetros:

- $P1(S)$ = Probabilidad de que el próximo periodo sea de sequía (Indicador IESP negativo)
- $P2(S)$ = Probabilidad de que los dos próximos periodos sean de sequía (Indicador IESP negativo en los dos próximos periodos)

- $P_3(S)$ = Probabilidad de que los tres próximos periodos sean de sequía (Indicador IESP negativo en los tres próximos periodos)
- ...
- $Pr(S)$ = Probabilidad de que los r próximos periodos sean de sequía (Indicador IESP negativo en los r próximos periodos)

El procedimiento de estimación consta de los siguientes pasos:

1. Crear una base de datos en la que los registros son de la forma $(t, I(t), I(t+1), \dots, I(t+v-1))$, donde
 - t es el mes del año
 - $I(t)$ es el valor del IESP en el mes t
 - $I(t+1)$ es el valor del IESP en el mes $t+1$
 - ...
 - $I(t+v-1)$ es el valor del IESP en el mes $t+v-1$

En otras palabras, los registros de esta base de datos, que llamaremos *cadena*, son vectores con $v+1$ componentes, siendo la primera componente un mes, y las restantes los valores del IESP en los meses $t, t+1, \dots, t+v-1$.

2. Dada la cadena c , buscar el conjunto C de k cadenas en la base de datos más próximas a c , donde la proximidad se mide según la distancia euclídea: la proximidad $d(t, t')$ entre las cadenas $(t, I(t), I(t+1), \dots, I(t+v-1))$ y $(t', I(t'), I(t'+1), \dots, I(t'+v-1))$ viene dada por

$$d(t, t') = w(s(t)-s(t'))^2 + (I(t)-I(t'))^2 + \dots + (I(t+v-1)-I(t'+v-1))^2,$$

donde $s(t)$ y $s(t')$ representan los valores de t, t' pasados, en primer lugar, a la escala 1,2,...,12, y luego estandarizados para que la serie tenga media cero y

desviación típica 1. De esta forma, si las k cadenas seleccionadas para la estimación coinciden en su último mes con el último mes de la cadena a predecir, $w=0$, mientras que cuanto más se alejen mayor será el valor de w ($w>0$) y por lo tanto la distancia calculada para cada cadena C será mayor, lo que supone que se le da mayor cercanía a las k cadenas que se encuentren en el mismo momento del año que la cadena objeto de predicción.

3. Estimar los parámetros como:

- $P1(S)$ = porcentaje de secuencias en C en las que el siguiente periodo fue de sequía
- $P2(S)$ = porcentaje de secuencias en C en las que los dos siguientes periodos fueron de sequía
- $P3(S)$ = porcentaje de secuencias en C en las que los tres siguientes periodos fueron de sequía
- ...
- $Pr(S)$ = porcentaje de secuencias en C en las que los r siguientes periodos fueron de sequía

Estos porcentajes son los que nos permiten asignar la predicción del IESP para los siguientes 12 meses ($r=1,..,12$) en función del método de clasificación basado en la regla de los k vecinos más cercanos (*k-nearest neighbors rule*) que asigna la clase (IESP negativo o IESP positivo) a partir de la mayoría –clase más frecuente o porcentaje $> 50\%$ - computada en las k cadenas vecinas más cercanas tomadas en cuenta.

La probabilidad de la predicción de un caso x puede establecerse a partir del cómputo de la función de densidad $p(x)$ o bien por aproximación posterior dependiendo de la distribución de las predicciones realizadas $-P(\hat{c}_i | x)$.

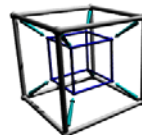
La función de densidad para un caso viene dada por

$$p(x) \approx \frac{k/n}{V}$$

donde k es el número de vecinos seleccionados, n es el conjunto de secuencias C disponibles y V es el volumen del hipercubo⁴ o de la hiperesfera centrado en x y acrecentado hasta que el número k de vecinos es incluido. En los análisis sobre variables unidimensionales enfrentadas a prototipos, lógicamente, el hipercubo o la hiperesfera se convierten en un cuadrado y un círculo respectivamente. Por tanto, la función de densidad es dependiente del volumen de dicho cubo. Si la distancia respecto a los vecinos es pequeña, la densidad será mayor –volumen inferior del cubo-, mientras que si los vecinos se encuentran a mayor distancia el hipercubo que los capturaría tendría un volumen mayor y por lo tanto la densidad sería inferior. Estudios acerca del comportamiento de la función de densidad calculada sobre algoritmos de knn no recomiendan el cálculo de la misma por varias razones. Entre ellas cabe destacar por un lado, la necesidad de tamaños muestrales muy altos ya que, en teoría, solo si un número infinito de muestras está disponible, podríamos construir una serie de estimaciones que converjan con la función de densidad usando algoritmos de knn , y, por otro lado, que normalmente las distribuciones de estimaciones no tienen patrones claros presentando numerosas discontinuidades.

Es por ello, que se suele realizar una asignación de probabilidades a las estimaciones realizadas por aproximación posterior dependiendo de la distribución de las predicciones realizadas $-P(c_i|x)$, es decir de la predicciones realizadas de que un caso x haya sido asignado a la clase c_i .

⁴ Un hipercubo n -es la generalización del cubo dentro de n dimensiones, es decir desplazado o proyectado en las n dimensiones que tenga el espacio en el que se desarrolla.



Atendiendo a la expresión anterior de la función de densidad

$$p(x) \approx \frac{k/n}{V}$$

la probabilidad de que un caso de estimación x haya sido asignado a la clase c_i sería

$$p(c_i, x) \approx \frac{k_i/n}{V}$$

donde k_i sería el número de secuencias o número de vecinos entre los más próximos k seleccionados asignados a la clase c_i .

Usando las expresiones de la probabilidad condicionada bayesiana obtenemos entonces:

$$p(c_i|x) = \frac{p(x, c_i)}{p(x)} = \frac{p(x, c_i)}{\sum_{j=1}^m p(x, c_j)} \approx \frac{k_i/n}{V \sum_{j=1}^m \frac{k_j/n}{V}} = \frac{k_i}{\sum_{j=1}^m k_j} = \frac{k_i}{k}$$

Es decir, la probabilidad de la asignación de un caso x a una clase c_i viene determinada por la relación o cociente entre el número de vecinos más próximos asignados a dicha clase c_i (k_i) y el número total de vecinos utilizados en la estimación (k).

V.1.6.5. Consideraciones finales

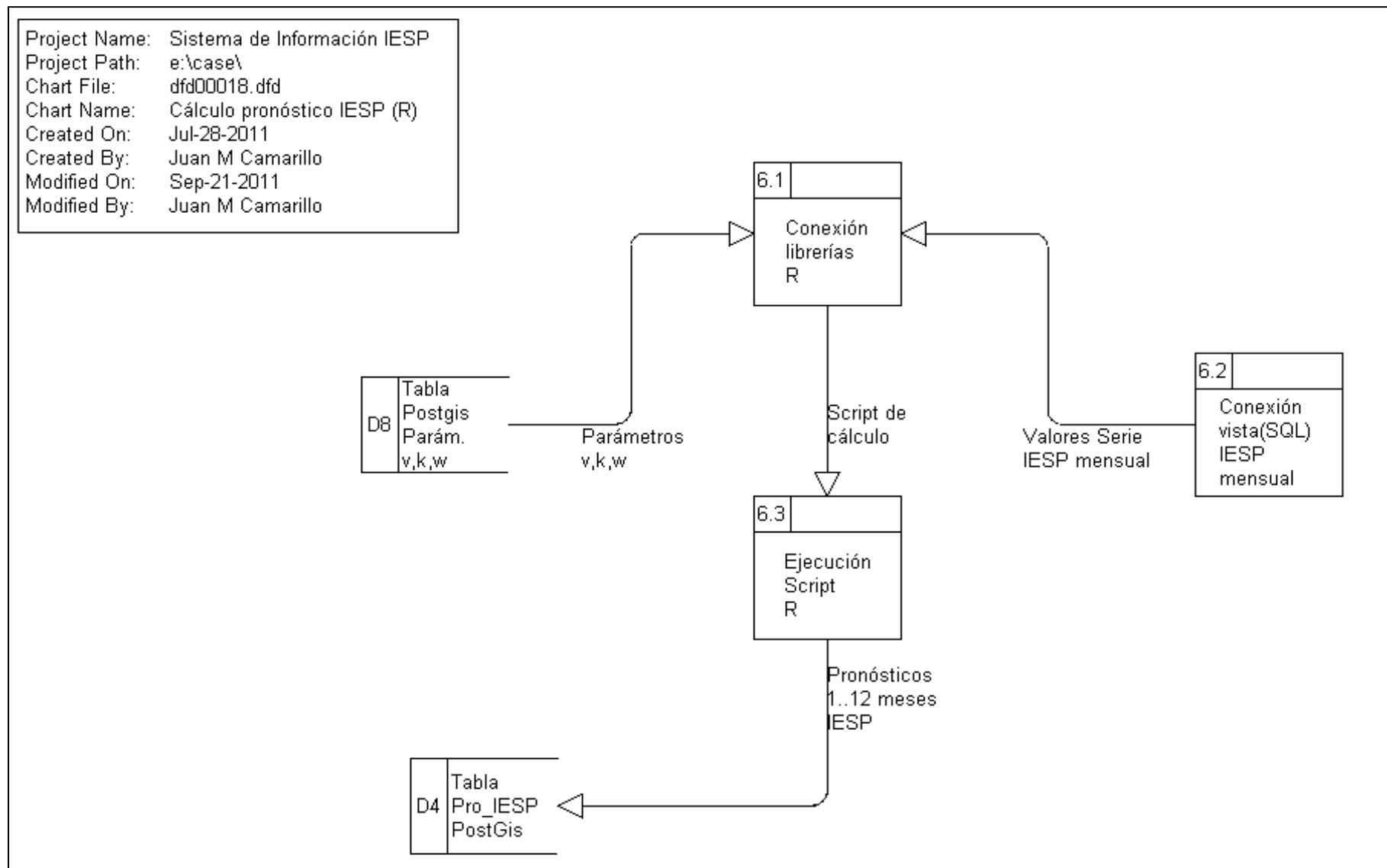
El método de clasificación supervisada knn es un método no paramétrico de asignación de estimaciones predictoras largamente utilizado para series de datos sobre los que existe un alto grado de incertidumbre en la asignación de los modelos de distribución de las mismas.

Junto a ello, existen ciertas cuestiones importantes a tener en cuenta, entre las que cabe destacar la alta sensibilidad al número de vecinos próximos seleccionados ya

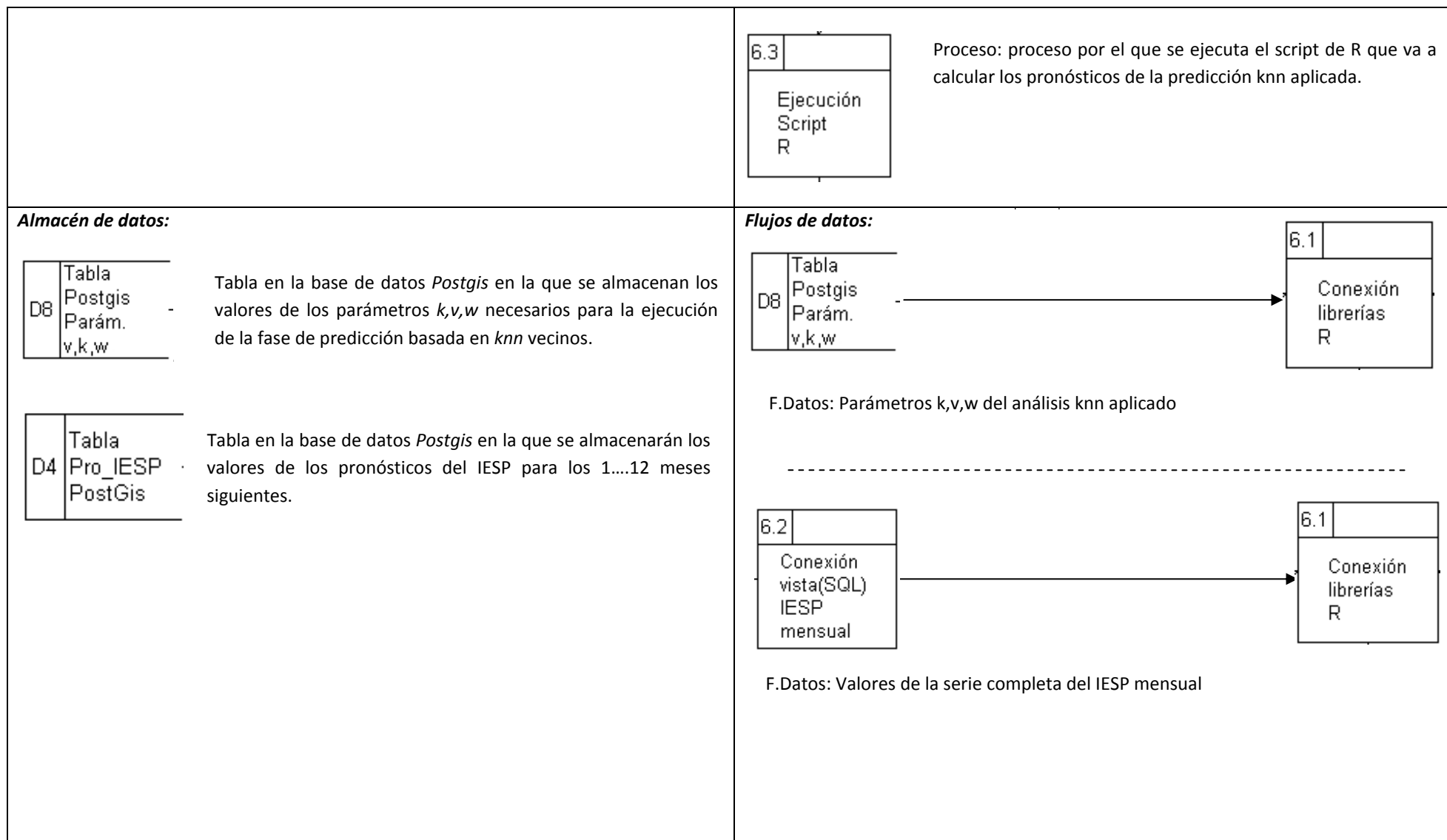
que un bajo número de vecinos puede generar un gran ruido en las regiones de frontera, mientras que un alto número de estos podría producir una suavización importante en las estimaciones en dichas regiones. Para ello, se ha utilizado una función de verosimilitud en que minimiza las distancias entre las estimaciones tomando los ternas de parámetros necesarios (v,k,w) para su aplicación.

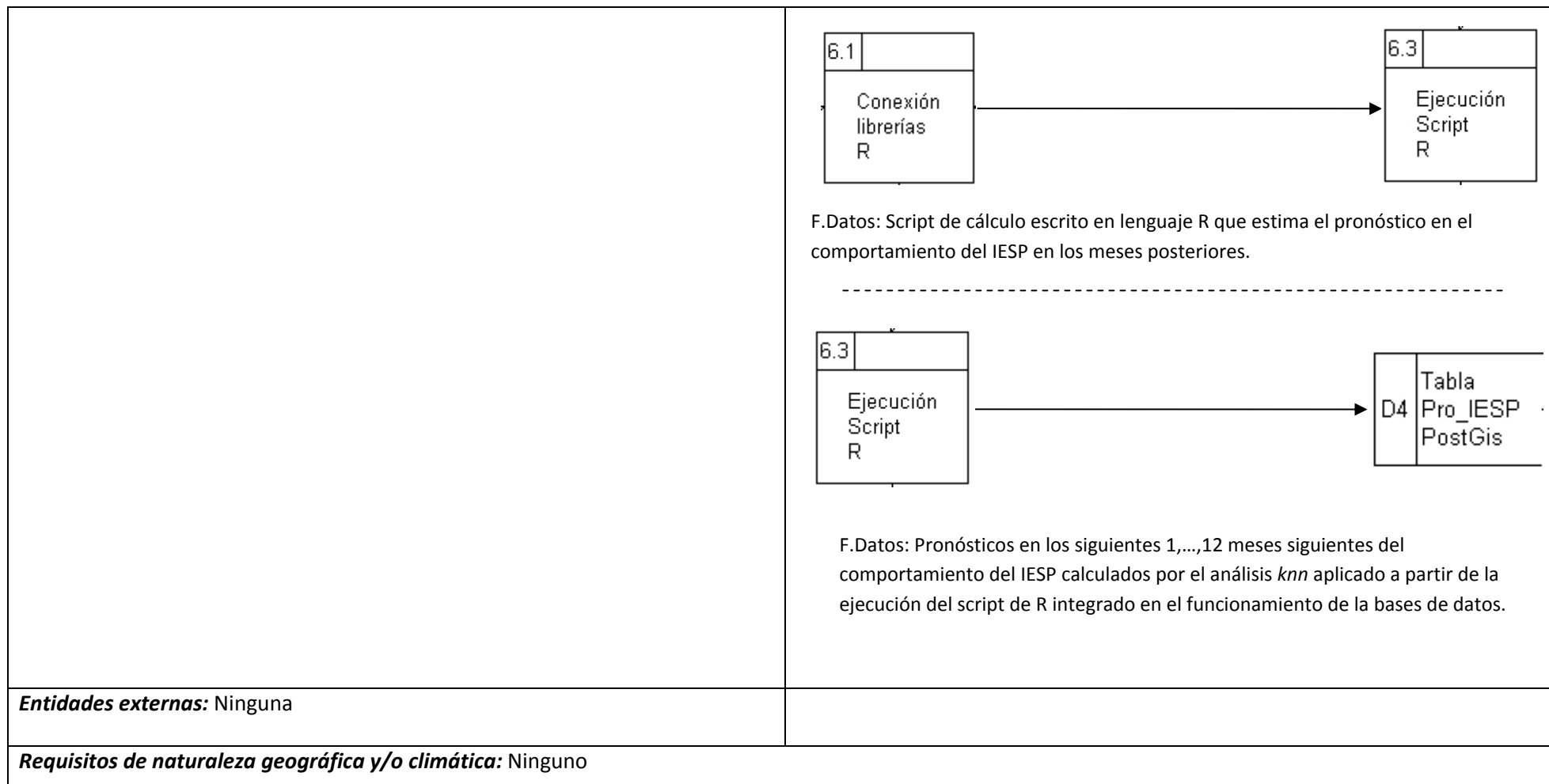
Se trata también de un método que, al estar basado en una fase de aprendizaje, permite la posibilidad de mejoras en las estimaciones al ir contando, progresivamente, con un mayor número de secuencias comparadas al mismo tiempo que permite revisar los valores óptimos de los parámetros anteriores.

