



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE COMPUTACIÓN

Albero: Visualización de Pronósticos Probabilísticos utilizando Técnicas de Análisis de Análogos

Tesis presentada para optar al título de
Licenciado en Ciencias de la Computación

Leandro Pelorosso

Director: Alexandra Diehl

Codirector: Juan Ruiz

Buenos Aires, 2018

ALBERO

Este trabajo presenta Albero, una solución de análisis visual que expone distintos aspectos de los pronósticos probabilísticos basados en la técnica de regresión por pronósticos análogos (Reforecast Analog Regression, RAR). Esta técnica comprende el análisis de grandes cantidades de pronósticos meteorológicos generados con un modelo numérico específico y sus observaciones asociadas, con el fin de obtener una mejor estimación de la incertidumbre de un pronóstico actual, y de los sesgos y errores sistemáticos del modelo. Albero permite a los pronosticadores el estudio de pronósticos probabilísticos, los análogos y observaciones utilizados en la generación de los mismos, sus errores e información estadística adicional por medio de vistas coordinadas. Albero presenta una análisis visual compuesto de tres ciclos de proceso anidados. El primer bucle se compone de la parametrización y el refinamiento progresivo de la técnica. El segundo bucle proporciona un resumen interactivo e iterativo de los pronósticos probabilísticos. El tercer y último bucle permite un análisis detallado de los análogos y observaciones utilizadas, y propiedades estadísticas asociadas con la técnica. Hemos diseñado nuestra solución siguiendo una metodología participativa junto con expertos del dominio. Varios meteorólogos de diferentes orígenes validaron nuestro enfoque. Dos casos de estudio ilustran las capacidades de la solución presentada. Albero efectivamente facilita el análisis de incertidumbre y sesgos sistemáticos del modelo de predicción meteorológica para mejorar la toma de decisiones y la optimización de la técnica.

Palabras claves: Reforecast, RAR, Incertidumbre, Pronósticos Probabilísticos, Visual Analytics, Optimización, Toma de decisiones.

Índice general

1.. INTRODUCCIÓN	1
1.1. Visual Analytics	1
1.1.1. Contribuciones de la presente tesis	3
1.2. Reforecast Analog Regression	4
2.. TRABAJOS PREVIOS	8
2.1. Visualización meteorológica y climatología	8
2.2. Visualización de la incertidumbre	9
2.3. Herramientas disponibles	9
2.3.1. ViSUS	9
2.3.2. UV-CDAT	10
2.3.3. ParaView	10
2.3.4. VisIt	11
2.3.5. COVISE	11
2.3.6. Ensemble-VIS	11
2.3.7. VTK	12
2.3.8. ArcGIS	12
2.3.9. AVS/Express	13
2.3.10. GrADS	13
2.3.11. Integrated Data Viewer (IDV)	14
2.3.12. VisTrails	14
3.. REFORECAST ANALOG REGRESSION (RAR)	21
3.1. Introducción	21
3.2. Descripción del método	21
3.2.1. Paso 1: Selección de análogos	21
3.2.2. Paso 2. Construcción del pronóstico probabilístico	22
3.2.3. Paso 3. Smoothing	23
4.. ALBERO	28
4.1. Descripción general	28
4.2. Fuentes de datos	28
4.3. Workflow de tres ciclos	30
4.3.1. Ciclo de Parametrización	30
4.3.2. Ciclo de Pronóstico Probabilístico	31
4.3.3. Ciclo de Análogos	33
4.4. Proceso de Diseño	33
4.4.1. Fase de Incepción	34
4.4.2. Fase de Diseño	35
4.5. Implementación	36
4.5.1. Arquitectura	37
4.5.2. Detalles de Implementación	43
4.5.3. BingMaps Tiling	49

4.5.4. Manipulación de Datos Observacionales	50
5.. CASOS DE ESTUDIO	52
5.1. Albero para pronóstico operacional	52
5.2. Albero para optimización de la técnica	53
6.. LECCIONES APRENDIDAS Y CONCLUSIONES	56
6.1. Lecciones Aprendidas	56
6.1.1. Proceso de Diseño	56
6.2. Conclusiones	57

1. INTRODUCCIÓN

1.1. Visual Analytics

Cook et al. [1] define Visual Analytics como la ciencia del razonamiento analítico soportado por interfaces visuales interactivas. En la actualidad se producen y almacenan datos a un ritmo y cantidad que sobrepasa la capacidad de analizarlos. En este contexto, la visualización de datos es una herramienta poderosa para resumir, extraer, y resaltar los aspectos importantes u ocultos en los datos.

En las últimas décadas se han desarrollado una gran cantidad de métodos automáticos de análisis de datos. Sin embargo, la naturaleza compleja de muchos problemas hace que sea indispensable incluir la inteligencia humana en una etapa temprana del proceso de análisis. Los métodos de Visual Analytics permiten a los responsables de la toma de decisiones combinar su creatividad y conocimiento previo con las enormes capacidades de almacenamiento y procesamiento de las computadoras de hoy en día, para obtener información sobre problemas complejos. Utilizando interfaces visuales avanzadas, los humanos pueden interactuar directamente con las capacidades de análisis de datos de la computadora actual, lo que les permite tomar decisiones bien informadas en situaciones complejas.

Keim et al. [2] ofrece una definición mas específica: Visual Analytics combina técnicas de análisis automatizadas con visualizaciones interactivas, para una comprensión, razonamiento y toma de decisiones efectiva sobre conjuntos de datos muy grandes y complejos. El objetivo del análisis visual es la creación de herramientas y técnicas para permitir al usuario:

- Sintetizar la información y extraer conocimiento de grandes conjuntos dinámicos de datos, a menudo ambiguos y contradictorios,
- Detectar lo esperado y descubrir lo inesperado.
- Proporcionar evaluaciones oportunas, defendibles y comprensibles.
- Comunicar la evaluación de manera efectiva para facilitar la toma de decisiones.