IV.1.7. Diagrama del proceso “Cálculo del pronóstico del IESP” (Figura IV.21)

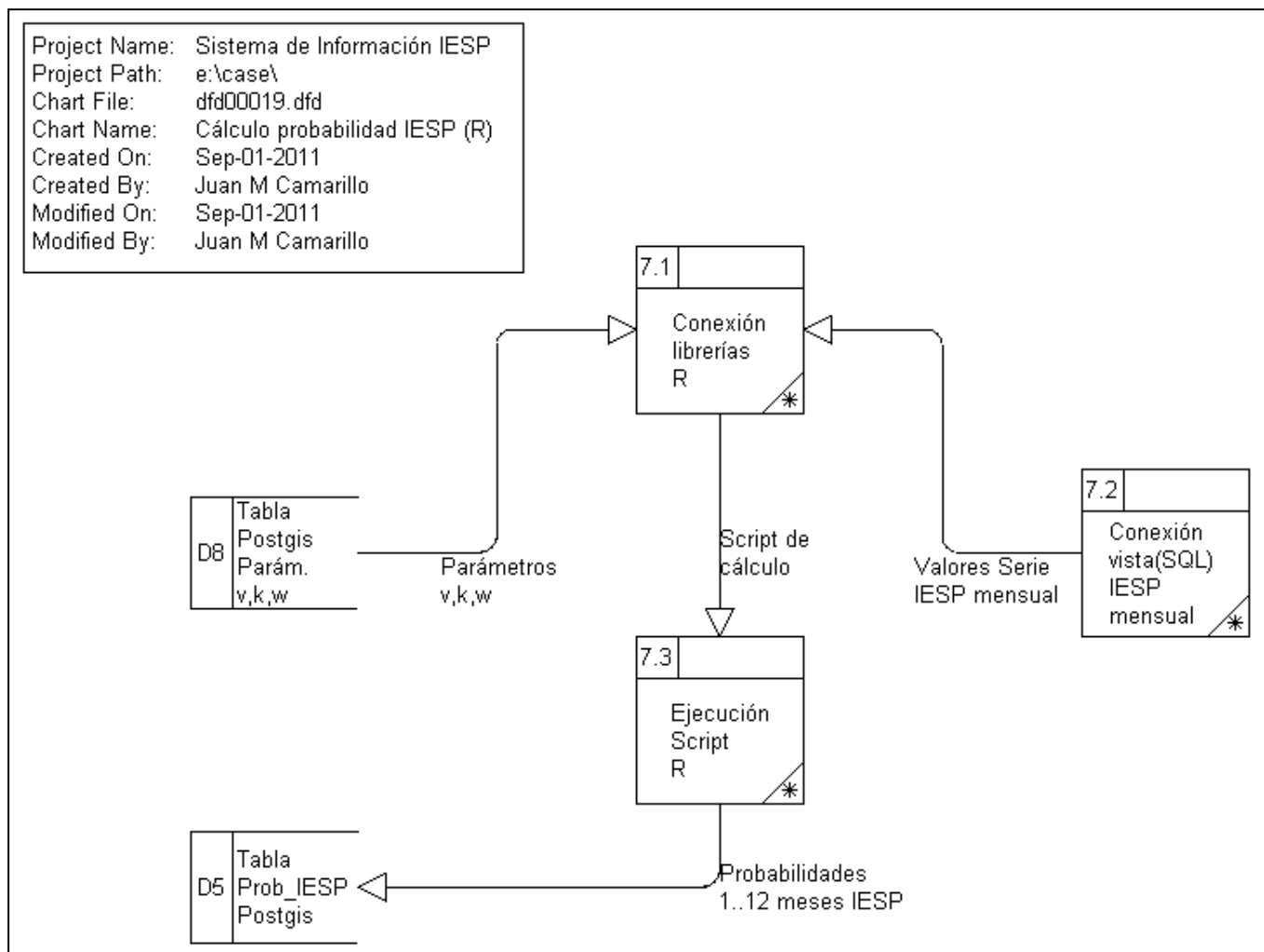


DFD nº 00018 Diagrama de Flujo de Datos del Proceso para el cálculo de los pronósticos a 12 meses del IESP. Nivel 2									
<p>Objetivos: El proceso descrito a continuación tiene como objetivo integrar en el sistema el procedimiento de cálculo de los pronósticos del comportamiento del Indicador según la metodología propuesta. El proceso está diseñado como Script del software R y precisa de una conexión a la base de datos PostGis a partir de la integración en la misma de las librerías de dicho software.</p> <p>Este esquema funcional permitirá incorporar como datos de entrada de dicho Script las vistas (SQL) que proporcional el cálculo de la serie histórica del IESP.</p>	<p>Datos de partida: Los datos de partida disponibles son las tablas que para cada estación de referencia almacena los valores de los parámetros k,v,w. Junto a ello, las vistas (SQL) necesarias para el cómputo de los pronósticos.</p>								
<p>Requisitos funcionales:</p> <ul style="list-style-type: none"> - Conexión con las librerías de R para su integración con la base de datos POstgis. - Definición de las vistas (SQL) cuya salida es la serie completa del valor de IESP. - Ejecución del script de cálculo. 	<p>Procesos de datos:</p> <div style="display: flex; align-items: flex-start; margin-bottom: 20px;"> <div style="border: 1px solid black; padding: 5px; margin-right: 10px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30px; text-align: center;">6.1</td> <td style="width: 100px;"></td> </tr> <tr> <td colspan="2" style="text-align: center;">Conexión librerías R</td> </tr> </table> </div> <div style="flex-grow: 1;"> <p>Proceso: integración en la base de datos Postgis de las librerías del programa de análisis matemático R que deberán ser llamadas para el cálculo de los pronósticos del IESP.</p> </div> </div> <div style="display: flex; align-items: flex-start;"> <div style="border: 1px solid black; padding: 5px; margin-right: 10px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30px; text-align: center;">6.2</td> <td style="width: 100px;"></td> </tr> <tr> <td colspan="2" style="text-align: center;">Conexión vista(SQL) IESP mensual</td> </tr> </table> </div> <div style="flex-grow: 1;"> <p>Proceso: Definición de las sentencias SQL dentro del script de R para que funcionen a modo de tablas de entrada sobre las que se realizará la fase de predicción de la metodología knn aplicada.</p> </div> </div>	6.1		Conexión librerías R		6.2		Conexión vista(SQL) IESP mensual	
6.1									
Conexión librerías R									
6.2									
Conexión vista(SQL) IESP mensual									



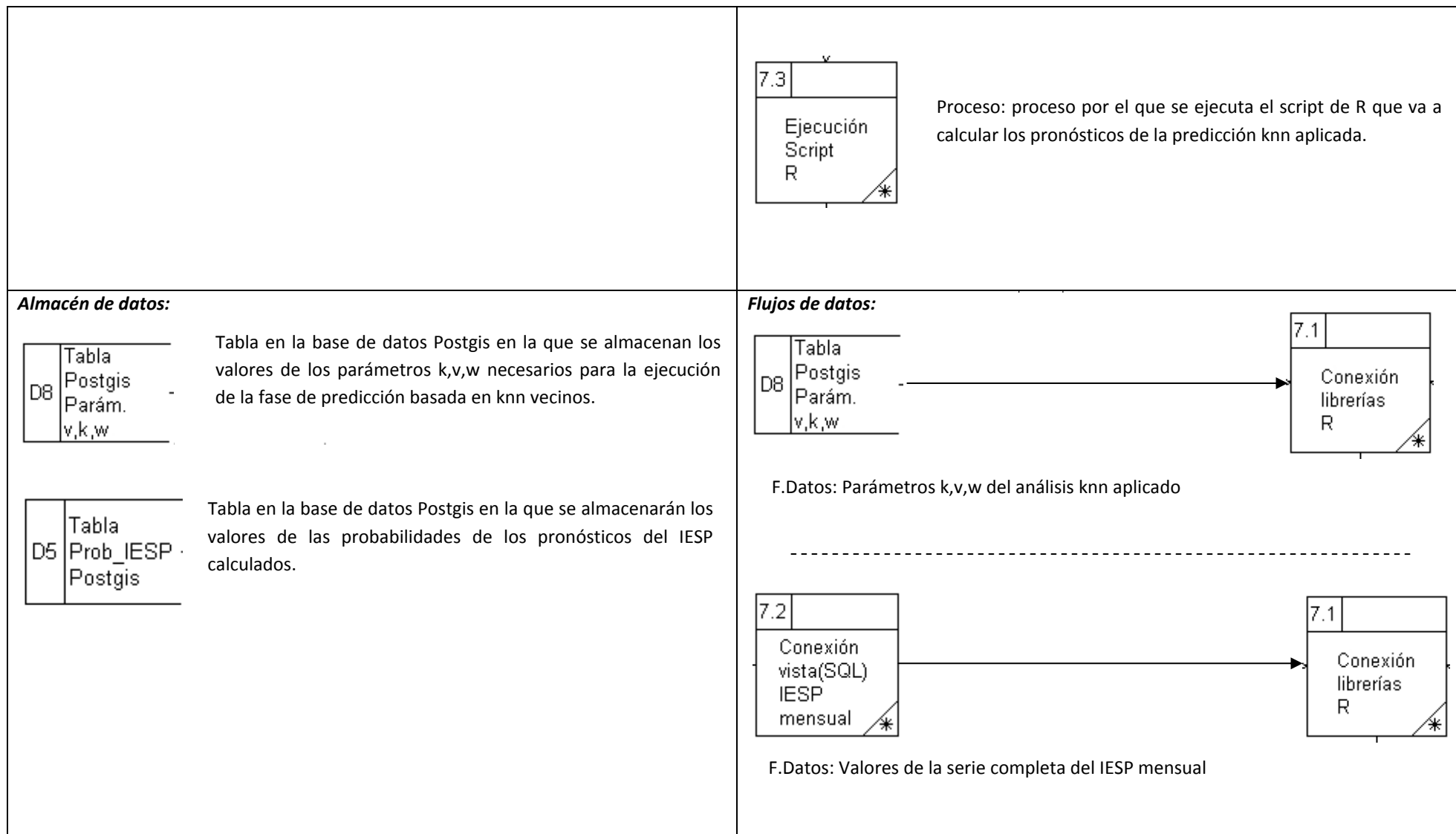


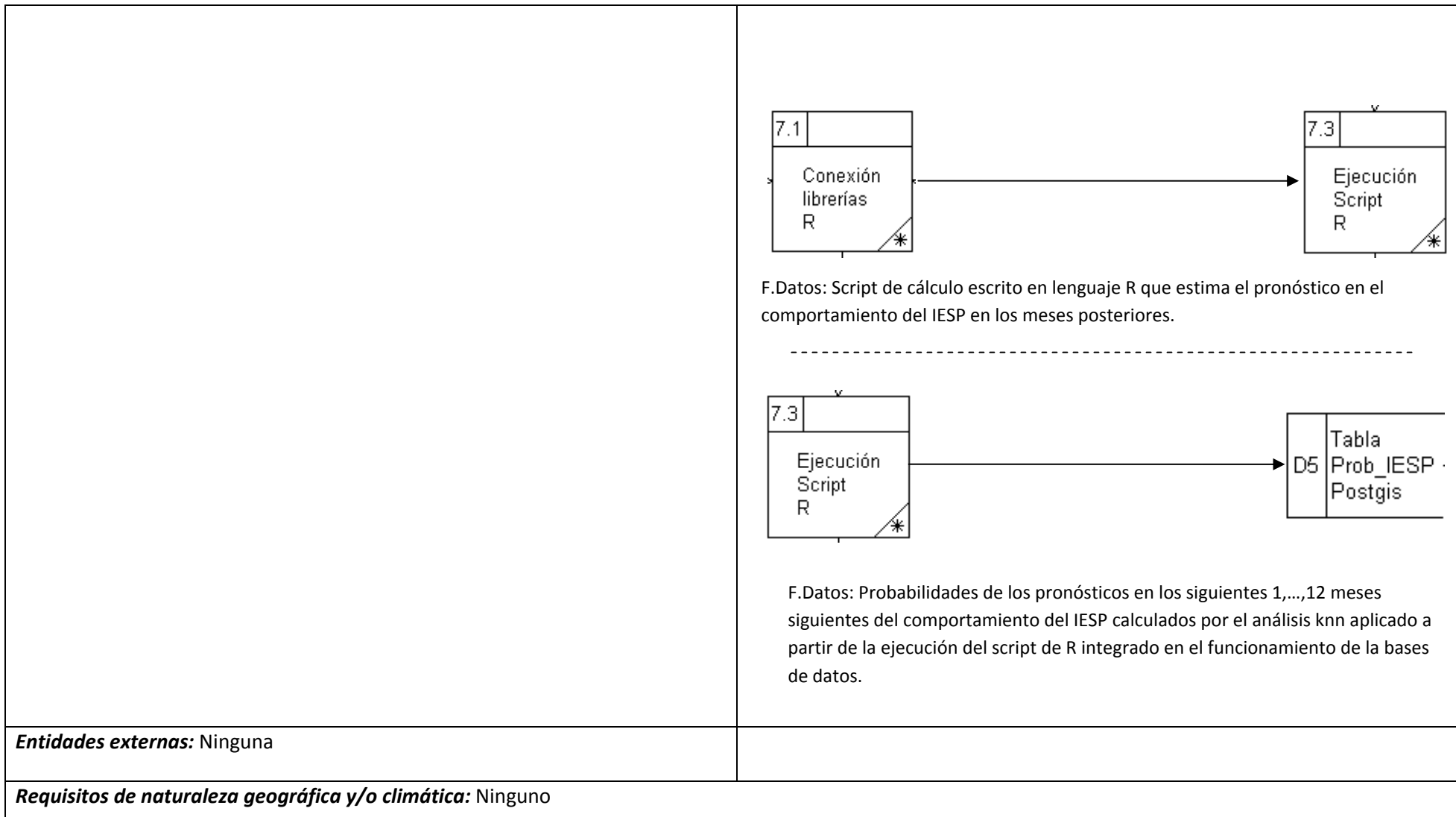
IV.1.8. Diagrama del proceso “Cálculo de la probabilidad del pronóstico del IESP” (Figura IV.22)



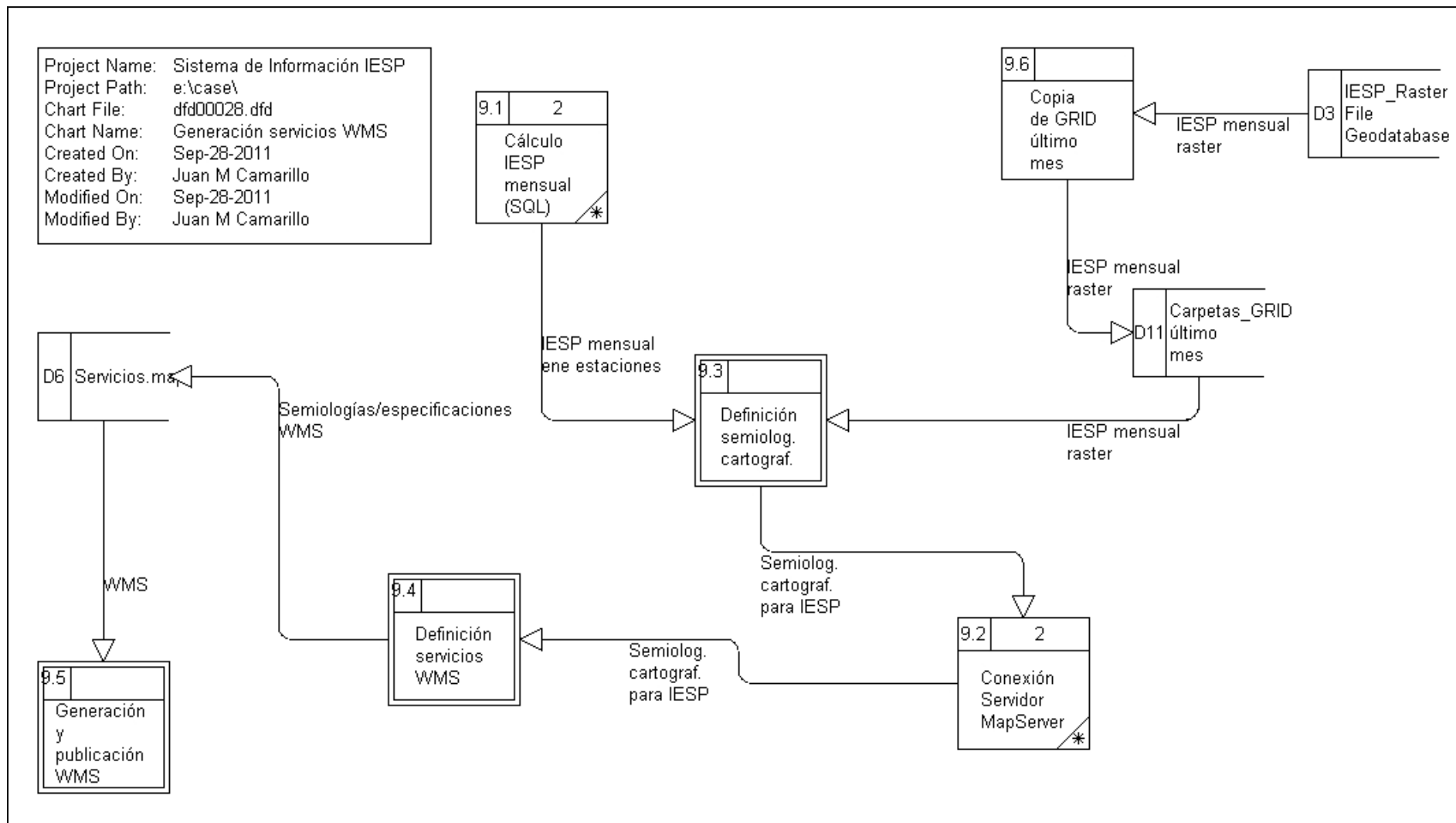
Fuente: elaboración propia

<u>DFD nº 00019 Diagrama de Flujo de Datos del Proceso para el cálculo de las probabilidades de los pronósticos a 12 meses del IESP. Nivel 2</u>									
<p>Objetivos: El proceso descrito a continuación tiene como objetivo integrar en el sistema el procedimiento de cálculo de las probabilidades asociadas a los pronósticos del comportamiento del Indicador según la metodología propuesta. El proceso está diseñado como Script del software R y precisa de una conexión a la base de datos PostGis a partir de la integración en la misma de las librerías de dicho software.</p> <p>Este esquema funcional permitirá incorporar como datos de entrada de dicho Script las vistas (SQL) que proporcional el cálculo de la serie histórica del IESP.</p>	<p>Datos de partida: Los datos de partida disponibles son las tablas que para cada estación de referencia almacena los valores de los parámetros k,v,w. Junto a ello, las vistas (SQL) necesarias para el cómputo de los pronósticos.</p>								
<p>Requisitos funcionales:</p> <ul style="list-style-type: none"> - Conexión con las librerías de R para su integración con la base de datos POstgis. - Definición de las vistas (SQL) cuya salida es la serie completa del valor de IESP. - Ejecución del script de cálculo. 	<p>Procesos de datos:</p> <div style="display: flex; align-items: flex-start; margin-bottom: 20px;"> <div style="border: 1px solid black; padding: 5px; margin-right: 10px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30px; text-align: center;">7.1</td> <td style="width: 100px;"></td> </tr> <tr> <td colspan="2" style="text-align: center;">Conexión librerías R</td> </tr> </table> </div> <div style="flex-grow: 1;"> <p>Proceso: integración en la base de datos Postgis de las librerías del programa de análisis matemático R que deberán ser llamadas para el cálculo de los pronósticos del IESP.</p> </div> </div> <div style="display: flex; align-items: flex-start;"> <div style="border: 1px solid black; padding: 5px; margin-right: 10px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30px; text-align: center;">7.2</td> <td style="width: 100px;"></td> </tr> <tr> <td colspan="2" style="text-align: center;">Conexión vista(SQL) IESP mensual</td> </tr> </table> </div> <div style="flex-grow: 1;"> <p>Proceso: Definición de las sentencias SQL dentro del script de R para que funcionen a modo de tablas de entrada sobre las que se realizará la fase de predicción de la metodología knn aplicada.</p> </div> </div>	7.1		Conexión librerías R		7.2		Conexión vista(SQL) IESP mensual	
7.1									
Conexión librerías R									
7.2									
Conexión vista(SQL) IESP mensual									





IV.1.9. Diagrama del proceso “Generación de servicios WMS” (Figura IV.23)



DFD nº 000028 Diagrama de Flujo de Datos del Proceso para la generación de los servicios WMS. Nivel 2

Objetivos: El proceso descrito a continuación tiene como objetivo especificar el flujo de datos y los procesos que deben funcionar en la configuración y generación de los servicios WMS que mostrarán los valores tanto en estaciones como raster del indicador IESP.

El proceso completo se basa en la definición de las semiologías para expresar el índice, en la configuración formal de dichas semiologías en el archivo *.map de configuración del servidor de mapas MapServer y en la definición y configuración de las características del servicio interoperable WMS producido sobre las que se basarán las capacidades del servicio –ie.get feature info-.

Datos de partida: Los datos de partida disponibles son la vista (SQL) del indicador IESP del último mes de cálculo y la superficie raster del mismo. La primera es una consulta SQL en la base de datos geográficas asociada a una entidad con geometría: esta_sel_capa (estaciones meteorológicas de referencia) y la segunda es una superficie raster almacenada en una File Geodatabase de ArcGis.