A gran escala, las soluciones de Visual Analytics proporcionan tecnología que combina las fortalezas del procesamiento de datos humano con el de los dispositivos electrónicos. La visualización se vuelve el medio sobre el que se desarrolla un proceso analítico semiautomatizado, donde humanos y máquinas pueden cooperar utilizando sus respectivas capacidades, para obtener los mejores resultados posibles. El usuario es quien guía la dirección de análisis, mientras que el sistema debe proporcionar medios efectivos de interacción que faciliten dicho análisis.

El proceso de Visual Analytics combina métodos de análisis automáticos y visuales con un acoplamiento estricto a través de la interacción humana para obtener conocimiento. La figura Figura 1.1 muestra un resumen general de las diferentes etapas (representadas mediante óvalos) y sus transiciones (flechas) en el proceso de Visual Analytics.

En muchas aplicaciones, las fuentes de datos heterogéneas deben integrarse antes de ser posible un análisis visual o automático. Por lo tanto, el primer paso suele consistir en pre

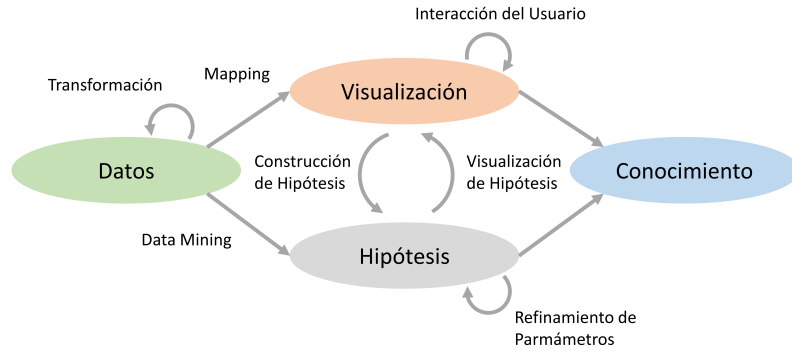


Fig. 1.1: Diagrama de proceso de Visual Analytics. Adaptado de *Visual Analytics: Scope and Challenges*. [2]

procesar y transformar los datos para derivar diferentes representaciones que faciliten una mayor exploración (como lo indica la flecha de transformación en la figura). Otras tareas típicas de pre procesamiento incluyen la limpieza de datos, la normalización, la agrupación o la integración de fuentes de datos heterogéneas. Después de la transformación, el analista puede elegir entre aplicar métodos de análisis visuales o automáticos.

Si se utiliza primero un análisis automático, se aplican métodos de análisis automáticos para generar una hipótesis. Una vez obtenida, el analista debe evaluar y refinar dicha hipótesis. Múltiples visualizaciones permiten a los analistas interactuar con los métodos automáticos, ya sea modificando parámetros o seleccionando otros algoritmos de análisis. Estas visualizaciones facilitan la evaluación de la hipótesis generada. La alternancia entre los métodos visuales y automáticos es parte del proceso de Visual Analytics y conduce a un refinamiento continuo y a la verificación de los resultados preliminares. Los resultados confusos en un paso intermedio pueden así descubrirse en una etapa temprana, dando lugar a mejores resultados y una mayor confianza en el análisis. Si se realiza primero una exploración de datos visuales, el usuario debe confirmar las hipótesis generadas mediante un análisis automático. La interacción del usuario con la visualización es necesaria para revelar información útil, por ejemplo, amplificando en diferentes áreas de datos o considerando diferentes visualizaciones de los mismos. Los hallazgos en las visualizaciones pueden ser usados para orientar la construcción de una hipótesis en el análisis automático.

Keim et al. [2] resume el proceso de Visual Analytics mediante el siguiente mantra (en el idioma publicado):

*“Analyse First -
Show the Important -
Zoom, Filter and Analyse Further -
Details on Demand”*

Sacha et al. [3] establece que los modelos previamente desarrollados son guiados por el sistema, y no describen en detalle el papel del razonamiento humano en el proceso de visual analytics. Los autores plantean la necesidad de que los investigadores posean un framework integrado de todos los procesos y modelos relevantes para la generación de conocimiento con visual analytics. Intentan, además, dar los primeros pasos en la definición de un modelo de generación de conocimiento con visual analytics, expandiendo los modelos existentes de Keim [2].

El modelo propuesto se compone de dos partes: el sistema computacional con datos, visualizaciones y modelos analíticos, y el componente humano modelando el proceso cognitivo asociado a una sesión analítica. En este modelo (Figura 1.2) no existe una clara separación entre la parte humana y la parte computacional, dado que ambas son necesarias para el análisis de los datos. La generación de conocimiento en visual analytics se compone de procesos de razonamiento abductivo, deductivo e inductivo [4] [5]. El proceso de razonamiento abductivo formula hipótesis a partir de observaciones inesperadas, o que no pueden ser explicadas por medio del conocimiento existente. Asumiendo que estas hipótesis son verdaderas, comportamientos y expectativas de efectos, patrones o relaciones en los datos analizados son deducidos. A través de interacciones con el sistema de visual analytics, el investigador intenta encontrar evidencia y detectar patrones en los datos, con el fin de verificar o invalidar las hipótesis, lo que compone el proceso de razonamiento inductivo. El modelo humano cognitivo es representado por medio de ciclos, dado que el análisis no sigue reglas determinísticas, sino que posee una naturaleza caótica y espontánea.