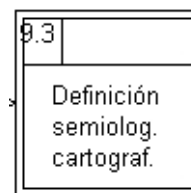
Requisitos funcionales:

- La vista del IESP en estaciones debe relacionar el valor del indicador y las estaciones meteorológicas. Esta última debe ser una tabla con geometría (*well known bynary*) en PostGis.
- Conexión del servidor de mapas MapServer a la base de datos PostGis y a la carpeta de sistema GRID_ultimo_mes
- Servidor de mapas –MapServer-
- Servidor de aplicaciones –Apache-

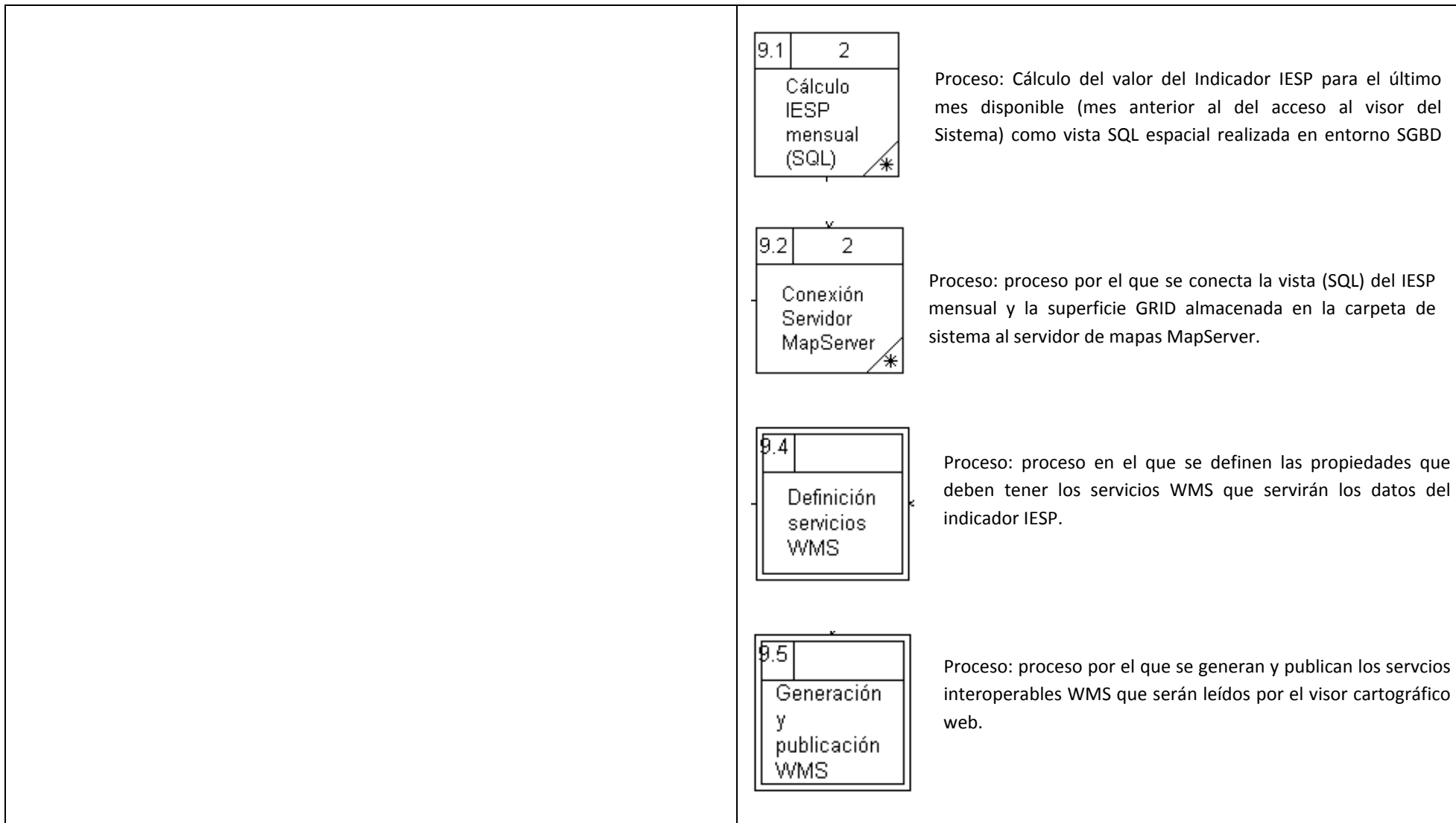
Procesos de datos:

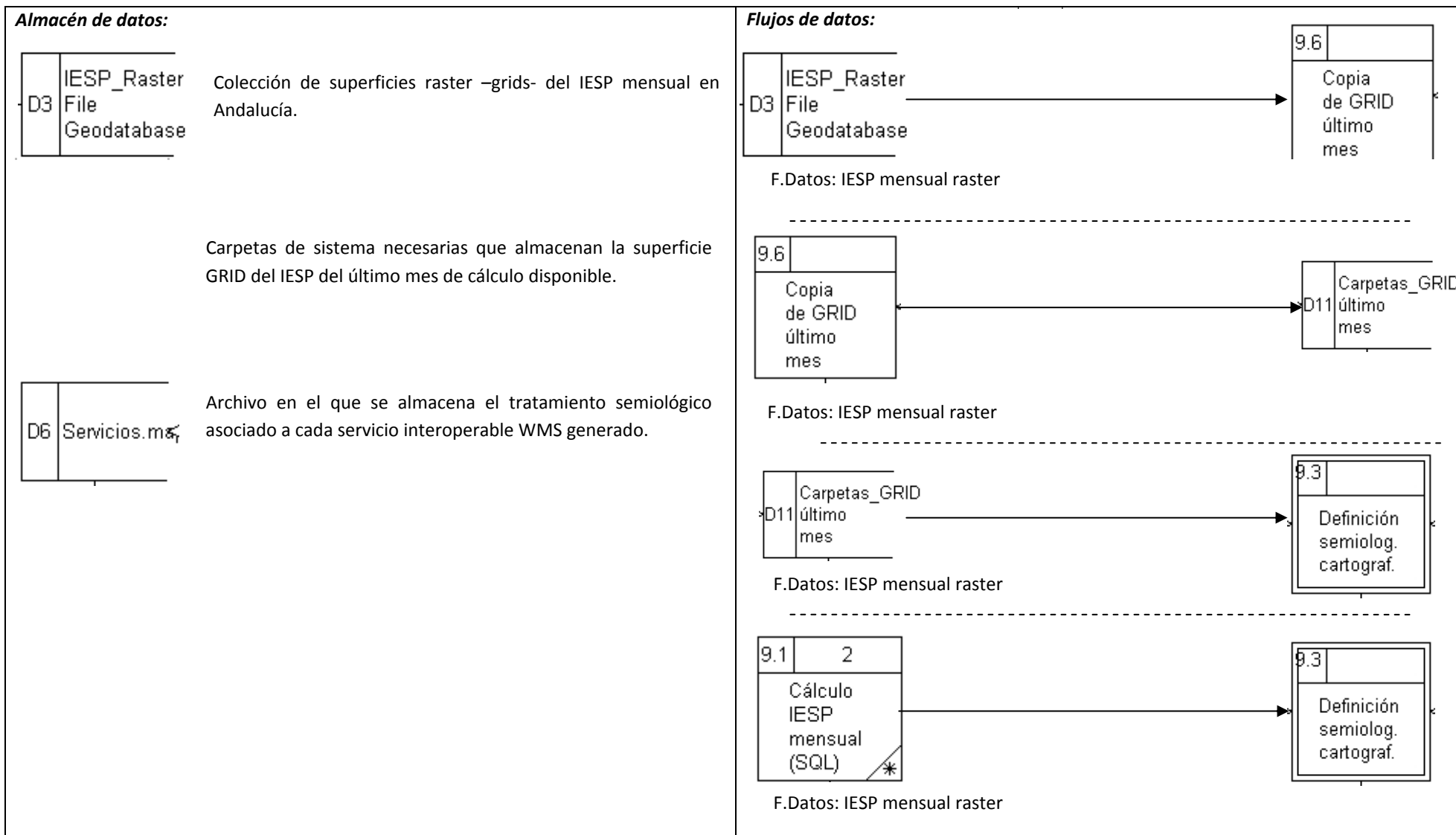


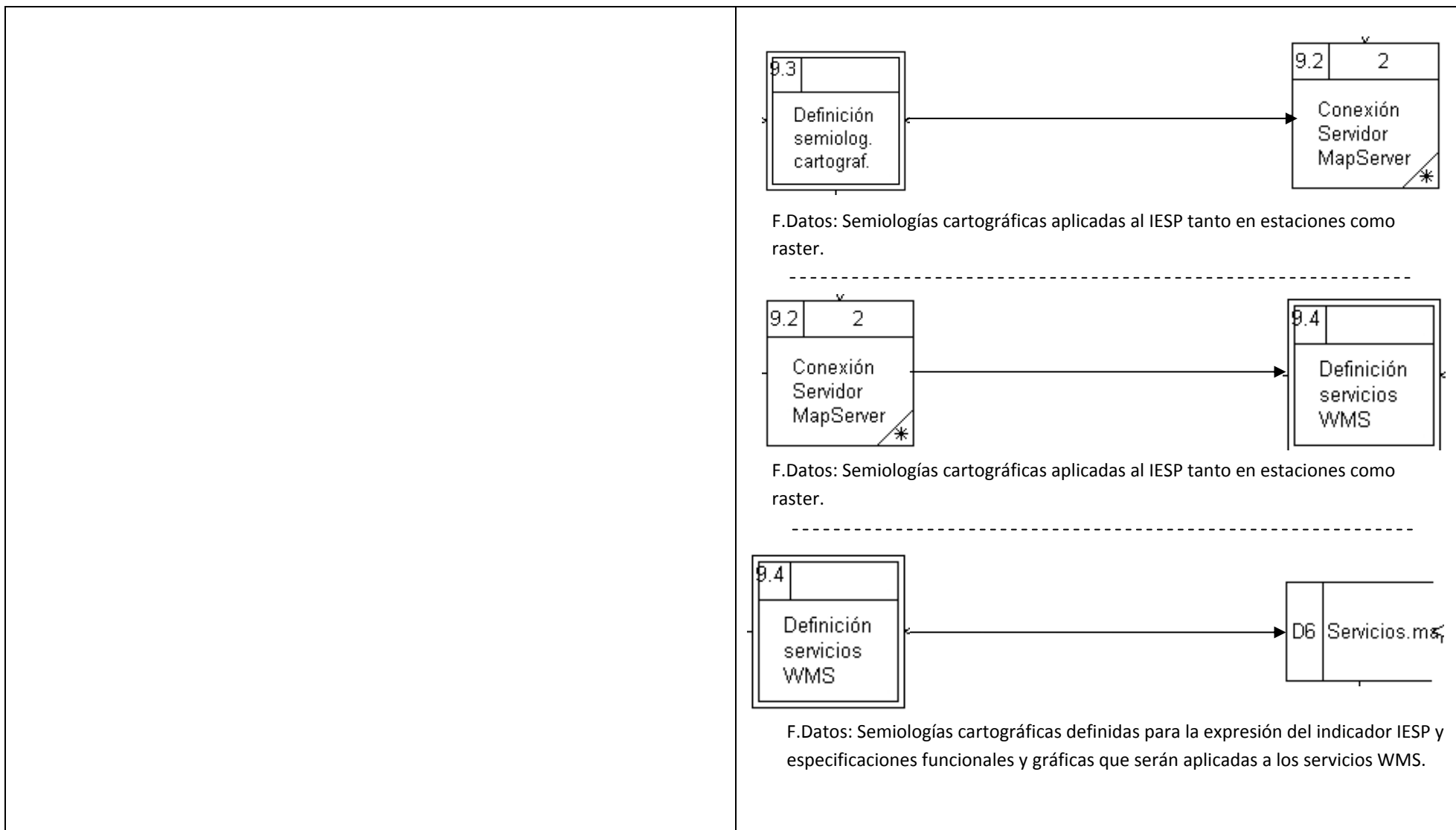
Proceso: copia de las superficies raster del valor del indicador IESP en una carpeta de sistema en la que se almacenan las superficies GRID del último mes de cálculo disponible para poder ser leídas por MapServer.

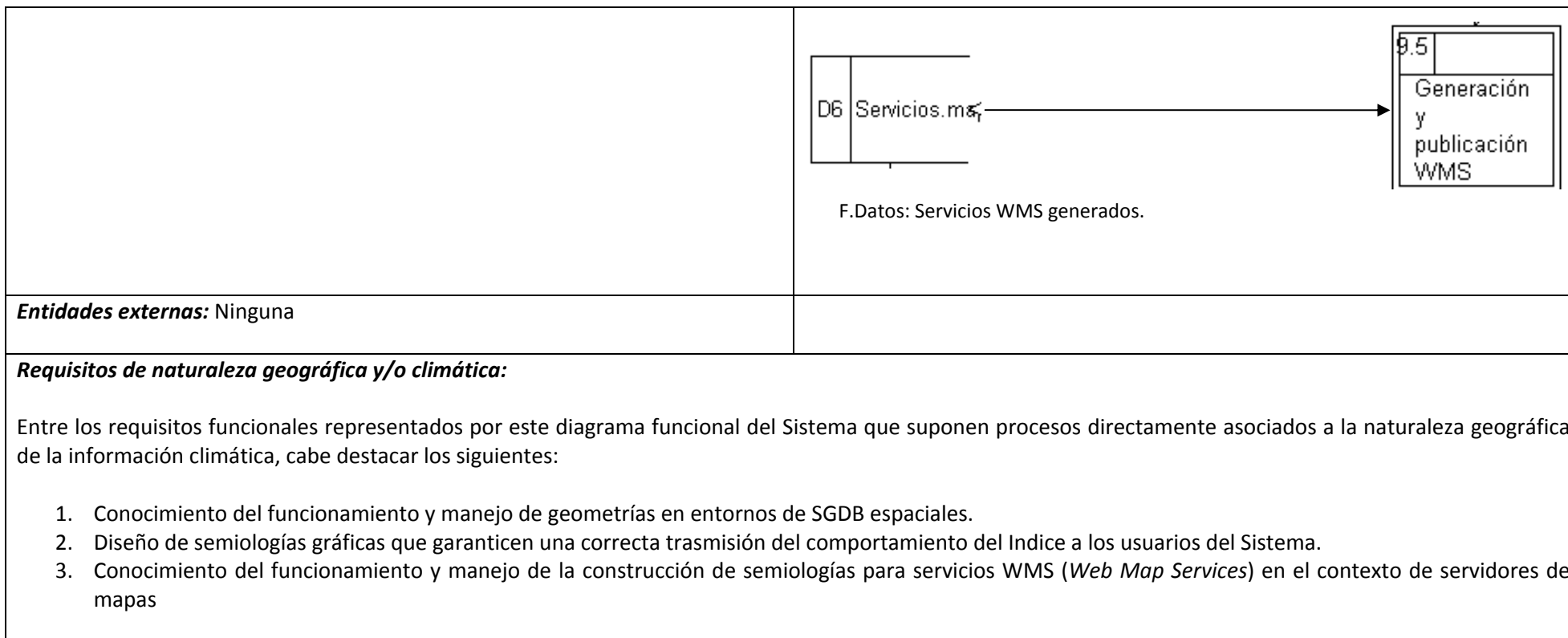


Proceso: Definición de las semiologías cartográficas que serán aplicadas al indicador.









IV.2. Modelado lógico de datos (Logical Data Modelling –LDM) o diseño del modelo Entidad-Relación

El modelado lógico de de datos hace referencia al modelo de datos entidad-relación (E-R) que supone la estructura lógica en la que se van a almacenar todos los datos necesarios para el cálculo del IESP mensual dentro del sistema propuesto.

El modelo entidad-relación es el modelo o estructura organizativa de los datos más utilizado en el presente en todas las esferas que impliquen gestión dinámica de datos. Es decir, en la actualidad, en la mayoría de las aplicaciones de gestión que utilizan sistemas de gestión de bases de datos, es el principal modelo organizativo de la información aplicado, fundamentalmente debido a su gran versatilidad, potencia analítica implícita, formulaciones matemáticas de base que lo sustentan –teoría de conjuntos, lógica de predicados y algebra relacional-, madurez tecnológica y amplia disponibilidad de software (Xia et al 2002).

IV.2.1. Introducción conceptual al modelo relacional

Los principios del modelo de datos relacional fueron establecidos por E.F.Codd, empleado del gigante *IBM*, a finales de los años 60⁶ y fue definido como modelo entidad relación (E-R) especialmente orientado al diseño del primero por Chen en 1976⁷. Hoy día, especialmente a partir de la década de los 80 cuando comienzan a comercializarse los primeros SGDB con rendimientos aceptables, se trata del modelo de organización, estructuración, almacenamiento y gestión de datos más utilizados en el ámbito de los sistemas de información de cualquier orden.

⁶ “A relational model for large shared data banks”; publicado en Communications of the ACM.

⁷ “The entity-relationship model—toward a unified view of data” publicado en ACM Transactions on Database Systems (TODS) - Special issue: papers from the international conference on very large data bases

El principal objetivo del modelo de datos relacional es facilitar que la base de datos sea percibida o vista por el usuario como una estructura lógica que consiste en un conjunto de entidades, relaciones o tablas y de interrelaciones entre los registros de las mismas y no como una estructura física de implementación. Esto ayuda a conseguir un alto grado de independencia de los datos. Es decir, la forma de almacenar los datos no debe influir en su manipulación lógica.

Un objetivo adicional del modelo es conseguir que esta estructura lógica con la que se percibe la base de datos sea simple y uniforme. Con el fin de proporcionar simplicidad y uniformidad, toda la información se representa de una única manera: mediante valores explícitos que contienen las entidades o tablas, superando de esta forma la fase anterior en la que las relaciones entre los datos se incorporaban apuntando físicamente a la dirección en el disco de los registros. Con el mismo propósito, todos los valores de datos se consideran atómicos; es decir, no es posible descomponerlos.

Además, el modelo debe garantizar la independencia lógica, de tal manera que las aplicaciones que tienen como datos de entrada el modelo utilizado no necesiten ser modificadas cuando cambian los elementos de la base de datos.

Debe garantizar también un alto grado de flexibilidad y versatilidad en términos de explotación de la información contenida en la base de datos a partir de la definición de vistas (consultas) independientes de los datos físicos propiamente dichos.

La terminología básica implicada en la formulación y desarrollo de los modelos relacionales y que, por tanto, serán manejadas en el desarrollo del epígrafe es la siguiente:

IV.2.1.1. Relación

Es la estructura básica del modelo relacional. Con una relación es posible representar tanto instancias de una entidad del universo real como interrelaciones

entre entidades de distinto tipo. Es capaz de recoger interrelaciones de cardinalidad múltiple. Su representación informal es una tabla. También se denomina **entidad** en el modelo E_R, que será la nomenclatura utilizada en este trabajo, especialmente para poder referirnos a las interrelaciones entre los registros de las tablas en las que se materializan las entidades simplemente como relaciones.

La entidad puede definirse como cualquier objeto (real o abstracto) que existe en la realidad y acerca del cual queremos almacenar información en la base de datos (Chen, 1976).

IV.2.1.2 Dominio

Es el conjunto válido de valores de referencia para definir propiedades o atributos. Un dominio es un conjunto nominado y homogéneo de valores. Cada dominio puede definirse de dos maneras:

- Por **extensión** (enumeración de sus elementos): días de la semana = {lunes, martes, miércoles, jueves, viernes, sábado, domingo}
- Por **intensión** (mediante una propiedad que recoja el recorrido de sus valores admisibles): precipitación mensual = $0 \leq \text{pmm} \leq 2000$

IV.2.1.3. Atributo

Representa las propiedades de la entidad o relación. Un atributo necesariamente ha de definirse sobre un dominio. Su representación informal es una columna.

Un atributo (A) es la interpretación de un determinado dominio en una entidad.