El primer ciclo, el ciclo de exploración, describe como los investigadores interactúan con el sistema de visual analytics con el fin de generar nuevas visualizaciones o modelos, y analizar los datos. El investigador explora los datos interactuando con el sistema y observando los resultados. Las acciones llevadas a cabo en el ciclo de exploración dependen de los descubrimientos, o de un objetivo de análisis concreto. En caso de que no se cuente con un objetivo particular, el ciclo de exploración se convierte en una búsqueda de descubrimientos, que puede llevar a un objetivo de análisis concreto.

El ciclo de verificación guía la exploración para confirmar hipótesis o formular nuevas. Para verificar una hipótesis concreta, un análisis de confirmación es llevado a cabo, y el ciclo de exploración es guiado para revelar descubrimientos que verifiquen o invaliden la hipótesis. Los investigadores obtienen conocimiento en la medida en que logren interpretar dichos descubrimientos, en el contexto del dominio del problema. Estos descubrimientos pueden llevar a nuevas hipótesis que requieran investigación adicional.

En el último ciclo, el ciclo de generación de conocimiento, el investigador forma hipótesis a partir de su conocimiento acerca de un dominio del problema, y obtiene nuevo conocimiento formulando y verificando hipótesis durante el proceso de visual analytics. Cuando el investigador confía en los resultados obtenidos, nuevo conocimiento del dominio del problema es obtenido, que puede a su vez influenciar la formulación de nuevas hipótesis en el siguiente proceso de análisis.

1.1.1. Contribuciones de la presente tesis

Como sucede habitualmente en el campo de la visualización de la información, las contribuciones presentadas en esta tesis son el resultado de colaboraciones entre expertos de múltiples disciplinas.

El autor de esta tesis es co-autor de las publicaciones P1 a P3, llevando a cabo el desarrollo completo del software de P3, y contribuyendo partes sustanciales del diseño, investigación, y generación de resultados.

La Dr. Alexandra Diehl, autora de las publicaciones P1 a P3, realizó invaluable aportes en el diseño y workflow de trabajo. Dirigió las reuniones con los expertos del dominio, permitiendo llevar a cabo un proceso de construcción iterativo e incremental. Sus conocimientos en el área de la visualización de la información resultaron imprescindibles para la creación de este trabajo.

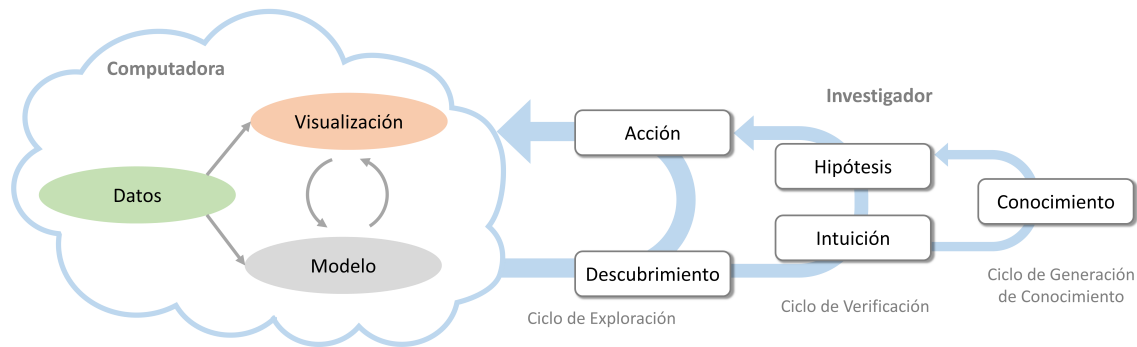


Fig. 1.2: Modelo de generación de conocimiento. Adaptado de *Knowledge Generation Model for Visual Analytics*. [3]

El Dr. Juan Ruiz, co-autor de las publicaciones P1 a P3, principal stakeholder y experto del dominio, contribuyó incansablemente en la dirección del rumbo del proyecto, estableciendo requerimientos y necesidades, evaluando prototipos, y compartiendo didácticamente su conocimiento en el área de pronósticos probabilísticos y por ensambles.

Prof. Meister Gröller y Prof. Stefan Bruckner contribuyeron con sus aportes durante todo el ciclo de desarrollo, desde la inepción hasta la evaluación de Albero. Avalaron el diseño de las herramientas con ideas y sugerencias. Son co-autores de las publicaciones P1 a P3.

La evaluación de Albero fue realizada por expertos del dominio del Centro de Investigaciones del Mar y la Atmósfera (CIMA), y el Servicio Meteorológico Nacional: Dr. Juan Ruiz, Dr. Yanina Garcia Skabar, Laura Aldeco y Cynthia Matsudo.

- P1: 2015 A. Diehl, L. Pelorosso, K. Matković, C. Delrieux, J. Ruiz, M. E. Gröller, S. Bruckner. “Albero: A Visual Analytics Tool for Probabilistic Weather Forecasting”. Poster presented at the workshop “Big Data and Environment: Seizing the Data Deluge in Environmental Sciences”. Buenos Aires, Argentina, November 2015.
- P2: 2016 L. Pelorosso, A. Diehl, K. Matković, C. Delrieux, J. Ruiz, M. E. Gröller, S. Bruckner. “Visualizing Uncertainty for Probabilistic Weather Forecasting based on Reforecast Analogs”. Poster accepted for the EGU General Assembly 2016. Session: “Information in earth sciences: visualization techniques and communication of uncertainty”.
- P3: 2017 A. Diehl, L. Pelorosso, C. Delrieux, K. Matković, J. Ruiz, M. E. Gröller, S. Bruckner. “Albero: A Visual Analytics Tool for Probabilistic Weather Forecasting”. Computer Graphics Forum. 36(7), pp. 135-144. ISSN 0167-7055. eISSN 1467-8659. Available under: doi: 10.1111/cgf.13279

1.2. Reforecast Analog Regression

La generación de pronósticos meteorológicos involucra datos de fuentes de complejidad heterogénea y distinta resolución espacial y temporal, requiriendo múltiples niveles de abstracción para facilitar su análisis.