$D = \text{Dom}(A) \Rightarrow D$ es el dominio de A

Un atributo está siempre asociado a una entidad, mientras que un dominio tiene existencia propia con independencia de las entidades que existan en el modelo.

Representa por tanto una propiedad de una entidad tomando los valores de un dominio establecido. Otra característica importante a destacar es que varios atributos distintos (de la misma o de diferentes entidades) pueden tomar sus valores del mismo dominio.

IV.2.1.4. Esquema de la entidad

Consiste en un nombre de entidad o relación R y un conjunto de atributos $\{A_1, A_2, \dots, A_n\}$.

IV.2.1.4. Tupla: es una ocurrencia o instancia dentro de una relación o entidad. Una tupla permite referenciar una instancia de una entidad en el universo o la interrelación específica o concreta entre instancias de entidades. Su representación informal es una fila. Por tanto, una entidad o relación tiene un **“conjunto”** de tuplas.

De esta forma, cada entidad viene definida por un esquema de tuplas entendido como conjunto de pares de la forma:

$\tau = \{(A_1, D_1), (A_2, D_2), \dots, (A_n, D_n)\}$ donde

$\{A_1, A_2, \dots, A_n\}$ ($n > 0$) es el conjunto de nombres de atributos y

D_1, D_2, \dots, D_n son los dominios asociados.

Por tanto, una tupla t de esquema de tupla $\tau = \{(A_1, D_1), (A_2, D_2), \dots, (A_n, D_n)\}$ sería un conjunto de pares de la forma:

$t = \{(A_1, v_1), (A_2, v_2), \dots, (A_n, v_n)\}$ tal que $v_i \in D_i$

o lo que es lo mismo, los valores (v) adoptados por los diferentes atributos de la entidad correspondiente son de tal forma que para cualquier valor de un atributo se cumple que este pertenece al dominio (D) definido para el mismo.

En una relación o entidad una tupla no puede repetirse ya que un conjunto no tiene por definición elementos repetidos.

IV.2.1.5. Grado de la entidad

Es el número de atributos de su esquema

IV.2.1.6. Cardinalidad de la entidad

Es el número de tuplas que la forman

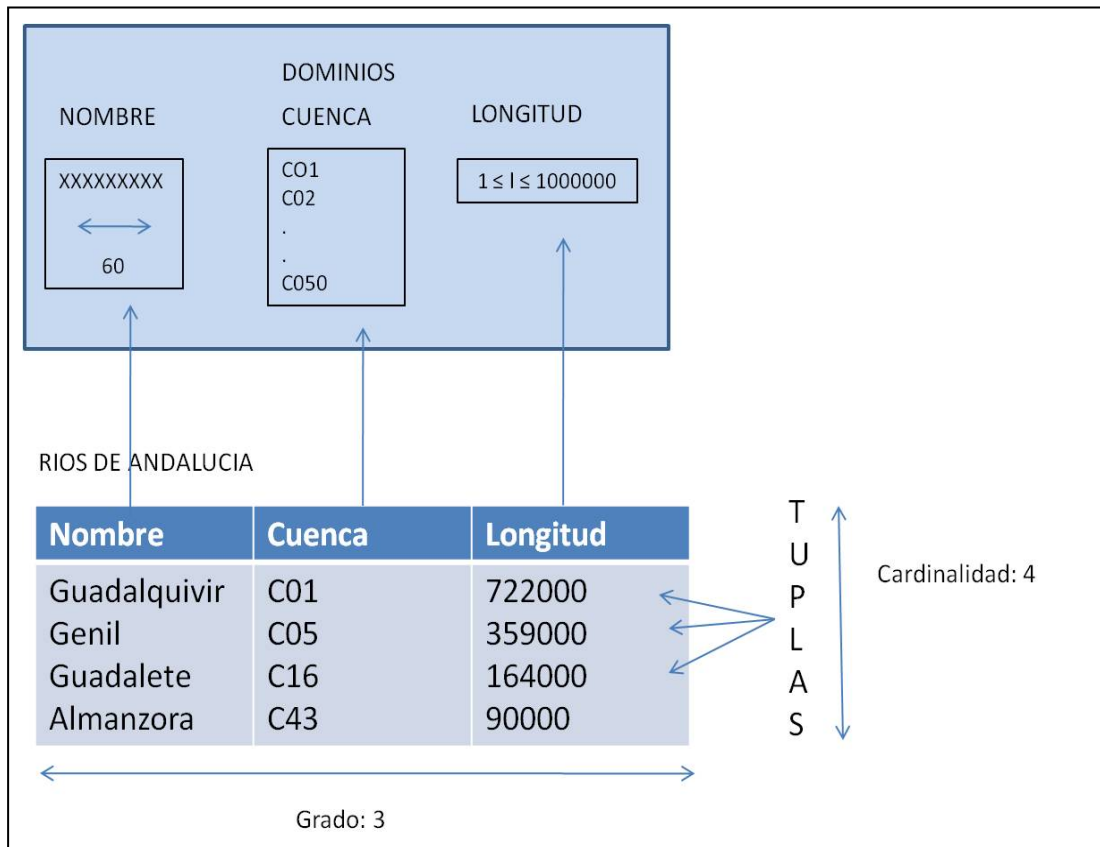


Figura IV.24. Ejemplo de esquema de entidad

Fuente: elaboración

IV.2.1.7. Interrelación (relación en el modelo E-R)

Se entiende por interrelación *–relationship–* la asociación, vinculación o correspondencia entre entidades (Chen, 1977).

Estas relaciones entre entidades se establecen entre las tuplas de las mismas.

IV.2.1.8. Cardinalidad de las interrelaciones

Hace referencia a la correspondencia existente entre las tuplas de las entidades interrelacionadas. Las entidades en el mundo real no son siempre independientes. A menudo hay interrelaciones entre las entidades: los ríos y sus embalses, las estaciones meteorológicas y los datos que registran, etc... Las diferentes cardinalidades posibles son las siguientes:

- n exactamente n
- 1 uno y solo uno
- 0 —————→ 1: de cero a 1
- 0 —————→ *: de cero a muchos
- 1 —————→ *: de uno a muchos
- m —————→ n: de muchos a muchos

IV.2.1.9. Rol

Se denomina rol al papel o función que juegan las tuplas existentes en el extremo de una relación. Cada relación posee dos roles, uno en cada extremo, describiendo estos como una entidad ve a la otra a través de la relación.

IV.2.1.10. Integridad relacional

Con este nombre se designan aquellas reglas que han de ser aplicadas a una base de datos para asegurar que los datos introducidos sean consistentes con la realidad que

pretenden modelar. Existen dos reglas *generales* que aporta el modelo relacional que son las siguientes:

- Regla de integridad de las entidades: todas las tuplas de una entidad deben estar referenciadas unívocamente; o lo que es lo mismo, se debe garantizar que dichas tuplas presenten un valor unívoco en alguno de los atributos de su estructura o, en su defecto, que la composición del conjunto de valores que toma cada tupla en el conjunto de atributos de la entidad sea unívoco y por lo tanto distinto a la de cualquier otra tupla de esta.

Cualquiera de los dos casos materializa lo que se conoce como clave primaria de la entidad.

- Regla de integridad referencial: las interrelaciones entre las tuplas de dos entidades son implícitas y se concretan en la correspondencia entre los valores del atributo con valores unívocos para cada tupla presente en ambas entidades.

IV.2.1.11. Claves primarias (*primary key*) y claves foráneas (*foreign key*)

Como hemos visto anteriormente, una relación o entidad necesita un identificador único para cada una de las tuplas, ya que estas son irrepetibles en el conjunto que representa la entidad. Esta es la clave primaria de la entidad, que se define como un subconjunto C de los atributos de R , cuyos valores no pueden ser repetidos. Una clave primaria debe ser *mínima*, en el sentido de que en su composición no intervengan más que los atributos estrictamente requeridos para identificar las tuplas de forma única. Puesto que una relación es un conjunto de tuplas, se debe dar la condición de que toda relación deba tener una clave primaria; al menos el conjunto de los atributos de una relación conforma la clave de esa relación. Además, una clave primaria puede ser simple (formada por un solo atributo) o compuesta (formada por más de uno).

Entendemos por clave foránea el atributo o conjunto de los atributos de una entidad que pertenece al mismo dominio D que el atributo o conjunto de atributos que conforman la clave primaria de otra entidad presente en la base de datos y cuya correspondencia con esta garantiza la *integridad referencial* anteriormente citada.

Por definición los valores del o de los atributos que conforman la clave primaria de una entidad no pueden presentar valores nulos para garantizar la regla de *integridad de entidades* anterior.

IV.2.1.12. Vistas o consultas

Una vista es una tabla virtual en una base de datos cuyos contenidos están definidos por una consulta realizada a través de algún lenguaje específicamente orientado a la consulta de tablas sobre un modelo de datos. Para los usuarios la vista aparece con la apariencia de una tabla constituida por sus correspondientes filas y columnas. Sin embargo la vista no existe físicamente como un conjunto almacenado de valores; es decir, los datos y columnas visibles en la vista son resultado de la consulta que define la vista. El lenguaje estándar e internacional de consulta a las tablas presentes en un modelo relacional es el lenguaje SQL.

IV.1.13. Lenguaje SQL (Structured Query Language)

Conceptualmente SQL es un lenguaje de definición de datos (*LDD*), un lenguaje de definiciones de vistas (*LDV*) y un lenguaje de manipulación de datos (*LMD*), que posee también capacidad para especificar restricciones y evolución de esquemas y a través del cual podemos definir el conjunto de vistas de los datos almacenados en el modelo necesarias para el funcionamiento del sistema de información.

IV.2.1.14. Diseño conceptual del modelo entidad-relación (E-R)

El modelo E-R (Chen, 1976) fue propuesto para crear esquemas conceptuales del flujo de datos dentro de un sistema de información. Se trata por tanto del diseño o construcción de un esquema que modela la información implicada en el sistema de información que se esté diseñando, así como el flujo de la misma necesario para la producción y resolución de toda la información derivada que será de utilidad en el mismo. Este diseño conceptual se realiza independientemente de cualquier consideración física en términos de arquitectura del sistema y además deberá incorporar la semántica de la información.

Las tareas a realizar en el diseño conceptual son las siguientes:

- Identificación de las entidades del modelo
- Identificación de las relaciones entre entidades
- Identificación de los atributos de cada entidad
- Determinación de los dominios de los atributos
- Determinación de los identificadores (claves primarias y claves foráneas)
- Dibujo del diagrama entidad-relación

IV.2.1.15. Diseño lógico del modelo E-R

Se trata del diseño cuyo objetivo es transformar el esquema conceptual obtenido en la etapa anterior, adaptándolo al modelo de datos en el que se apoya el *SGBD* que se va a utilizar, en nuestro caso cualquier *SGDB* relacional.

IV.2.2. Diseño conceptual del modelo E-R para el sistema de información propuesto

El diseño de un modelo conceptual de datos debe estar sustentado en una serie de premisas básicas sin las cuales el modelo puede ser inestable en su comportamiento futuro o incluso inválido desde el punto de vista de la explotación de la información que contiene.

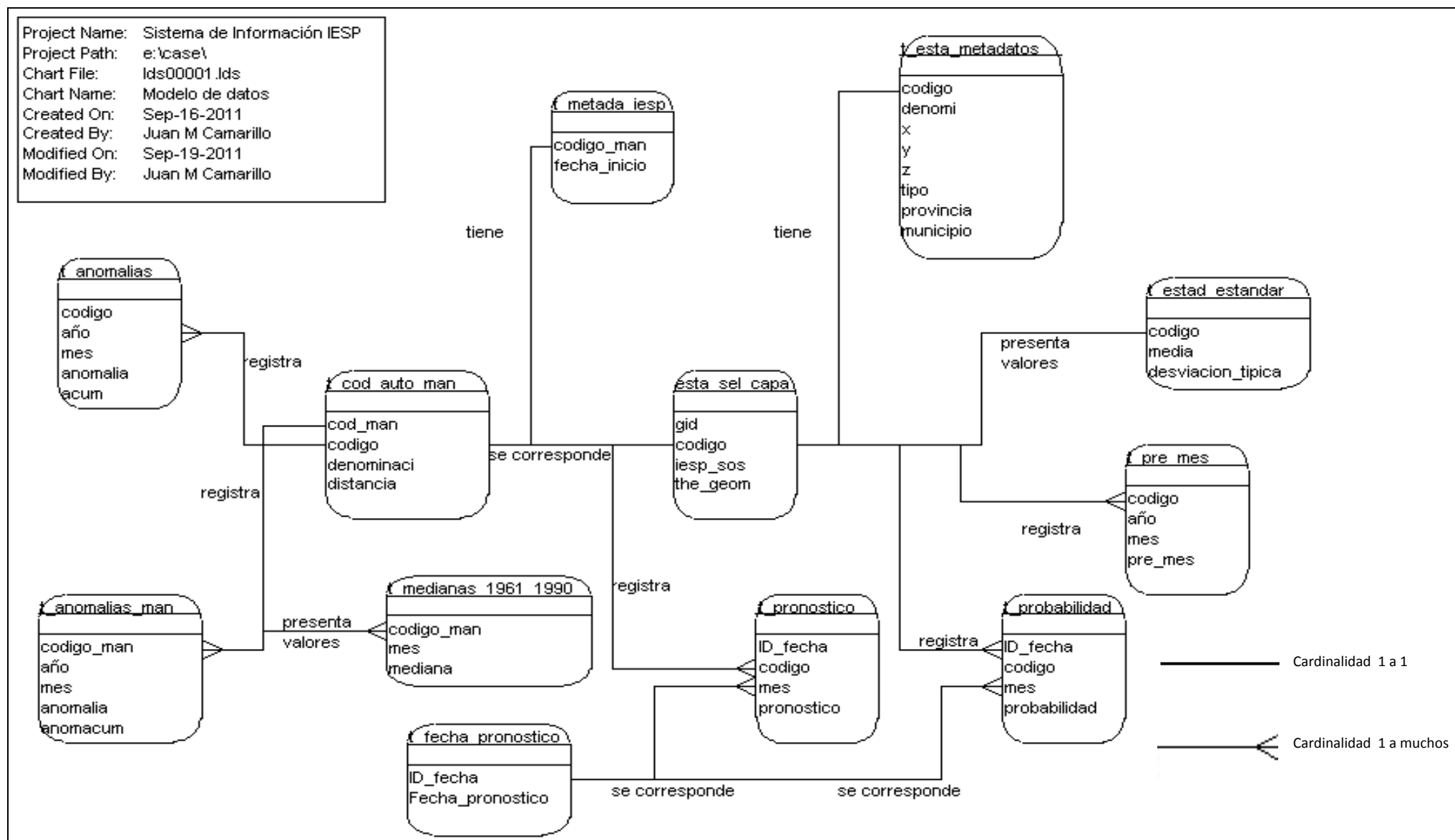
Desde esta perspectiva, el elemento esencial que debemos mantener presente en la fase de diseño conceptual del modelo es el establecimiento de los objetivos a los que debe responder. Desde esta óptica debemos hacer una reflexión que consideramos de suma importancia. Al contrario que los sistemas de gestión puros, como por ejemplo aquellos diseñados para almacenar, en nuestro caso, los billones de registros producidos por las distintas redes de observación cuyo objetivo principal es construir un modelo sólido y robusto especialmente orientado al almacenamiento masivo de datos, en el caso que ocupa a este trabajo de investigación, estaríamos hablando de una nueva generación de sistemas cuyo objetivo principal no es el almacenamiento y gestión de la información medida, sino, por el contrario, la flexibilidad y adaptabilidad en términos de explotación y cálculo de información derivada.

Son por tanto modelos que deben tener en cuenta, en su fase de diseño, las preguntas y respuestas para las que debe estar preparado; en nuestro caso, el cálculo del IESP como indicador ambiental de estado así como de las predicciones y probabilidades asociados al comportamiento de este. Es y debe ser por tanto un modelo *orientado al cálculo*, que debe garantizar el correcto flujo de datos para permitir la construcción de vistas vivas enlazadas, que muestren el valor del IESP en el conjunto de estaciones de referencia en cada mes de cálculo continuado.

El valor del IESP por tanto no se calculará sobre una tabla fija de realidad física, sino que mantendrá una realidad virtual que es, precisamente, la que garantiza su mantenimiento continuado a través del paso de la unidad temporal de cálculo en la escala mensual.

Esta segunda generación de modelos aplicados a la producción, cálculo y gestión de indicadores ambientales, exige en su diseño, y esta es una perspectiva que nos gustaría resaltar sobremanera, un importante conocimiento temático de la naturaleza de la propia información climática de partida, así como de los procedimientos de cálculo del indicador, pero incluso aún más importante, de preguntas tales como: ¿para qué sirve un indicador ambiental de estado como el IESP?, ¿en qué contexto se inserta su cálculo?, ¿de qué manera va a difundirse?, ¿va a ser utilizado por otros servicios o clientes humanos o máquinas?, ¿necesita ser interoperable?, ¿tendrá que ser servido para mostrar su variabilidad espacial y temporal diacrónica?,... Estas son las cuestiones que deben presidir el diseño del modelo conceptual y que realmente incorporan una fuente de valor añadido importante al concepto global del sistema. Y en esta esfera es precisamente, tal y como se defiende en este trabajo de investigación, donde se inserta el papel del experto en información climática y ambiental como vehículo de transmisión de objetivos, pero también plenamente con un papel activo en la fase de modelado y de construcción de las vistas y consultas necesarias para el funcionamiento completo del sistema.

IV.2.2.1. Modelo conceptual de datos entidad-relación propuesto (Figura IV.25).



IV.2.2.2. Descripción de las entidades

Las entidades presentes en este modelo son descritas a continuación.



Es la entidad central del modelo sobre la que pivotan el resto de entidades necesarias y cuyo identificador está presente como clave foránea en gran número del resto de entidades del modelo.

Esta entidad almacena la selección de estaciones meteorológicas de referencia en las que se calcula el comportamiento del IESP.

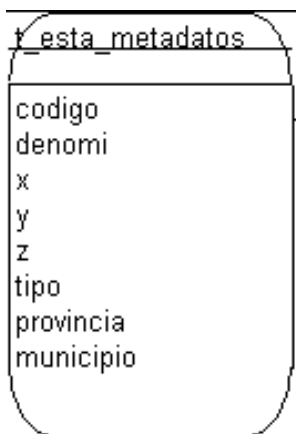
Atributos:

gid: identificador geográfico. Necesario ya que se trata de una entidad con geometría espacial cuyo manejo por la base de datos geográfica utilizada Postgis requiere de un identificador (*integer*) unívoco.

codigo: clave primaria de la entidad. Es el código identificador de la estación.

iesp_sos: atributo binario (0/1) que permitirá el control sobre las estaciones que mostrarán el comportamiento del indicador en el mes de cálculo. Necesario para tener control sobre las estaciones y los posibles errores o inconsistencias que se derivaran de los registros originales de precipitación del mes.

the_geom: atributo que almacenará la información de la geometría espacial de las estaciones.



Esta entidad almacena los metadatos básicos de las estaciones de referencia en las que se calcula el comportamiento del IESP.

Atributos:

codigo código identificador de la estación. Clave foránea de la relación con esta_sel_capa.

denomi: atributo que registra el nombre identificativo de la estación.

x: atributo para el valor de la longitud de la estación.

y: atributo para la latitud de la estación.

z: altura de la estación

tipo: tipo de estación meteorológica.

provincia: provincia en la que se ubica.

municipio: municipio en el que se ubica.

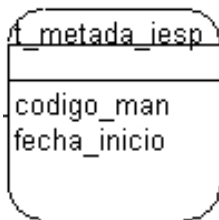


Esta entidad almacena las correspondencias necesarias entre las estaciones manuales en las que se inicia el cálculo del IESP(ver **anexo I**) y las estaciones automáticas de referencia para el cálculo del Indicador que se utilizarán desde la fase de desarrollo del sistema..

Atributos:

codigo código identificador de la estación. Clave foránea de la relación con esta_sel_capa.

cod_man: clave primaria fecha de inicio desde la que está disponible el valor del indicador mensual en las estaciones manuales de referencia.

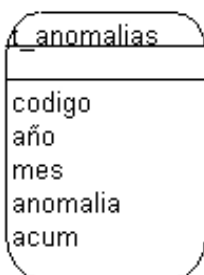


Esta entidad almacena los metadatos referidos al cálculo del Indicador IESP en cada una de las estaciones manuales de referencia.

Atributos:

codigo código identificador de la estación. Clave foránea de la relación con t_cod_auto_man.

fecha_inicio: fecha de inicio desde la que está disponible el valor del indicador mensual en las estaciones manuales de referencia.



Esta entidad almacena los valores de las anomalías pluviométricas del mes de cálculo del Indicador en las estaciones meteorológicas de referencia.

Atributos:

codigo: código identificador de la estación. Clave foránea de la relación con t_esta_sel_capa.

año: año del registro.

mes: mes del registro.

anomalia: anomalía pluviométrica del mes de cálculo

acum: anomalía pluviométrica acumulada con el mes anterior según formulación de cálculo del IESP.



Esta entidad almacena los valores de las anomalías pluviométricas del mes de cálculo del Indicador en las estaciones meteorológicas manuales. Es necesaria su inclusión ya que el sistema tiene pretensiones de ofrecer el servicio de los valores históricos del comportamiento del Indicador.

Atributos:

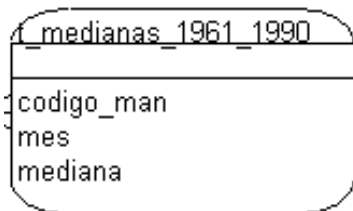
codigo_man: código identificador de la estación. Clave foránea de la relación con t_cod_auto_man.

año: año del registro.

mes: mes del registro.

anomalia: anomalía pluviométrica del mes de cálculo

anomacum: anomalía pluviométrica acumulada con el mes anterior según formulación de cálculo del IESP.



Esta entidad almacena los valores de la mediana de la serie normal para cada uno de los 12 meses del año en cada estación manual. Estas medianas son las que se utilizan para el cálculo de las anomalías pluviométricas.

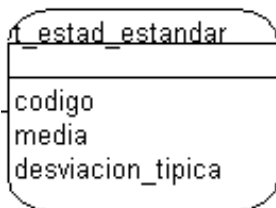
Atributos:

codigo_man: código identificador de la estación.

Clave foránea de la relación con t_cod_auto_man.

mes: mes de referencia para la mediana.

mediana: valor de la mediana en el periodo normal.



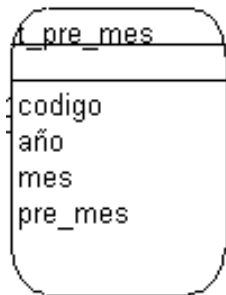
Esta entidad almacena los valores de los estadísticos (media y desviación típica) de las series del IESP para cada estación de referencia necesarios para el cálculo del IESP

Atributos:

codigo: código identificador de la estación. Clave foránea de la relación con t_esta_sel_capa.

media: valor de la media aritmética de la serie.

desviacion_tipica: valor de la desviación típica de la serie.



Esta entidad almacena los valores de precipitación mensual registrados en las estaciones de referencia. Los datos proceden de una consulta de inserción de registros realizadas en la geodatabase sobre la que se han volcado los datos diarios después del proceso de interpolación.

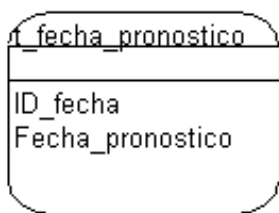
Atributos:

codigo: código identificador de la estación. Clave foránea de la relación con t_esta_sel_capa.

año: año del registro.

mes: mes del registro.

pre_mes: precipitación registrada en el mes

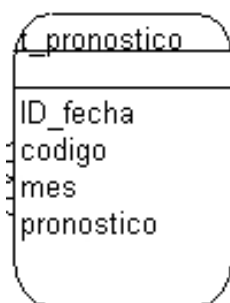


Esta entidad almacena las fechas para los pronósticos a 1,...,12 meses que se realizan para cada valor mensual del IESP calculado por el sistema.

Atributos:

ID_fecha: clave primaria para cada fecha de pronóstico.

Fecha_pronostico: fecha de pronóstico del IESP.



Esta entidad almacena los valores de los pronósticos realizados para el comportamiento del IESP en los 1,...,12 meses posteriores al cálculo del Indicador en el mes. Los datos proceden de la ejecución del *script* en R diseñado a tal efecto.

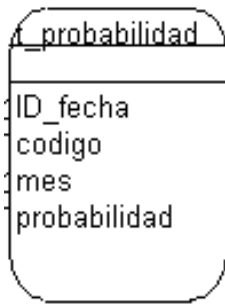
Atributos:

ID_fecha: clave foránea de la relación con la entidad t_fecha_pronostico.

codigo: código identificador de la estación. Clave foránea de la relación con t_esta_sel_capa.

mes: mes del pronóstico.

pronostico: pronóstico para el comportamiento del Indicador (seco/húmedo).



Esta entidad almacena los valores asociados de probabilidad de los pronósticos realizados para el comportamiento del IESP en los 1,...,12 meses posteriores al cálculo del Indicador en el mes. Los datos proceden de la ejecución del *script* en R diseñado a tal efecto.

Atributos:

ID_fecha: clave foránea de la relación con la entidad t_fecha_pronostico.

codigo: código identificador de la estación. Clave foránea de la relación con t_esta_sel_capa.

mes: mes del pronóstico.

pronostico: probabilidad del pronóstico para el comportamiento del Indicador (seco/húmedo).

IV.2.2.3. Vistas y definición de consultas (SQL)

Hemos afirmado anteriormente que el diseño del modelo propuesto está orientado fundamentalmente al cálculo del IESP, así como a los requerimientos derivados de su inserción en un sistema de indicadores climático-ambientales . Para ello se hará necesario definir las consultas en lenguaje SQL en el esquema de la base de datos previsto para ello. El conjunto y descripción de dichas consultas realizadas se encuentran en el Anexo 3 del trabajo de investigación y constituyen, soportada por el diseño del modelo, el núcleo central del funcionamiento operativo del sistema.

IV.3. Arquitectura del sistema

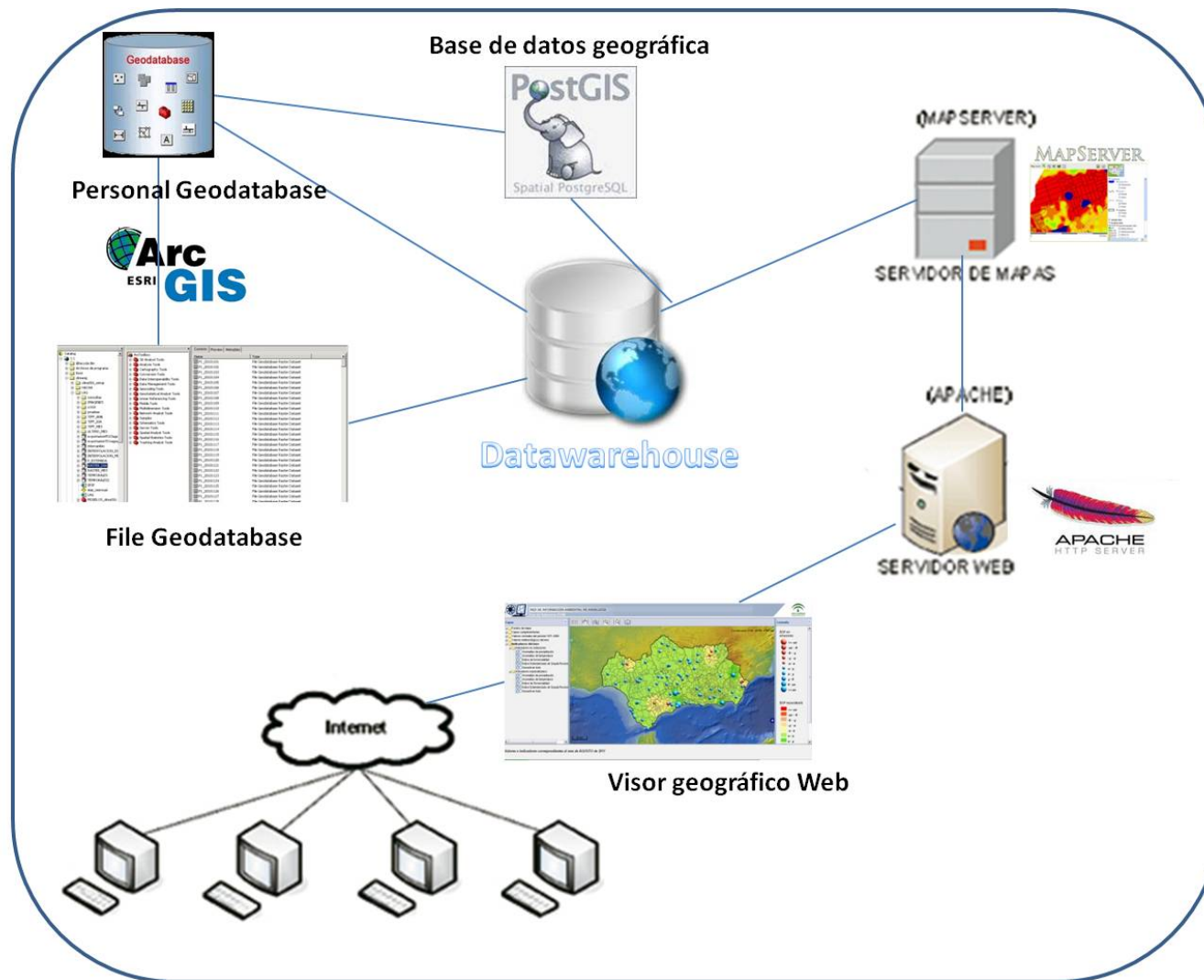


Figura IV.26. Arquitectura del sistema

Fuente: elaboración propia

IV.3.1. Documentación de la arquitectura del sistema

IV.3.1.1. Almacén de datos –*Datwarehouse*–

El almacén de datos del sistema está compuesto por tres *SGDB*:

- ***Geodatabases de ArGis:*** en ella se guardan los datos necesarios para la interpolación de la precipitación diaria a partir de los datos diarios de precipitación (PD_23) extraídos del Subsistema CLIMA. Almacenan también los datos de precipitación diaria para las 84 estaciones de referencia una vez interpolados los valores diarios ausentes. Seleccionada por la dependencia de lectura y escritura de los procedimientos espaciales insertados en el *model builder* diseñado para efectuar dicha interpolación.
- ***Filegeodatabse de ArcGis:*** almacena las superficies raster de los valores mensuales del IESP calculados. Su elección se sustenta en la mejora y rapidez de acceso sobre datos *raster* que presenta respecto a la personal *geodatabase* y cualquier gestión de archivos de sistema.
- ***Base de datos geográfica PostGis:*** almacena el modelo de datos para el cálculo del IESP y los datos que alimentan al mismo. Guarda las consultas SQL que proporcionan las vista del comportamiento mensual del IESP. Es una base de datos geográficas para asociar dichas vistas a la geometría de las estaciones meteorológicas. Se trata de software libre, apropiado para las dimensiones del sistema diseñado, sin coste operativo alguno. Su elección está basada también en la propia estrategia de la Junta de Andalucía en relación a la extensión del software libre en las aplicaciones de la administración.

IV.3.1.2. Software de análisis espacial

El software de análisis espacial seleccionado en la arquitectura del sistema es *ArcGis*. La capacidad analítica que proporciona, las herramientas de conexión a bases de datos *-geodatabases-* y, fundamentalmente la posibilidad de diseñar modelos de procesos encadenados mediante la herramienta *model builder* proporcionan una enorme operatividad casi automática en el lanzamiento y ejecución de los procesos que son necesarios para el correcto funcionamiento del sistema.

IV.3.1.3. Servidor de mapas

El servidor de mapas seleccionado, *MapServer*, se encuentra entre una de las opciones técnicas más utilizadas en las arquitecturas de sistemas de información geográfica actuales. El hecho de ser software libre y, por lo tanto libre de licencias – en contraposición a la versión comercial de *ArcGis server-*, la multitud de formatos de entrada que lee, su rápida conexión a bases de datos y especialmente a *PostGis*, compensa el trabajo de asignación de semiologías que debe realizarse antes de la generación del servicio *WMS*.

IV.3.1.4. Servidor Web

El servidor de aplicaciones web seleccionado, *Apache*, se encuentra englobado también dentro del software libre y es también uno de los más utilizados hoy día.

IV.3.1.5. Visor de mapas Web

El visor cartográfico web, si bien su diseño no entra dentro del diseño funcional del sistema, ya que no juega un papel crítico en los objetivos de este trabajo, ha sido desarrollado por el equipo de investigación bajo la arquitectura, también de software libre, de *OpenLayer* y de su extensión *GeoExt*.

PARTE TERCERA: RESULTADOS, CONCLUSIONES Y LINEAS DE TRABAJO FUTURAS

CAPITULO V: RESULTADOS DE LA INVESTIGACION OBTENIDOS

Los resultados obtenidos por el trabajo de investigación se expresaran en términos de consecución de objetivos concretos alcanzados por un lado y en términos de satisfacción de las hipótesis de partida por otro.

V.1. Resultados concretos y objetivos alcanzados por el trabajo de investigación

Entre los que creemos más interesantes debemos destacar los siguientes:

1. Desarrollo de un sistema de información para el seguimiento de la sequía plenamente operativo dentro del Subsistema de Indicadores climático-ambientales de la Red de Información Ambiental de Andalucía (REDIAM).

El resultado operativo del mismo es el cálculo del IESP mes a mes para Andalucía y su difusión en un visor cartográfico dinámico web que permite su visualización espacial así como el acceso a los datos sobre las estaciones de referencia y la cartografía continua proporcionada, cumpliendo así con las necesidades especificadas en el objetivo número 1 del capítulo correspondiente.

Este visor web puede consultarse en la web www.climasig.es/inndex_phtml y algunas de sus funciones y aspecto pueden observarse en la figura V.1.

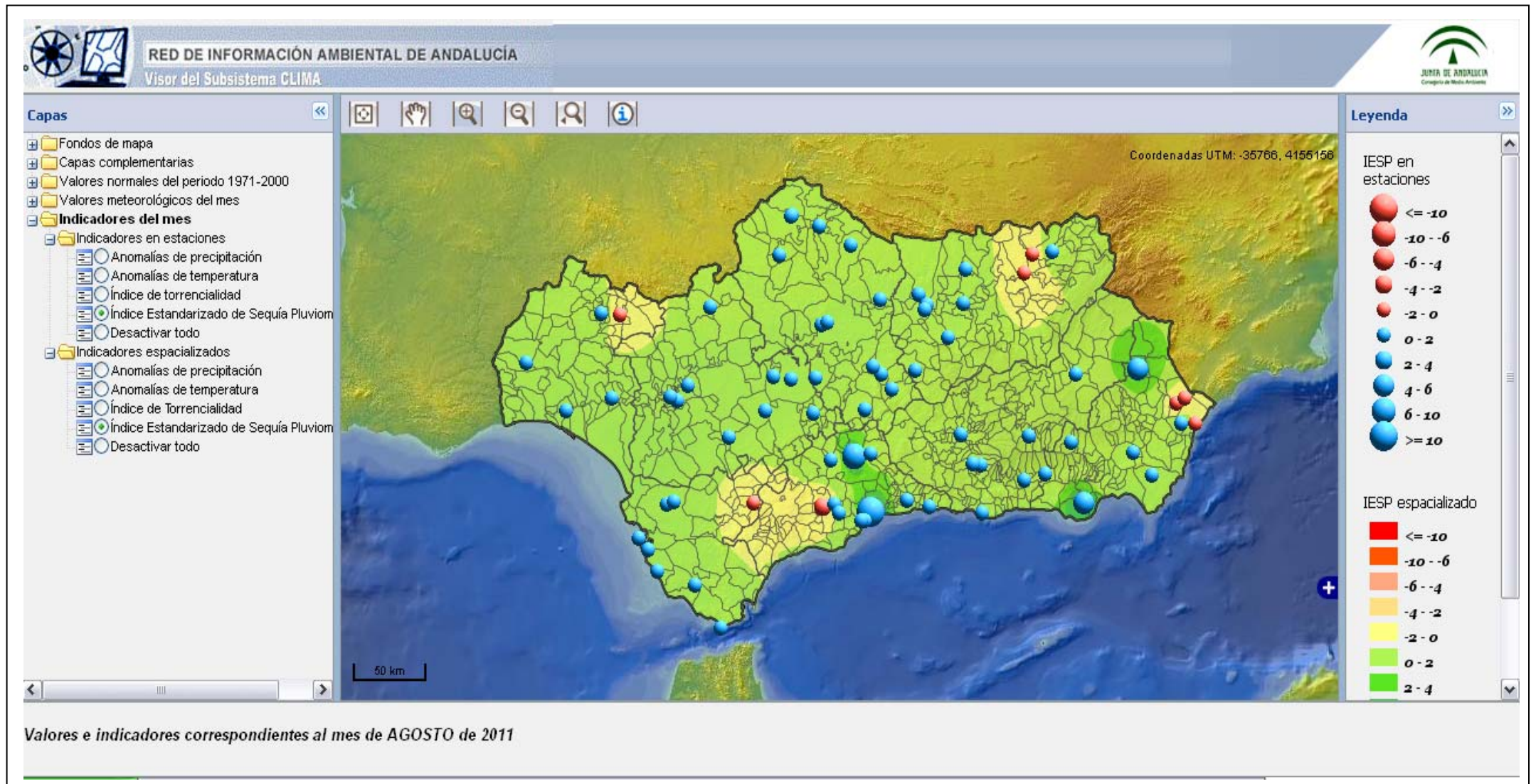


Figura V.1 Captura de la cartografía del IESP en el visor de mapas WebIESP

Fuente: Consejería de Medio Ambiente. Junta de Andalucía

2. Se ha analizado y puesto de manifiesto en todas las fases del diseño funcional del sistema las numerosas características especiales y particulares que presenta la información meteorológica y climática en el contexto de su inclusión en sistemas operacionales de alerta temprana. Estas características obligan a un tratamiento particularizado y específico de las mismas –interpolación de la precipitación diaria, PD₂₃, procesos de espacialización y de interpolación espacial, inserción de procedimientos de cálculo de pronósticos, etc...- sin los cuales se haría imposible el funcionamiento operacional del sistema diseñado.
3. Se aporta un modelo de datos entidad-relación (E-R) universal y válido para el tratamiento de la información climática necesario para el cálculo en tiempo real o *near real time* del Índice Estandarizado de Sequía Pluviométrica IESP. Este modelo incluye el conjunto de entidades y relaciones necesarias para un flujo correcto de la información en términos de cálculo operacional del indicador.
4. El trabajo proporciona una metodología completa para el diseño y desarrollo de un sistema de seguimiento de la sequía basado en el cálculo del IESP mensual. Este esquema metodológico queda estructurado a partir de los diagramas de flujo de datos del análisis funcional del sistema, el modelo de datos entidad-relación propuesto y la arquitectura del sistema adoptada. De esta forma se facilita en el futuro los posibles diseños basados en este indicador que puedan generarse o aplicarse en otros ámbitos geográficos.
5. Se ha generado una transferencia real y tangible de conocimiento desde el ámbito de la investigación al ámbito de la gestión, poniéndose de manifiesto la simbiosis necesaria y plausible, creemos, en el contexto de nuestra disciplina. El binomio investigación básica e investigación aplicada creemos que se pone de manifiesto en este trabajo ya que

participa de ambos elementos: por un lado ha sido necesario investigar una serie de procedimientos y análisis aplicados a la información meteorológica, así como una evaluación de las necesidades, requisitos y problemáticas asociadas a la misma y por otro, los resultados de lo anterior ha supuesto la implementación operativa de un sistema de información en el ámbito de la gestión.