El acceso inmediato a la información permite a los pronosticadores tomar decisiones rápidamente, las cuales deben ser eficientemente comunicadas en tiempo, con el objetivo de prevenir posibles desastres, salvando vidas y pérdidas económicas.

Dada una representación adecuada del estado atmosférico en un tiempo determinado, su futura evolución puede ser estimada utilizando modelos de predicción numérica (NWP). En un modelo de pronóstico determinístico, una única corrida del modelo es realizada utilizando las mejores condiciones iniciales conocidas, para obtener un único resultado [6]. En este caso, el pronóstico no provee información alguna de la incertidumbre de la predicción.

A diferencia del modelo de predicción determinístico, los pronósticos probabilísticos reconocen que los datos utilizados para correr el NWP son imprecisos, que el modelo es una aproximación, y que existen variables adicionales que no están siendo consideradas. Un modelo probabilístico resume múltiples pronósticos determinísticos, utilizando distintos métodos e información para calcular la probabilidad de ocurrencia de un evento. Este tipo de modelo permite la evaluación de la incertidumbre. Los modelos de predicción por ensambles son modelos probabilísticos que corren uno o más modelos NWP múltiples veces, utilizando ligeras variaciones de las condiciones iniciales para obtener un ensamble o colección de pronósticos. El ensamble generado representa únicamente algunos de los escenarios posibles. Los pronósticos por ensamble usualmente subestiman el grado de incertidumbre en la predicción de los futuros estados de la atmósfera [7] [8], debido a observaciones imprecisas, a nuestras limitaciones al modelar procesos físicos de la atmósfera, y a la naturaleza caótica de la misma.

Hamill y Whitaker [9] introdujeron la técnica de predicción por pronósticos retrospectivos (Reforecast Analog Regression, RAR) para pronósticos probabilísticos, con el objetivo de superar algunas limitaciones de los pronósticos determinísticos y por ensamble. El algoritmo propuesto utiliza una extensa base de datos de pronósticos meteorológicos generados utilizando un modelo NWP específico [10], y las observaciones asociadas [11].

La técnica RAR consiste en un procedimiento de tres pasos. El primer paso consiste en generar pronósticos para los próximos días, utilizando un modelo numérico específico.

En un segundo paso se buscan pronósticos históricos similares al actual, generados por el mismo modelo de predicción numérica, para una determinada región geográfica. Se utiliza una función para evaluar la similitud, tal como el error cuadrático medio de la diferencia. Estos pronósticos similares son conocidos como “análogos”.

Como tercer y último paso, el método hace uso de las observaciones correspondientes a las fechas análogas, para estimar la función de densidad de probabilidad de la variable meteorológica para el periodo del pronóstico actual, con el fin de obtener un pronóstico cuyo error sistemático sea menor al del modelo numérico utilizado.

Este procedimiento de dos pasos compensa los sesgos generados por el modelo numérico. La información se puede presentar entonces de varias formas: como probabilidades superando límites concretos, como valores extremos, como la media o la mayoría de los resultados probables, etc. Los errores sistemáticos del modelo pueden también ser obtenidos comparando los pronósticos análogos y sus correspondientes observaciones. Hamill y Whitaker proponen diversas variantes de la técnica. Este trabajo utiliza la técnica “Smoothed Rank Analog”. Aldeco et al. [12] verifica experimentalmente el rendimiento de la técnica para Argentina, utilizando datos observacionales obtenidos de estaciones meteorológicas locales.

Sin embargo, conjuntos de datos con propiedades estadísticas casi idénticas pueden

presentar comportamientos muy diferentes, como ha sido ilustrado por Anscombe [13]. La técnica RAR puede proveer pronósticos probabilísticos similares, frente a conjuntos dispares de ensambles de análogos y sus observaciones.

En las visualizaciones existentes, se le ofrece al pronosticador únicamente la probabilidad obtenida por el método de análogos. Una pregunta surge de manera inmediata: ¿Es esta toda la información que puede ser aprovechada por el usuario, o puede el método permitir un análisis más profundo de los datos?. Para responder esta pregunta, es necesario ir mas allá de las visualizaciones tradicionales, y desarrollar herramientas que permitan al usuario explorar en detalle toda la información provista por el método.

El estudio de esta información permite a los investigadores expertos optimizar la técnica RAR y a los pronosticadores mejorar sus procesos de toma de decisiones. Un workflow visual interactivo puede ser utilizado en cada paso de la técnica de RAR para visualizar los resultados intermedios, descubriendo así información que de otra forma permanecería oculta.

Otro aspecto importante es que los meteorólogos necesitan herramientas eficaces que les permita el acceso inmediato a los datos, y realizar un rápido análisis de los resultados. Necesitan herramientas que les permita tomar decisiones en un corto tiempo. Podemos sustituir el workflow automático de Hamill y Whitaker (véase Figura 1.3) utilizado por nuestros expertos con un nuevo workflow interactivo (véase Figura 1.4), que permite a los pronosticadores obtener nuevos conocimientos sobre los pronósticos probabilísticos, eventos pasados, la incertidumbre y los errores sistemáticos del modelo.