6. Se aporta adicionalmente una metodología analítica completa para el relleno de lagunas en los datos primarios de precipitación diaria procedentes de las estaciones meteorológicas automáticas utilizadas. Este método mejora los resultados de los procesos de relleno que venían aplicándose por los servicios de la Consejería, se encuentran en consonancia con los resultados alcanzados por otros estudios en Andalucía y ha sido implementado operativamente en el sistema de información.
7. De la misma forma, se aporta una metodología para el cálculo y asignación de pronósticos y probabilidades de ocurrencia del comportamiento del IESP en horizontes futuros, cumpliendo de esta forma con uno de los elementos comunes presentes en los sistemas de alerta temprana por sequía operativos, la proyección del comportamiento de los indicadores. Esta metodología aportada también ha sido implementada desde un punto de vista operacional.
8. El sistema se integra dentro de un esquema más amplio, como puede observarse en los resultados del visor geográfico operativo, constituido por el sistema de información de indicadores climático-ambientales integrados en la REDIAM andaluza.

V.2. Satisfacción de las hipótesis propuestas

1. Entendemos que la primera hipótesis ha sido completamente verificada, ya que el cálculo operacional en tiempo real de un indicador se sustenta en su inserción dentro de un sistema de información configurado a través de diferentes etapas básicas: análisis y diseño funcional del mismo, diseño del modelo de datos, proposición de la arquitectura física del mismo e implementación operativa.
2. La segunda hipótesis se encuentra testada como metodología de base que sustenta el diseño de sistemas de información. Su validez aplicada a sistemas de información de naturaleza climatológica y de producción de indicadores y concretamente del IESP queda explícitamente probada con los resultados alcanzados en este trabajo de investigación.
3. Entendemos que han quedado demostradas la especificidad e implicaciones que la naturaleza de la información climática impone en el diseño de sistemas de información que la manipulan y que, recordemos, eran las siguientes:
 - Diseño de un modelo de datos previo que garantice el correcto flujo de información orientado al cálculo de un indicador climático..
 - Incorporación a ese modelo de datos de la componente espacial, convirtiéndolo, de esta manera, en un modelo de datos geográfico.
 - Necesidad de diseño de procesos de interpolación espacial específicos que deben quedar insertados en el sistema.
 - Necesidades de diseño de semiologías gráficas adecuadas para su correcta interpretación y difusión
4. La verificación de la cuarta hipótesis es, sin duda, uno de los aspectos seguramente más relevante y de mayor proyección de la tesis. En este

sentido, la inclusión o participación de un experto en información climática entendemos que demostrada con la conclusión de este trabajo de investigación, puede aportar numerosas ventajas en el diseño e implementación de sistemas de información climáticos. Por un lado, el conocimiento de la naturaleza de la información y de los procedimientos de análisis que le son exigidos; por otro, el conocimiento profundo de los objetivos que deben perseguir estos sistemas. Además entendemos que, trabajar con el mismo lenguaje y herramientas propias de la ingeniería de sistemas proporciona una serie de ventajas en términos de productividad y garantía de éxito operativo de los sistemas indudable. Nuestra propia experiencia en estos diseños así nos lo indica también.

Por último, el proceso de diseño conceptual del modelo de datos se ve reforzado desde el momento en el que, como expertos analistas en la información climática, ofrecemos una visión global y completa de las preguntas a las que debe responder dicho modelo que, no lo olvidemos, en el contexto de este trabajo, se orienta hacia la explotación y el cálculo de variables derivadas mucho más que al almacenamiento de datos primarios.

CAPITULO VI: CONCLUSIONES, DISCUSION Y LINEAS DE TRABAJO FUTURAS

VI.1. Conclusiones

Las conclusiones principales que se derivan de la materialización de este proyecto de tesis doctoral, en consonancia con los resultados logrados y los objetivos alcanzados, son las siguientes:

1. La importancia creciente de los sistemas de alerta temprana por sequías están propiciando una nueva fase en el diseño de sistemas de información climatológica. Se trata de sistemas de información orientados a la explotación y a la difusión de derivados climáticos –índices-.
2. Uno de los aspectos más importantes desde el punto de vista anterior, pero que también encierra las máximas dificultades en términos de producción, es la necesidad de que estos índices se calculen con la periodicidad y temporalidad suficiente como para verdaderamente ser indicadores del estado y evolución del fenómeno. En nuestro caso, esa periodicidad debe ser mensual ya que el IESP es un índice de sequía mensual. Por lo tanto es obligado el uso de información procedente de estaciones meteorológicas automáticas. Y para ello, son comunes a los diseños de los sistemas los siguientes elementos:
 - conexión directa a estaciones o a los servidores de los sistemas de producción de los organismos que las gestionan.
 - diseño de una capa de control de la calidad de los datos especialmente para detectar los errores sistemáticos, los errores aleatorios y los datos aberrantes sobre la variables pero también sobre su distribución espacial incoherente.

- diseño de procedimientos operativos de relleno de lagunas en los valores diarios de la precipitación. Estos procedimientos deben aunar varias características: deben ser operables, es decir, deben permitir la posibilidad de integrarlos operativamente en el sistema, deben garantizar unos resultados científicamente válidos y aceptables y la relación coste-beneficio de su desarrollo debe ser buena.
3. Un aspecto fundamental en la concepción de estos sistemas de información como sistemas de alerta temprana es, como hemos visto anteriormente, la importancia de la disponibilidad y difusión del comportamiento del indicador –de ahí por qué son sistemas de monitorización y seguimiento-. Y desde este punto de vista debemos destacar el papel que juega la cartografía dinámica servida a través de Internet en entornos de visores web. Esta cartografía dinámica implica algunos elementos que nos parece interesante y necesario destacar:
- implica una conexión viva de las capas disponibles en el visor a la base de datos y a las vistas (SQL) que muestran el valor del indicador en el momento de acceso a la información. Esta conexión garantiza que el usuario está accediendo al último y más reciente cálculo del indicador que se encuentra en el sistema, en contraposición a las cartografías estáticas que dependen de procesos de actualización reiterativos.
 - implica el acceso interactivo e inmediato al dato que se muestra en el visor, elemento este inaccesible en una cartografía estática que lo que nos muestra es una imagen preprocesada.
 - implica la posibilidad de ofrecer al usuario la forma en la que ve esa información, especialmente basada en la posibilidad de combinar la información del indicador con alguna de las otras capas disponibles.
 - Implica en el contexto tecnológico actual la integración de estándares de servicios de información geográfica por Internet –*WMS (Web Map Service)*, *WFS (Web File Service)* o *WCS (Web Coverage Service)*-. Desde esta perspectiva, el valor del indicador en cualquier momento y fecha

puede ser servido como servicio interoperable para ser integrado en otros sistemas de información o usado por usuarios individuales.

4. Entendemos que la experiencia de incorporarnos a la fase de diseño operativo de un sistema de estas características aporta elementos novedosos, tanto desde la perspectiva de la disciplina como desde la perspectiva del funcionamiento operativo conseguido.

5. El esfuerzo de diseño conceptual y de modelado de datos que sustenta el almacenamiento, la gestión y el flujo de datos necesarios para el cálculo óptimo del indicador, ha servido para aportar un mayor y mejor conocimiento de la propia sequía, de su medición en términos de magnitud, intensidad y duración. Y lo que es especialmente destacable, en términos de su concepción operativa, de seguimiento y de monitorización.

6. Pensar en el dato climatológico y geográfico desde una orientación de explotación y difusión del mismo obliga a relacionarnos con conceptos y lenguajes propios de la ingeniería de sistemas, las bases de datos, los lenguajes de consulta a los mismos estándares, los protocolos y estándares de difusión de la información geográfica por Internet, etc... de la misma manera que pensar en el dato climatológico desde la orientación del análisis exploratorio de los mismos nos obliga a relacionarnos con el lenguaje matemático, las herramientas gráficas o los programas estadísticos o de información geográfica. Todos ellos no son sino herramientas que tenemos disponibles para lograr los objetivos propuestos, un análisis temporal de la evolución de alguna variable climática o el cálculo operativo de un índice de sequía.

7. La puesta en marcha y plena operatividad de un sistema como el propuesto en este trabajo para Andalucía creemos que, para el ámbito de la gestión de sequías y del tratamiento de indicadores ambientales, coloca

a la región en una buena posición científico-tecnológica que abre numerosas vías de futuro y que coloca unas bases sólidas para seguir avanzando en la definición de un sistema completo de alerta temprana por sequías en la Comunidad.

VI.2. Líneas de trabajo de futuro

Las líneas futuras de trabajo de este trabajo de investigación pueden ser resumidas en los siguientes apartados, de las más específicas a las más genéricas:

1. Desarrollar una capa de control en el sistema para la detección y posible corrección de los errores sistemáticos, aleatorios y aberrantes en los datos de precipitación diaria procedentes de las estaciones meteorológicas automáticas. La experiencia nos dice que se producen errores aberrantes no sobre las magnitudes de las variables, sino particularmente sobre la coherencia espacial que estos deben asegurar. Es por ello que errores de medición incompatibles con la estructura espacial de la precipitación en un entorno en un día en concreto deben ser aislados y corregidos de forma sistemática.
2. Mejorar los procedimientos de relleno de lagunas del proceso “Interpolación PD_23”. Evaluar otros algoritmos de interpolación espacial en el cálculo de la precipitación regional y verificar si existe algún patrón en los semivariogramas de la precipitación diaria para diseñar un procedimiento basado en *kriging* será una línea de investigación futura.
3. Insertar el trabajo realizado en otros proyectos de ámbito regional, nacional e internacional. Entre ellos cabe destacar el Proyecto internacional “Arizona”, codirigido por el Dr. Leandro del Moral Ituarte y que tiene como objetivo precisamente el diseño de herramientas de

gestión y de monitorización de la sequía de forma integral y abarcando todos los tipos desde la meteorológica a la social. Asimismo, la participación del autor en el Proyecto concedido del Plan Nacional de I+D+i “Directiva marco del agua y riesgos hídricos: gestión y mitigación de sequías”, dirigido por la Dra. Pilar Paneque Salgado, garantiza la aplicación del sistema operativo desarrollado y la difusión de los conocimientos adquiridos a contextos de investigación y aplicados orientados específicamente a la gestión de las sequías.

4. La inmediata redacción de dos artículos dirigidos a revistas internacionales de impacto que permitan evaluar con criterios de excelencia las bases conceptuales y teóricas que sustentan este proyecto de tesis doctoral.

REFERENCIAS BIBLIOGRAFICAS

ALBERCH FUGUERAS, Ramón; CRUZ MUNDET, José Ramón. **La aventura de la información : de los manuscritos del Mar Muerto al imperio Gates**. Madrid: Alianza Editorial, 2004. ISBN 84-206-5736-0

ALEXANDERSON, H. "A Homogeneity test applied to precipitation data". *Journal of Climatology*. Nº 6 (1986), p. 661-675. ISSN 0894-8755

BANNAYAN, M.; HOOGENBOOM, G. "Weather analogue : a tool for real-time prediction of daily weather data realizations based on a modified k-nearest neighbor approach". **Environmental Modelling and Software**. Nº 23 (2008), p. 703-713. ISSN 1364-8152

BERTALANFFY, Ludwig von. **General System theory : foundations, development, applications**. Revised ed. New York: George Braziller, 1976. ISBN 0-8076-0453-4

BHALME, H.N; MOOLEY, D.A. "Large scale droughts/floods and mosson circulation". **Monthly Weather Review**. Nº 188 (1980), p. 1197-1211. ISSN 0027-0644

BRABEC, B.; MEISTER, R. "A nearest-neighbor model for regional avalanche forecasting". **Annals of Glaciology**. Nº 32 (2001), p. 130-134. ISSN 0260-3055

CARBONE, Gregory J. et al. "A Regional-Scale Drought Monitoring tool for the Carolinas". **Bulletin of the American Meteorological Society**. Vol. 89, Issue 1 (jan. 2008). p. 20-28. ISSN 1520-0477

CASTELLS, Manuel. **La era de la información : economía, sociedad y cultura**. Madrid: Alianza, 2001-2005. ISBN 84-206-4246-0

CASTILLO REQUENA, José Manuel. "Evolución de la precipitación anual en las regiones pluviométricas de Andalucía : observaciones de Geografía comparada sobre recursos hídricos". **Cuadernos Geográficos**. Nº 30. p. 123-141. ISSN 0210-5462

CLAUSIUS, R. “Über die bewegende Kraft der Wärme”. **Annalen der Physik und Chemie**. Nº 79 (1850). p. 368-397, 500-524
<http://gallica.bnf.fr/ark:/12148/bpt6k15164w/f384.table>

CLAUSIUS, R. “Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie”. **Annalen der Physik und Chemie**. Nº 125 (1865). p. 353-400
<http://gallica.bnf.fr/ark:/12148/bpt6k152107/f369.table>

COLLAZOS, Guillermo. “Sistema Soporte de Decisión para la evaluación y optimización económica de Sistemas de Recursos Hídricos” Director: J. Andreu. Universidad Politécnica de Valencia, Departamento de Ingeniería Hidráulica y Medio Ambiente, 2004

DAMAS HERMOSO, Miguel. “Integración de sistemas para la supervisión, control y optimización de redes de abastecimiento de agua (aplicación a la ciudad de Granada)”. Director: Director: Gonzalo Olivares Ruiz. Universidad de Granada, Facultad de Ciencias, 1999

DATE, C.J. **Introducción a los sistemas de bases de datos**. 7ª ed. México: Prentice-Hall, 2001 ISBN 968-444-419-2
<http://unefazuliasistemas.files.wordpress.com/2011/04/introduccion-a-los-sistemas-de-bases-de-datos-cj-date.pdf>

DE MIGUEL, A.; PIATTINI, M.; MARCOS, E. **Diseño de bases de datos relacionales**. Paracuellos del Jarama (Madrid): Ra-Ma, 1999. ISBN 84-7897-385-0

ELÍAS CASTILLO, F.; RUIZ BELTRÁN, L. **Precipitaciones máximas en España**. Madrid: Servicio de Publicaciones Agrarias, Ministerio de Agricultura, 1979. ISBN 84-7479-029-8

ELMASRI, R.; NAVATHE, S.B. **Sistemas de bases de datos : conceptos fundamentales**. 3ª ed. Madrid: Addison-Wesley Iberoamericana, 2000. ISBN 84-7829-051-6

ESCALONA CUARESMA, María José. “Metodologías para el desarrollo de sistemas de información global: análisis comparativo y propuesta” Directores: Manuel Mejías Risoto, Jesús Torres Valderrama. Universidad de Sevilla, Escuela Técnica Superior de Ingeniería Informática, 2001

<http://www.lsi.us.es/docs/informes/EstadoActual.pdf>

GERMAIN, Richard James. “Drought management using a geographical information system”. Director: T. J. Chang. Ohio University, Civil Engineering, 1996

GIBBS, W.J. **A Drought Watch System**. [Geneva]: World Meteorological Organization, 1987. WMO/TD-No.193 WCP-134

GRAU GADEA, Gonzalo Francisco. “Propuesta de modelo integrado de sistema de información para la gestión de unidades básicas de enfermería en el marco IE-GIP. Aplicación al sistema sanitario del ejército español”. Directores: Lorenzo Ros McDonnell, Ángel Ortiz Bas. Universidad Politécnica de Valencia, Departamento de Organización de Empresas, 2004

GUTIÉRREZ, S. Raúl. **Introducción al método científico**. 18ª ed. México: Esfinge, 2006. ISBN 970-782-176-0

HEIM, R.R. “A review of twentieth-century drought indices used in the United States”. **Bulletin of the American Metrological Society**. N° 83 (2002), p. 1149-1165. ISSN 1520-0477

HURTADO MORENO, G. “Índices de sequía y su aplicación operativa en Colombia”. **Atmósfera**. Número 5 (marzo 1986). ISSN 0187-6236

INDICADORES ambientales. Una propuesta para España. Madrid: MINISTERIO DE Medio Ambiente, 1996

KATZ, R.W.; GLANTZ, M. “Anatomy of a rainfall index”. **Monthly Weather Review**. N° 114 (1986), p. 764-771. ISSN 0027-0644

KENDALL, Kenneth E.; KENDALL, Julie E. **Análisis y diseño de sistemas**. 6ª ed. México: Pearson Educación, 2005. ISBN 970-26-0577-6

LAUDON, Kenneth C., LAUDON, Jane P. **Sistemas de información gerencial**. 8ª ed. México: Pearson, 2004. ISBN 970-260-528-8

LONG, Larry; LONG, Nancy. **Introducción a las Computadoras y a los sistemas de información**. 5ª ed. México: Prentice Hall, 1997. ISBN 970-17-0262-X

McKEE, T.B.N.; DOESKEN, J.; KLEIST, J. "The relationship of drought frequency and duration to time scales". **Eight Conference On Applied Climatology**. Anaheim, C.A. American Meteorological Society (1993), p. 179-184.

MARCOS, O. "Sequía: definiciones, tipologías y métodos de cuantificación" **Investigaciones Geográficas**. Nº 26 (2001), p. 59-80

MATTELART, Armand. **Historia de la sociedad de la información**. Barcelona: Paidós, 2002. ISBN 84-493-1191-8

MÁXIMAS lluvias diarias en la España peninsular. Madrid: Ministerio de Fomento, Dirección General de Carreteras, 1999. ISBN 84-498-0419-1
http://epsh.unizar.es/~serreta/documentos/maximas_Lluvias.pdf

OUARDA, T. et al. "Generation of ESP Forecast Series with the Non-parametric Nearest Neighbor Approach". En: **Water 99: Joint Congress; 25th Hydrology & Water Resources Symposium, 2nd International Conference on Water Resources & Environment Research; Handbook and Proceedings Water**, 1999. p. 1154-1159. ISBN 1858257165

PÉREZ CUEVA, A.J. "Precipitaciones extraordinarias en la España Peninsular". **Agricultura y sociedad**. Nº 28 (1983), p. 189-203. ISSN 0211-8394

PINTÉR, J.D. "Continuous Global Optimization Software: A Brief Review", **Optima : Mathematical Programming Society Newsletter**. Nº 52 (dec. 1996), p. 1-8.
<http://www.mathopt.org/Optima-Issues/optima52.pdf>

PITA, M.F. Los Impactos del Cambio Climático Sobre el Riesgo de Sequía en España. **Libro del III Congreso Andaluz de Desarrollo Sostenible y VII Congreso Andaluz de Ciencias Ambientales. Huelva.** Federación Andaluza de Ciencias Ambientales y Universidad de Huelva. 2008. Pag. 126-139

PITA, M.F. “Recomendaciones para el Establecimiento de un Sistema de Indicadores para la Previsión, el Seguimiento y la Gestión de la Sequía”. **La Sequía en España. Directrices para Minimizar su Impacto.** Madrid: Ministerio de Medio Ambiente. (2007), p. 107-132

PITA, M.F. et al. “La covariación espacial de la precipitación en la Península Ibérica y su contribución al establecimiento de mecanismos de compensación interterritorial como instrumento de ayuda en la gestión de la sequía”. En: **La directiva Marco del Agua : realidades y futuros : actas del III Congreso Ibérico sobre Planificación y Gestión de Aguas (Sevilla 2002).** Zaragoza: Fundación Nueva Cultura del Agua, 2003. p. 110-115. ISBN 84-7820-700-7
http://www.upo.es/ghf/giest/documentos/clima_y_territorio/PitayOtros2002_covariacionespacial.pdf

PITA, M.F. “Sequías en la cuenca del Guadalquivir”. En: **Causas y consecuencias de las sequías en España.** Alicante: Instituto Universitario de Geografía de la Universidad de Alicante, 2001. p. 303-343. ISBN 84-7908-600-9

PITA, M.F. “Un nouvel indice de sécheresse pour les domaines méditerranéens. Application au bassin du Guadalquivir (SudOuest de l’Espagne)”. Nice: **Publications de l’Association Internationale de Climatologie**, 2001, vol. 13. p. 225-234
http://www.upo.es/ghf/giest/documentos/clima_y_territorio/Pita_unnouvelindexdesecheresse.pdf

PITA LOPEZ, María Fernanda. **Las sequías: análisis y tratamiento.** Sevilla: Consejería de Medio Ambiente de la Junta de Andalucía, 1995. ISBN 84-87294-67-7

QUIRING, Steven M., 2009 “Developing Objective Operational Definitions for Monitoring Drought”. **Journal of Applied Meteorology and Climatology.** Vol. 48, Issue 6 (jun. 2009). p. 1217–1229. ISSN 1558-8424

RODRÍGUEZ BALLESTEROS, C. “Las estaciones termopluviométricas de la red climatológica”. En: **Calendario meteorológico 2009**. Madrid: Ministerio de Medio Ambiente y Medio Rural y Marino, 2009. ISBN 978-84-8320-480-1

RODRÍGUEZ MOREDA, Ángel. “Propuesta de una arquitectura de red corporativa y su aplicación a la reingeniería administrativa”. Director: Rafael Mompo Gómez. Universidad de Valladolid. Escuela Técnica Superior Ingenieros de Telecomunicación de Madrid, 2003

SCHUMACHER, Marion. **The use of SSADM (Structured Systems Analysis and Design Methodology) as a standard methodology on Information Systems Projects**. 2001. ISBN 978-3-640-04313-2 (e-book)

SUÁREZ SÁNCHEZ, Ana. “Sistema integrado de información para la gestión de la auditoría reglamentaria de prevención de riesgos laborales”. Director: Luis Escanciano Montousse. Universidad de Oviedo, 2002

SVOBODA, M. et al. “The Drought Monitor”. **Bulletin of the American Meteorological Society**. Nº 83 (2002). ISSN 1181-1189

VAN ROOY, M.P. “A rainfall anomaly index independent of time and space”. **Notos**. Nº 14 (1965), p. 43-48

VARELA PET, José. “Sistema de información con capacidad de análisis espacial y multicriterio para ayuda a la toma de decisión en la gestión de los medios aéreos usados en la lucha contra incendios forestales”. Directores: Juan Enrique Arias Rodríguez, José Manuel Cotos Yáñez. Universidad de Santiago de Compostela, Facultad de Física, 2001

VILLAZÁN GARCIA DE LA PEDROSA, Luis Javier. “Sistema de información de gestión avanzada en el sector de las telecomunicaciones”. Directora: María Luisa Cuadrado Ebrero. Universidad Politécnica de Madrid, Departamento de Matemática aplicada a las tecnologías de la información, 1997

WHILHITE, D.; GLANTZ, M.H. “Understanding the drought phenomenon: the role of the definitions”. **Water International**. Nº 10 (1985). p. 111-120. ISSN: 0250-8060

WHILHITE, D.; SVOBODA, M. **Drought Early Warning Systems in the Context of Drought Preparedness and Mitigation.** 2005

http://www.unisdr.org/files/1882_VL102149.pdf

WHITTEN, J.L.; BENTLEY, L.D.; DITTMAN, K.C. **Systems analysis & Design methods.** [S.l.]: McGraw-Hill, 2004. ISBN 0-07-305233-7

XIA, Yulu; STINNER, Roland E.; PING-CHU, Chu. “Database integration with the Web for biologists to share data and information” **Electronic Journal of Biotechnology.** Vol.5 No.2, Issue of August 15, 2002. p. 154-161. ISSN 0717-3458 (versión online)

<http://www.scielo.cl/pdf/ejb/v5n2/a08.pdf>

ZAHUMENSKÝ, Igor. **Guidelines on Quality Control Procedures for Data from Automatic Weather Stations.** World Meteorological Organization Commission for Instrument and Methods of Observation. OPAG-SURFACE. Expert Team on Surface Technology and Measurement Techniques. First Session. Geneva, Switzerland, 13-16 October

Anexo I. Selección de estaciones meteorológicas para el cálculo y seguimiento histórico del IESP

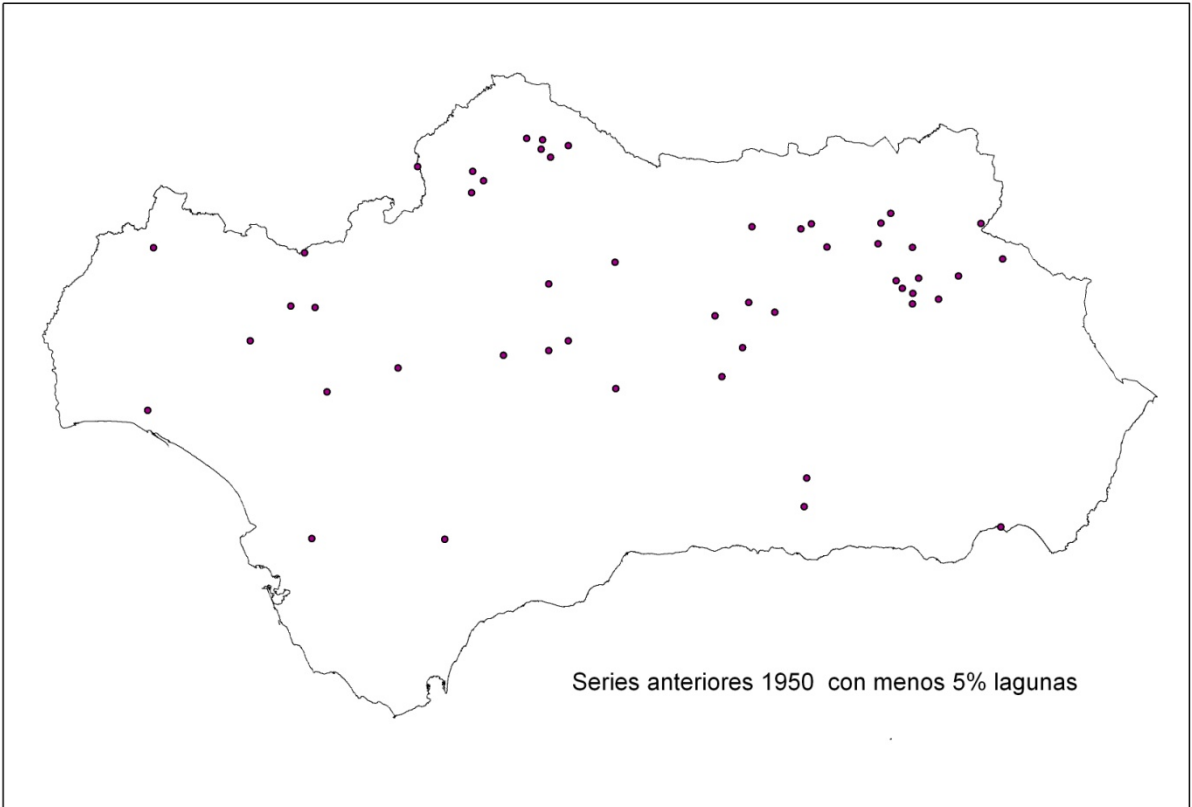
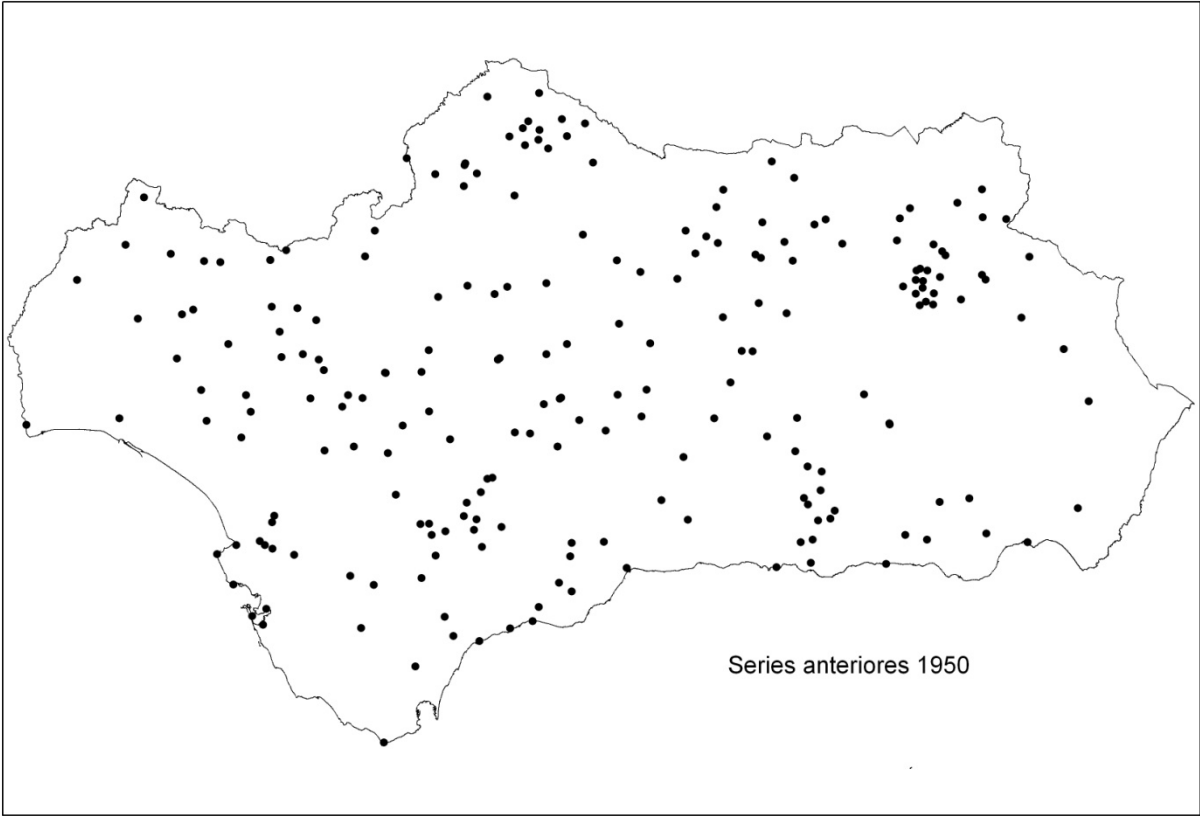
I.I. Series anteriores a 1950.

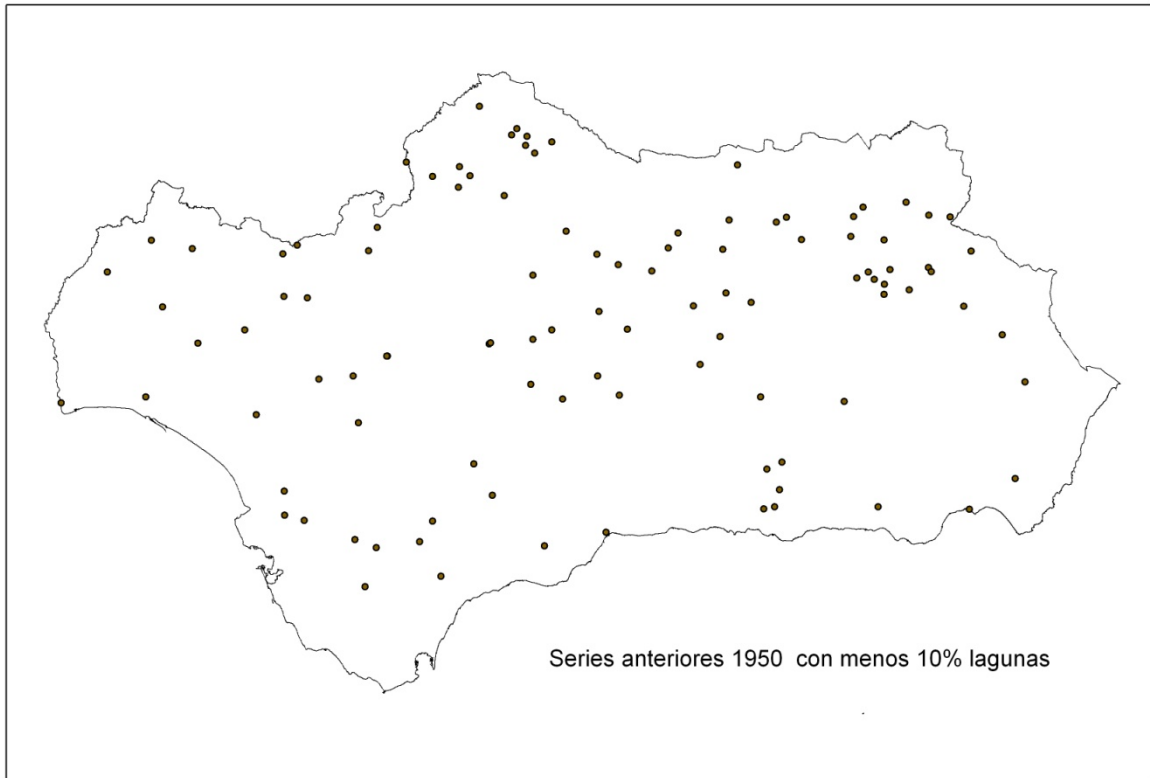
Partimos de 373 estaciones que inician sus registros mensuales de precipitación antes de 1950 de las cuales 196 llegan hasta la actualidad (finalizan en el año 2000 o en años posteriores). Para que nos sirva de orientación, desde 1930 al año 2000 tenemos *aproximadamente* 70 años de datos lo que supone un total de 840 meses (para tener cifras redondas).

Porcentaje de lagunas	Nº de años con datos	Nº de meses laguna	Nº de años con lagunas
5 %	66	42	3,5
10 %	63	84	7
15%	59	126	10,5

a) Series que inician sus registros antes de 1950 (sin especificar año de finalización)

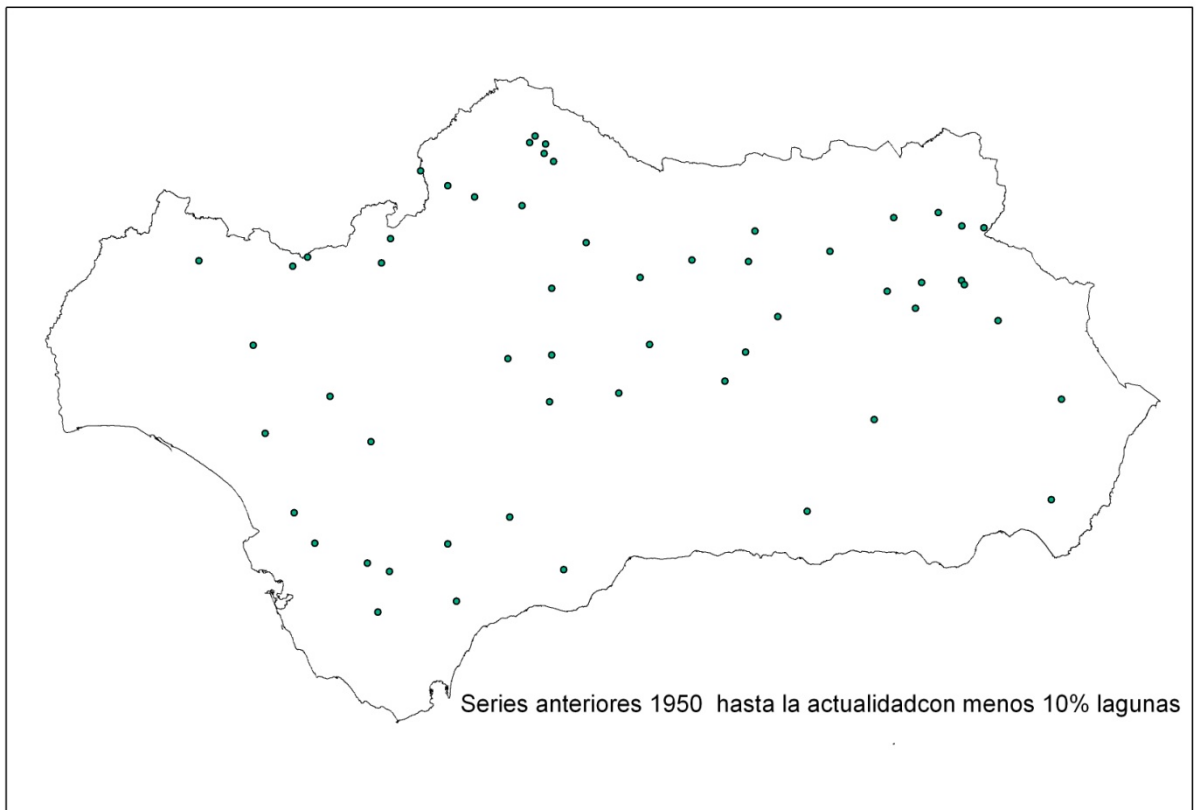
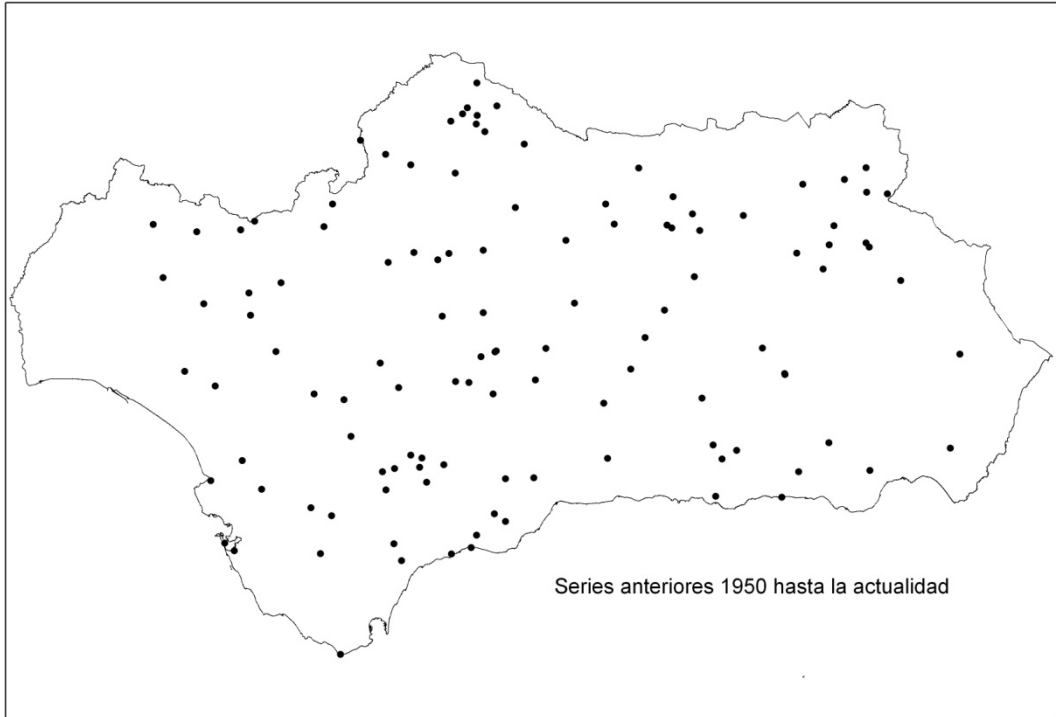
Características de las series	Número de series
Nº total	218
5 % de lagunas	50
10 % de lagunas	104





b) Series que inician sus registros antes de 1950 y se extienden hasta la “actualidad”
(finalizan a partir del año 2000)