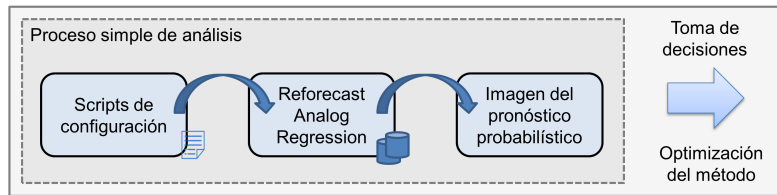


Fig. 1.3: Flujo de trabajo automático actual. Proporciona una flexibilidad reducida, y la necesidad de generar múltiples imágenes de manera individual.

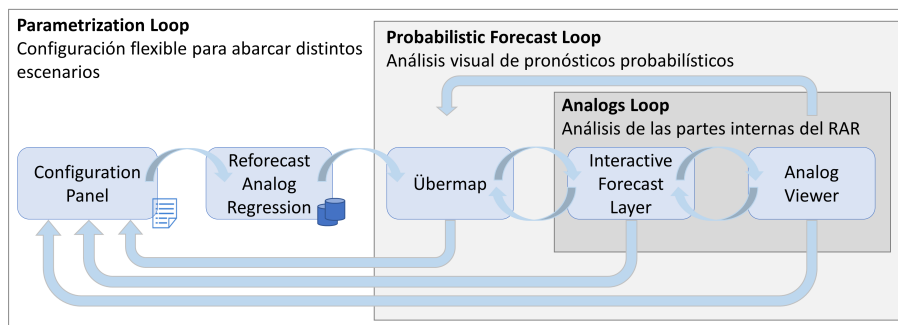


Fig. 1.4: Nuevo flujo de trabajo interactivo. Proporciona tres ciclos de análisis anidados para transmitir conocimientos sobre los pronósticos probabilísticos y la técnica RAR, para permitir una mejor toma de decisiones y la optimización de la técnica.

Nuestro workflow se compone de tres bucles principales, como muestra la Figura 1.4. Cada bucle consta de un proceso visual interactivo e iterativo entre el usuario y diferentes

vistas enlazadas, que actúan como una interfaz para los métodos computacionales subyacentes. Aunque las vistas enlazadas utilizan técnicas de visualización bien conocidas, su combinación para asistir a los meteorólogos en el análisis de la técnica RAR compone un enfoque novedoso. Este trabajo presenta dos principales contribuciones:

- Presentamos una solución de visualización integrada en el contexto de la técnica RAR. Las herramientas meteorológicas existentes no permiten un completo análisis visual, y requieren conocimientos de programación y configuraciones complejas. Al momento de la escritura de esta tesis, y según nuestro mejor conocimiento, esta es la primera solución de análisis visual que proporciona un análisis visual interactivo de los pronósticos probabilísticos basadas en la técnica RAR, es decir, más allá de las capacidades de las herramientas existentes para la predicción meteorológica. Nuestra solución, “Albero”, expone los aspectos internos de la técnica, permitiendo a los meteorólogos analizar no sólo los resultados de la misma, que son los pronósticos probabilísticos, sino también sus análogos asociados, observaciones e información estadística.
- Introducimos un nuevo flujo de trabajo interactivo que engloba tres bucles anidados. El primer bucle, el bucle de parametrización, permite a los usuarios definir parámetros temporales y de configuración del algoritmo RAR. En el segundo bucle interno, el bucle de pronóstico probabilístico, los usuarios pueden analizar los pronósticos probabilísticos con diferentes niveles de detalle y compararlas tanto temporal como espacialmente. Se resume la información técnica acerca de las observaciones, los errores y la incertidumbre. Este bucle permite a los usuarios detectar situaciones extremas en análogos y observaciones, tales como altas precipitaciones. El tercer bucle anidado, el bucle de análogos, permite a los usuarios analizar detalles del desempeño de la técnica RAR. Proporciona una vista de desglose que despliega información sobre conjuntos de pronósticos y observaciones para los análogos seleccionados, fechas y configuración del algoritmo. El bucle de análogos permite a los usuarios analizar la precisión esperada de la técnica para un evento determinado, y de acuerdo a los resultados obtenidos, analizar las posibles mejoras y realizar posteriores validaciones objetivas. Nuestro proceso cíclico de análisis visual proporciona a los expertos del dominio nuevas perspectivas sobre la técnica de regresión de pronósticos análogos, para mejorar los procesos de toma de decisiones y la consecuente preparación de acciones preventivas.

2. TRABAJOS PREVIOS

Múltiples herramientas de visualización han sido ampliamente utilizadas para el análisis de pronósticos y datos climatológicos. El análisis visual resulta una herramienta fundamental para los meteorólogos. En particular, el análisis visual de patrones espaciotemporales les ayuda en la comprensión de determinados fenómenos climáticos, comportamientos atípicos del modelo, y errores. Podemos clasificar los trabajos anteriores basados en diferentes aspectos que son relevantes para nuestro trabajo.

2.1. Visualización meteorológica y climatología

Se ha realizado una amplia labor en el área de sistemas de visualización analítica aplicada a la predicción meteorológica y la climatología.

En el marco de este trabajo, sólo los ejemplos más representativos se resumen como referencias a nuestro enfoque. La University Corporation for Atmospheric Research (UCAR) presentó una lista de herramientas de post procesamiento para la visualización meteorológica. Entre ellas se encuentran: Integrated Data Viewer (IDV) [14], Visualization and Analysis Platform for Ocean, Atmosphere, and

Solar Researchers (VAPOR) [15], y Grid Analysis and Display System (GrADS) [16]. VAPOR proporciona visualizaciones 3D interactivas de conjuntos masivos de datos, con foco en necesidades científicas. VAPOR permite el análisis visual de datos grillados y se puede ejecutar en computadoras equipadas con un GPU avanzado.

GrADS es un sistema de visualización que utiliza una mezcla de interfaz gráfica de usuario (GUI) y secuencias de líneas de comando para obtener visualizaciones de los datos procesados. La habilidad de aceptar comandos de comandos y características de programación puede ser visto como una ventaja en términos de flexibilidad, pero a su vez dificultan la interacción y experiencia del usuario. En su lugar, las interfaces visuales han demostrado ser más eficaces en términos de productividad cognitiva [17].

El Ultrascalse Visualization Climate Data Analysis Tools (UVCDAT) es un poderoso sistema desarrollado conjuntamente por varias instituciones, universidades y empresas privadas [18]. Integra un conjunto de herramientas de análisis de datos, análisis de ensamble, incertidumbre, métricas de cuantificación, y análisis visual aplicado a grandes volúmenes de datos climáticos. Asimismo, Song et al. [19] presenta un sistema de visualización que realiza el análisis de conjuntos de datos utilizando datos atmosféricos multidimensionales, visualizaciones, y representación de partículas. Los sistemas de visualización mencionados permiten un análisis visual completo y flexible, pero requieren ajustes de parámetros complejos y, en algunos casos, conocimientos de programación. En adición, dichos sistemas no proveen una interfaz web, lo que obliga al usuario a instalar y configurar localmente las aplicaciones, tareas que requieren, en general, conocimientos del sistema operativo utilizado.

Estamos especialmente interesados en sistemas de visualización disponibles como aplicaciones web. Microsoft Research FetchClimate [20] es un servicio de recuperación de datos climáticos que opera en la nube. Expone un webmap integrado con Bing Maps, varias vistas vinculadas y menús contextuales para visualizar e interactuar con los datos. Esta aplicación se centra en la climatología en lugar de la predicción meteorológica. Asi-

mismo, WeatherSpark Beta [21] proporciona múltiples vistas, como un sistema integrado de mapview, gráficos y glifos que representan las condiciones climáticas (soleado, nublado, lluvioso, etc.), y los datos históricos para una determinada estación meteorológica o una determinada ciudad en el mundo. Otros ejemplos son Weather.com [22], Wetter.de [23], Weather Underground [24], y el sistema de visualización meteorológica en 3D Terra3D [25].

Estos casos están orientados principalmente a la población general, en lugar de los expertos del dominio. En cuanto a las herramientas mencionadas, nuestros expertos del dominio declararon: “Las aplicaciones web disponibles no nos permiten realizar operaciones entre los pronósticos, no nos proporcionan una visión global y no nos permiten comparar múltiples pronósticos.”

Con respecto a los reforecast, una de las mayores limitaciones en las soluciones existentes (por ejemplo las implementadas por el Servicio Meteorológico Nacional), es que no se puede determinar el umbral, el plazo y el periodo de acumulación de manera arbitraria. Por otra parte, los gráficos estáticos no permiten el análisis en mayor profundidad un área en particular. Además, sólo hemos encontrado unas pocas opciones en estas aplicaciones para la visualización comparativa entre diferentes timesteps y/o capas.

2.2. Visualización de la incertidumbre

Potter et al. [26] describe un amplio panorama del estado del arte de la visualización de la incertidumbre. Describe una taxonomía de los diferentes enfoques para la visualización de la incertidumbre basado en las dimensiones de los datos y las dimensiones de la incertidumbre. Thomson et al. [27] propuso una tipología de incertidumbre que relaciona los datos, la incertidumbre y las tareas, con su adecuada representación visual. Numerosos trabajos de investigación están disponibles para los meteorólogos con el fin comprender y comunicar la incertidumbre de ensambles de pronósticos. Deitrick y Edsall [28] experimentalmente concluyeron que la incertidumbre de la visualización puede afectar el proceso de toma de decisiones. Potter et al. [29] presentó Ensemble-Vis, un marco que combina múltiples vistas vinculadas para facilitar el análisis visual de un conjunto de datos, centrado en la predicción meteorológica a corto plazo (cortos lapsos de tiempo) y la modelización del clima (lapsos de tiempo muy largos). Proporciona vistas espaciales y temporales globales (overviews) combinadas con vistas estadísticas detalladas. Estas visualizaciones dan una idea sobre la distribución de la incertidumbre de los modelos de pronósticos, y enfatizan las características probabilísticas del ensamble. Sanyal et al. [30] introdujo Noodles, un marco en el que se combina la visualización de la incertidumbre basada en cintas (ribbon) y glifos, spaghetti plots y mapas de color iso-presión para evaluar la incertidumbre en los datos. Su enfoque ha demostrado ser especialmente útil para encontrar datos anómalos en pronóstico por ensambles.

2.3. Herramientas disponibles

A continuación se describe brevemente las herramientas existentes analizadas.

2.3.1. ViSUS

ViSUS implementa técnicas de streaming de datos para proceso y visualización progresivo de grandes cantidades de datos científicos. Esto permite tres principales modalidades

de visualización. (i) Visualización interactiva en power-walls (display de muy alta resolución, construidas a partir de una matriz de pantallas). (ii) Visualización interactiva, en terminales personales, de grandes conjuntos de datos que podrían ser almacenados de forma local. (iii) Monitoreo inmediato de estaciones remotos desde terminales personales. Estas modalidades apuntan a varias fases del proceso de generación y exploración de grandes simulaciones de datos donde la interacción en tiempo real puede incrementar la productividad del científico. ViSUS portable visualization framework (Appkit) puede ser compilado como un plugin para Google Chrome, Microsoft Internet Explorer o Mozilla Firefox. Esto permite al framework ViSUS ser fácilmente integrado en aplicaciones Web.