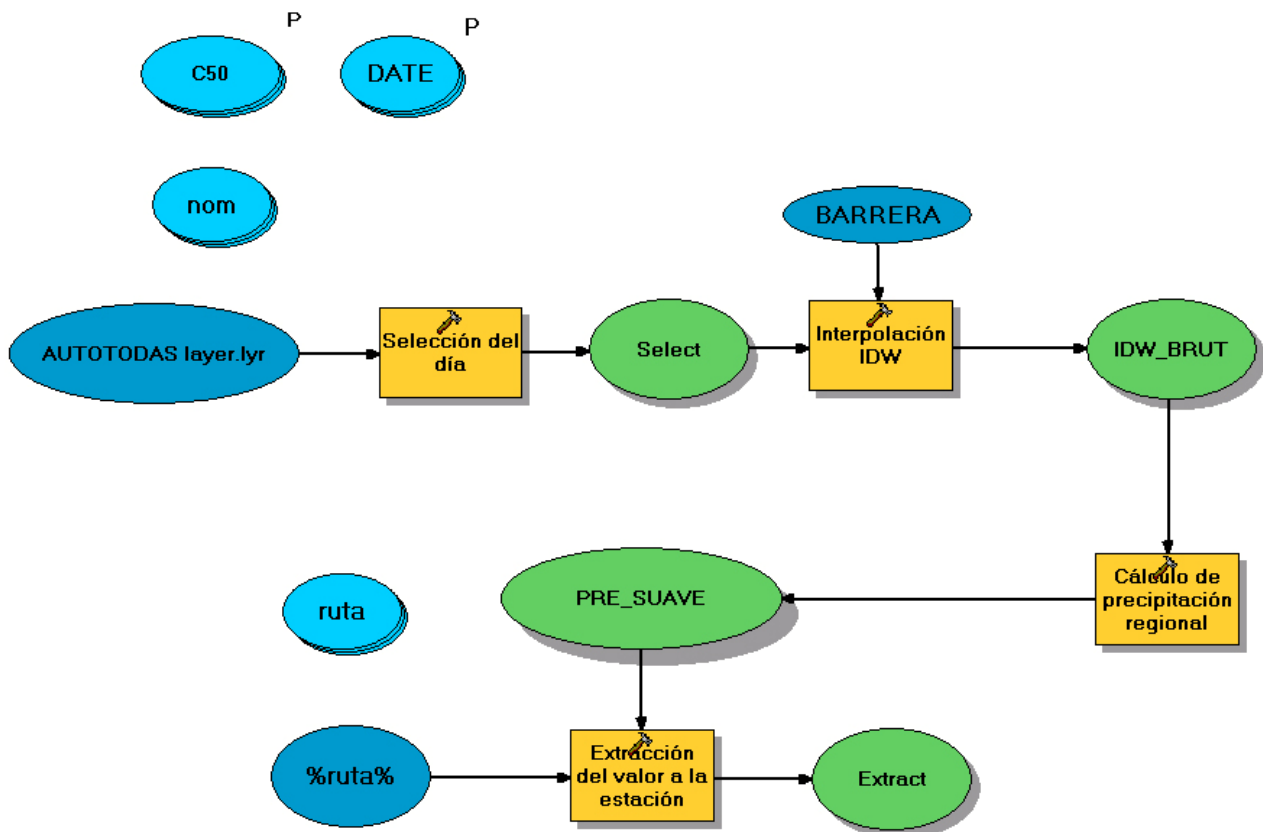
Características de las series	N° series hasta la actualidad
N° total	118
10 % de lagunas	56



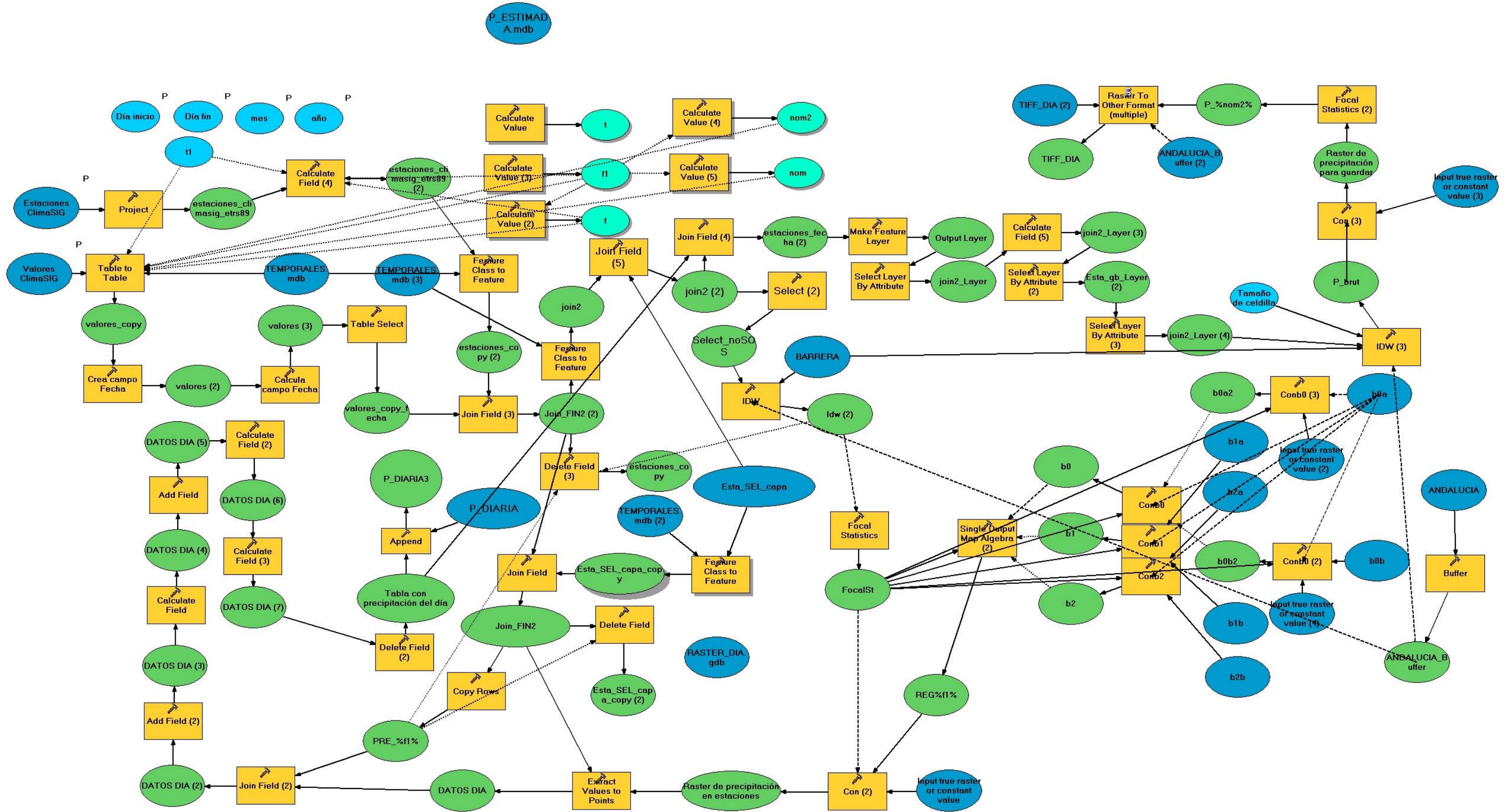
Las series seleccionadas fueron finalmente aquellas 56 anteriores a 1950 con hasta un 10% de lagunas, extendiéndose hasta las 84 finalmente elegidas para 28 estaciones que comenzaban en la década de los años 50 y que tampoco sobrepasaban el 10% de lagunas mensuales acotado.

Las lagunas mensuales fueron interpoladas mediante regresión con la estación vecina mejor correlacionada. Estas series reconstruidas fueron testadas para comprobar su homogeneidad a partir del Test de Alexanderson (Alexanderson y Moberg, 1991) disponible en la herramienta Anclim, no produciéndose ningún rechazo en dichas series por falta de superación del Test.

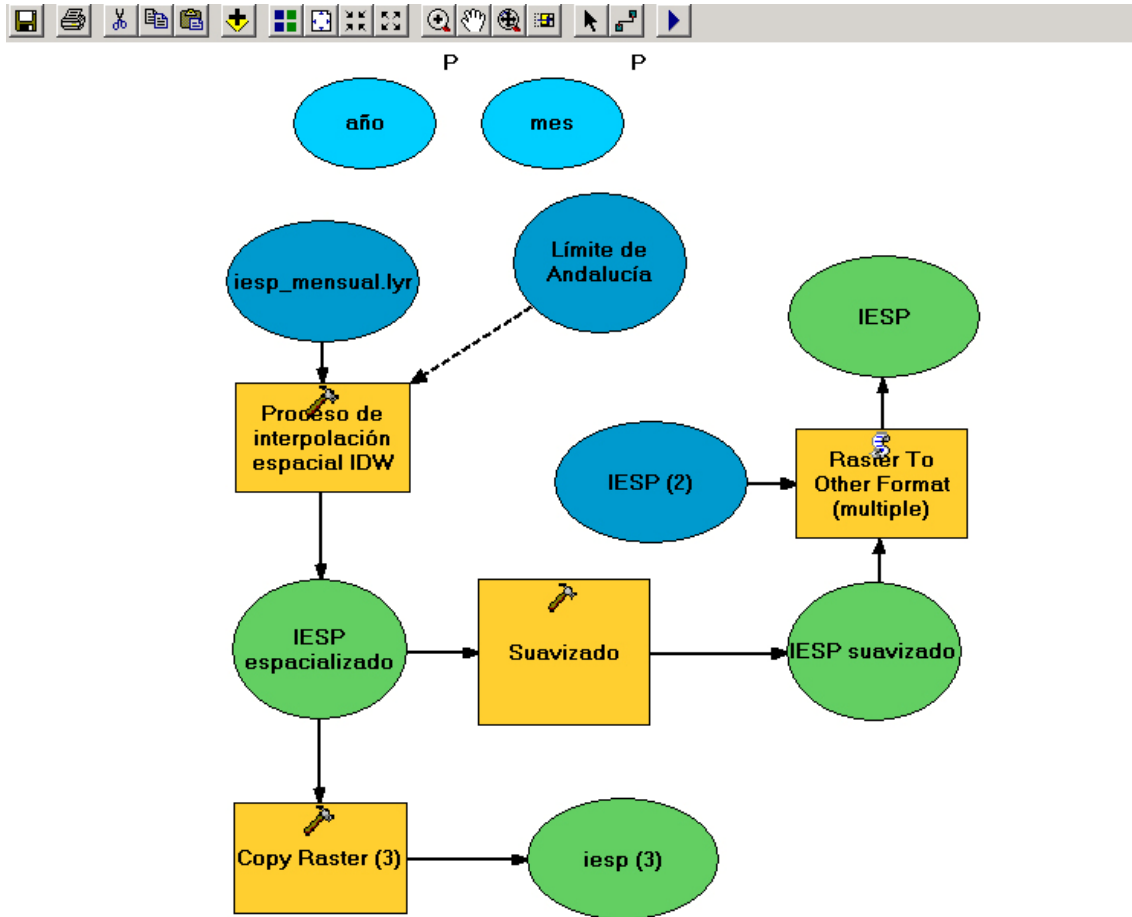
Anexo 2.1. Modelo de procesos para el cálculo de la precipitación regional (300 *84 realizaciones)



Anexo 2.2. Modelo de procesos para la interpolación de precipitación diaria PD_23



Anexo 2.3. Modelo de procesos para la espacialización del IESP mensual



Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta n°1: ***pre_mes_iesp***

```
SELECT a.codigo, Val(Mid(a.fecha,5,2)) AS mes, Val(Mid(a.fecha,1,4)) AS año,
Round(Sum(a.pre_dia),2) AS pre_mes
FROM p_diaria AS a, esta_SEL_capa AS b
WHERE Mid([a.fecha],1,6) Like "*201108*" AND a.codigo=b.codigo and b.IESP=1
GROUP BY a.codigo, Mid(a.fecha,5,2), Mid(a.fecha,1,4)
ORDER BY a.codigo, Val(Mid(a.fecha,1,4)), Val(Mid(a.fecha,5,2));
```

Consulta que agrupa los valores diarios como suma mensual de la precipitación

Consulta n°2: ***link_pre_mes***

```
INSERT INTO t_pre_mes ( codigo, año, mes, pre_mes )
SELECT codigo, año, mes, pre_mes
FROM pre_mes_iesp;
```

Consulta que inserta en la tabla t_pre_mes los valores mensuales calculados en la vista anterior.

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta: ***anomalías***

```
SELECT p.codigo, p."año", p.mes, p.pre_mes - c.mediana AS anomalia
FROM (
SELECT a.codigo AS cod_man, a.mes, a.mediana, b.codigo
FROM t_medianas1961_1990 a
JOIN t_cod_auto_man b ON a.codigo = b.cod_man)
c
JOIN t_pre_mes p ON p.codigo = c.codigo AND p.mes = c.mes
WHERE (p."año" * 100 + p.mes) = ((
SELECT max_mes.max_mes
FROM max_mes));
```

Consulta que calcula las anomalías pluviométricas mensuales para cada estación del último mes de cálculo

Consulta n°4: ***tt_anomalías***

```
SELECT a.codigo, a."año", a.mes, a.anomalia
FROM t_anomalías a
UNION
SELECT b.codigo, b."año", b.mes, b.anomalia
FROM anomalías b;
```

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta reproduce la vista de las anomalías pluviométricas de la serie histórica más las del último mes de cálculo –consulta *anomalías*–.

Consulta n°5: *anom_acum*

```
SELECT a.codigo, a."año", a.mes, a.anomalia, b.acum
```

```
FROM tt_anomalias a
```

```
LEFT JOIN t_anomalias b ON a.codigo = b.codigo AND a."año" = b."año"  
AND a.mes = b.mes;
```

Consulta que “introduce” el campo acum –anomalías pluviométricas acumuladas- de la tabla t_anomalias en la vista de anomalías completa –*tt_anomalías*–. Obviamente el campo acum –anomalías acumuladas- será nulo para el último mes de cálculo.

Consulta n°6: *acum_previo*

```
SELECT anom_acum.codigo, anom_acum."año", anom_acum.mes,  
lag(anom_acum.acum, 1) OVER (ORDER BY anom_acum.codigo, anom_acum."año",  
anom_acum.mes) AS lag
```

```
FROM anom_acum;
```

Consulta que, para cada estación, año y mes selecciona el valor de la anomalía pluviométrica acumulada del registro anterior mediante la función de ventana *lag*

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

–retardo-. Este valor es indispensable para el cálculo del indicador ya que este está basado en una condición sobre dicho valor de anomalía acumulada anterior.

Consulta n°7: ***anomalía_actual_1***

```
SELECT a."año", a.mes, a.anomalia, a.codigo
FROM anom_acum a
JOIN acum_previo b ON a.codigo = b.codigo AND a."año" = b."año" AND a.mes = b.mes
WHERE a.acum IS NULL AND a.anomalia < 0::double precision AND a.anomalia <
0::double precision AND b.lag >= 0::double precision;
```

Consulta que selecciona las estaciones en el último mes de cálculo –acum es nulo- que cumplen la primera condición para el cálculo del IESP: anomalía pluviométrica del mes negativa y anomalía acumulada anterior positiva –serían las estaciones en las que se inicia una secuencia seca-.

Consulta n°8: ***anomalía_actual_2***

```
SELECT a."año", a.mes, a.anomalia, a.codigo, b.lag
FROM anom_acum a
JOIN acum_previo b ON a.codigo = b.codigo AND a."año" = b."año" AND a.mes = b.mes
WHERE a.acum IS NULL AND (a.anomalia < 0::double precision AND b.lag < 0::double
precision OR a.anomalia >= 0::double precision AND b.lag >= 0::double precision OR
a.anomalia >= 0::double precision AND b.lag < 0::double precision);
```

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta que selecciona las estaciones en el último mes de cálculo –acum es nulo- que cumplen la segunda condición para el cálculo del IESP: anomalía pluviométrica del mes negativa y anomalía acumulada en el mes anterior negativa o anomalía pluviométrica positiva y anomalía acumulada anterior positiva también –serían las estaciones que en el mes anterior tienen una anomalía acumulada negativa y que la anomalía del mes de cálculo también es negativa y por tanto se suman o las estaciones en los que dos valores son positivos y por tanto también se adicionan-. En este caso, la consulta incorpora un nuevo campo que es el acumulado del registro anterior como campo de nombre lag.

Consulta n°9: ***acumulados***

```
SELECT b.codigo, b."año", b.mes, b.anomalia, b.anomalia AS acum
      FROM anomalia_actual_1 b
UNION
SELECT c.codigo, c."año", c.mes, c.anomalia, c.anomalia + c.lag AS acum
      FROM anomalia_actual_2 c;
```

Consulta que genera una sola vista para toda la serie disponible con los valores, para cada estación y mes, de la anomalía pluviométrica del mes y de la anomalía acumulada del último mes de cálculo en función de la condición del cálculo del IESP que han sido calculadas en las dos vistas anteriores - ***anomalia_actual_1*** y ***anomalia_actual_2***.

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta n°10: *link_t_anomalias*

```
INSERT INTO t_anomalias ( codigo, año, mes, anomalia, acum )
```

```
SELECT codigo, año, mes, anomalia, acum
```

```
FROM acumulados;
```

Consulta que actualiza la tabla t_anomalias con las anomalías pluviométricas mensuales y las anomalías pluviométricas acumuladas en función de las condiciones que impone el cálculo del IESP del último mes de cálculo.

Consulta n°11: *iesp*

```
SELECT p.codigo, p."año", p.mes, p.anomalia, p.acum AS anom_acum, (p.acum -  
c.media) / c.desviacion AS iesp, p."año" * 100 + p.mes AS fecha_join
```

```
FROM (
```

```
SELECT a.codigo, a.media, a.desviacion
```

```
FROM t_estad_estandar a)
```

```
c
```

```
JOIN t_anomalias p ON p.codigo = c.codigo
```

```
ORDER BY p.codigo, p."año", p.mes;
```

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta básica del Sistema. Muestra la vista de todos los valores del indicador IESP en cada una de las estaciones meteorológicas, desde 1950 hasta el último dato de entrada en el Sistema. Se calcula con los estadísticos media y desviación de la tabla t_estad_estandar y del valor de las anomalías pluviométricas acumuladas (campo anom de la tabla t_anomalias una vez actualizada por la consulta anterior – *link_t_anomalias*).

Consulta n°12: *ultimo_mes_iesp*

```
SELECT DISTINCT last_value(iesp."año") OVER (ORDER BY iesp."año") * 100 +  
last_value(iesp.mes) OVER (ORDER BY iesp."año", iesp.mes) AS last_mes  
  
FROM iesp;
```

Consulta que aplica una función de ventana para conseguir la serie de fechas disponibles para el indicador IESP ordenadas (a partir de la consulta *iesp*).

Consulta n°13: *max_mes_iesp*

```
SELECT max(ultimo_mes_iesp.last_mes) AS max_mes  
  
FROM ultimo_mes_iesp;
```

Consulta que selecciona el valor máximo del campo last_mes de la consulta anterior *ultimo_mes_iesp*, que no es otro de el valor mayor del siguiente cálculo: año * 100 + mes. El máximo valor será por tanto el correspondiente a la última fecha disponible.

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta n°14: *last_iesp*

```
SELECT a.codigo, a."año", a.mes, a.iesp  
FROM iesp a, max_mes_iesp b  
WHERE a.fecha_join = b.max_mes;
```

Consulta que muestra la vista de los valores del IESP del último mes disponible a partir de la vista del iesp anterior –*consulta max_mes_iesp*- y su relación con la fecha mensual máxima o última incorporada al sistema.

Consulta n°15: *iesp_cast*

```
SELECT last_iesp.codigo, last_iesp."año", last_iesp.mes, last_iesp.iesp::numeric AS iesp  
FROM last_iesp;
```

Consulta que lee los valores del indicador del último mes disponible (*last_iesp*) casteados para ser convertidos a valores numéricos. Esta conversión es necesaria para truncar los valores a dos decimales; operación que será realizada en la vista siguiente.

Anexo 3. Definición de consultas (SQL) necesarias para obtener la vista en tiempo real del IESP del último mes de cálculo que operan en el Sistema

Consulta n°16: *iesp_espacial*

```
SELECT a.gid, a.codigo, a.denomi, a.iesp_sos, a.sos, a.the_geom, b."año", b.mes,  
trunc(b.iesp, 2) AS iesp
```

```
FROM esta_sel_capa a
```

```
JOIN iesp_cast b ON a.codigo::bpchar = b.codigo
```

```
WHERE a.iesp_sos = 1::numeric;
```

Consulta que muestra el valor del indicador en el último mes de cálculo disponible asociados a la capa de estaciones meteorológicas de referencia y truncados los valores a dos decimales a efectos de difusión. Es la vista que lee directamente el servidor de mapas *Mapserver* para configurar el servicio WMS que muestra el visor y que, a partir de la opción *get feature info*, muestra el valor del indicador sobre el punto cuando se le realiza la petición.