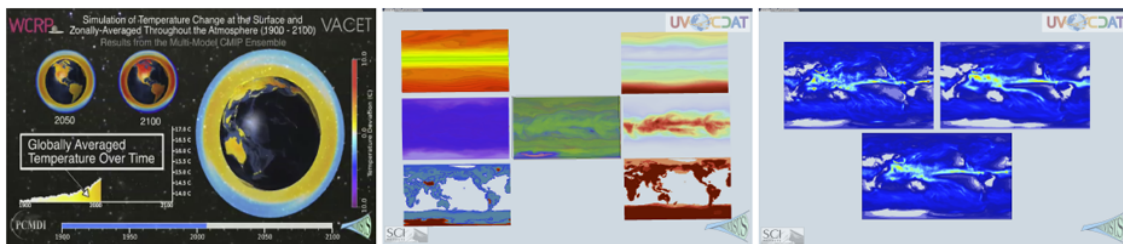


Fig. 2.1: Visualización del clima con ViSUS. Izquierda: el framework provee una visualización del ensamble de simulación de cambio de temperatura. Centro: dos conjuntos de datos de diferente resolución espacial mostrando la precipitación a nivel global. Derecha: simulación del clima mostrando las temperaturas globales y precipitación utilizando dieciséis modelos diferentes, fundidos utilizando una función de desviación standard definida por el usuario.

2.3.2. UV-CDAT

Integra múltiples paquetes de visualización open source ampliamente utilizados: CDAT, VTK, R, VisTrails, VisIt, ParaView, y DV3D. Entre las características mas importantes, se destacan: (i) Visualización y acceso a datos local y remoto. (ii) Visualización comparativa y análisis estadístico. (iii) Herramientas robustas de proyección y agregación de datos. (iv) Soporte para datos observacionales no grillados, incluyendo formatos comúnmente usados para conjuntos de datos observacionales.

2.3.3. ParaView

ParaView es una aplicación open source y multiplataforma para el análisis y visualización de datos. La aplicación permite la rápida construcción de visualizaciones para el análisis de datos utilizando técnicas cualitativas y cuantitativas. La exploración de los datos puede ser realizada de forma interactiva de manera 3D, o programáticamente. ParaView fue desarrollado para permitir el análisis de conjuntos de datos extremadamente grandes, utilizando recursos de memoria distribuida. Puede ser ejecutado en super computadoras, o en computadoras de escritorio cuando el conjunto de datos es relativamente pequeño. ParaView se ha convertido en una herramienta integral en muchos laboratorios, universidades, o complejos industriales, y ha sido adjudicado con premios relacionados a la alta performance computacional.

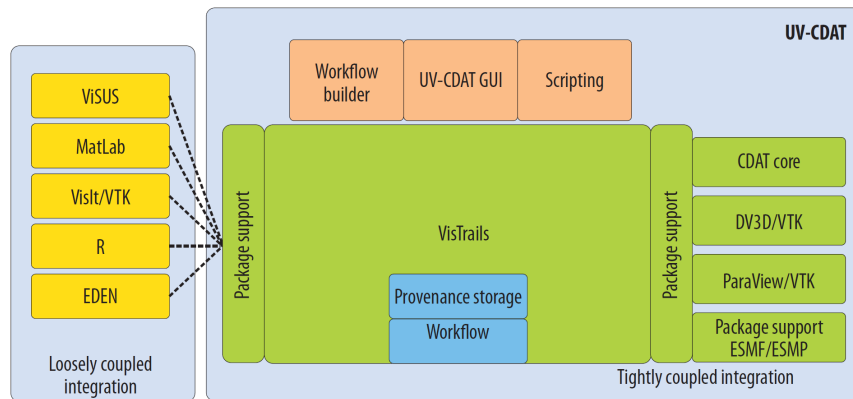


Fig. 2.2: El framework Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT). Uniendo múltiples protocolos y APIs, el framework permite a los científicos visualizar e interpretar datos climatológicos y observaciones en múltiples modelos.

2.3.4. VisIt

VisIt es una herramienta open source, interactiva y escalable de visualización, animación, y análisis de datos. Compatible con Unix, Linux, Windows y MacOs, permite a los usuarios generar visualizaciones, animarlas en el tiempo, manipularlas con una amplia gama de operadores y expresiones matemáticas, y guardar las imágenes y animaciones resultantes. VisIt permite el análisis y visualización de datos escalares y vectoriales, definidos en estructuras en dos y tres dimensiones, y datos no grillados.

2.3.5. COVISE

COVISE es un entorno de software distribuido y escalable para integrar simulaciones, postproceso y funcionalidades de visualización de datos de una manera sencilla, permitiendo el trabajo colaborativo entre ingenieros y científicos. La aplicación se encuentra dividida en múltiples etapas de proceso, representadas por módulos. Estos módulos, implementados como procesos independientes, pueden repartirse en distintas plataformas heterogéneas. Los módulos de visualización permiten entornos virtuales, cubriendo desde workbenches hasta power-walls, pantallas curvas o full domes (CAVEs). Los usuarios pueden visualizar los conjuntos de datos de manera intuitiva, en un entorno totalmente inmersivo, a través de técnicas como Volume Rendering, y Fast Sphere Rendering.

2.3.6. Ensemble-VIS

Ensemble-VIS es un framework consistente de una colección de vistas globales y estadísticas, unidas a través de un gran nivel interactividad. Esta herramienta permite a los científicos obtener información clave sobre la distribución de los resultados de simulaciones, así también como la incertidumbre asociada con los datos científicos, permitiendo al usuario dirigir las visualizaciones. Ensemble-VIS permite la exploración del rango de posibles predicciones, y la probabilidad de las mismas, así como la posibilidad de obtener información sobre el ensamble, por medio de distintas herramientas de visualización, co-

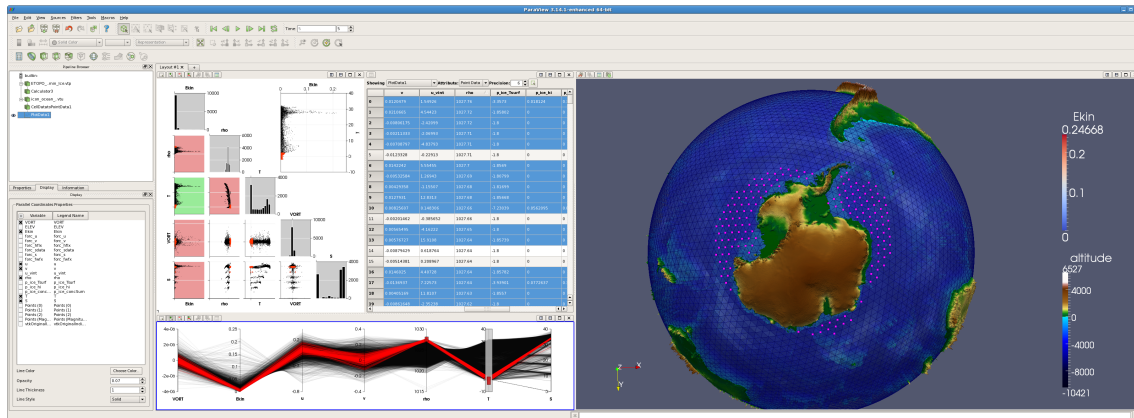


Fig. 2.3: Visualización de un conjunto de datos oceánicos ICON utilizando ParaView. En el panel de la izquierda se muestran los datos, las selección realizada y la topografía de la Tierra. Los tres viewports son utilizados para especificar la selección, basada en una matriz de scatterplot y coordenadas paralelas.

mo Height and Comparative Displays, Time Navigation Summaries, Trend Charts, Query Contours, and Spaghetti Plots.

2.3.7. VTK

El Kit de herramientas de visualización (VTK) es un sistema de software libre, disponible para la realización de gráficos 3D, procesamiento de imagen y visualización. VTK consiste en una biblioteca de clases de C++ y varias capas de interfaz, como Tcl/Tk, Java, y Python. VTK soporta una amplia variedad de algoritmos de visualización como: escalar vector Euclides, tensor, textura y métodos volumétricos; y avanzadas técnicas de modelado como: modelado implícito, reducción de polígonos, suavizado de malla (mesh smoothing), corte, contorneado y triangulación de Delaunay. VTK cuenta con un amplio marco de visualización de datos, soporta el procesamiento en paralelo y se integra con diversas bases de datos de herramientas GUI como Qt y Tk. VTK es multiplataforma y se ejecuta en plataformas Linux, Windows, Mac y Unix. VTK también incluye soporte auxiliar de widgets de interacción 3D, anotación bi y tridimensional y computación paralela. En su núcleo VTK es implementado como un conjunto de herramientas de C++, exigiendo a los usuarios crear aplicaciones combinado varios objetos en una aplicación. El sistema también soporta ajuste automatizado del núcleo de C++ en Python, Java y Tcl, para que también se puedan escribir aplicaciones VTK utilizando estos lenguajes de programación interpretados.

2.3.8. ArcGIS

ArcGIS permite a los usuarios construir mapas de manera sencilla, utilizando conocimiento geográfico. Es un software propietario, implementado por el Environmental Systems Research Institute (ESRI). Los usuarios finales puede acceder a mapas, datos a través de computadoras portátiles, teléfonos móviles, tablets, y navegadores de Internet. Los desarrolladores web pueden construir aplicaciones para múltiples plataformas (JavaScript, Flex,

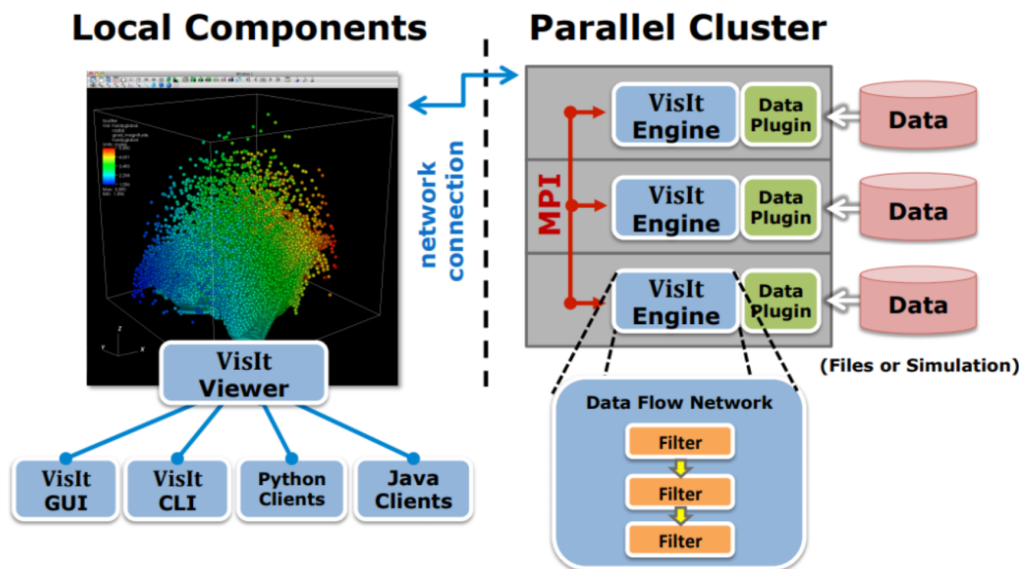


Fig. 2.4: VisIt implementa una arquitectura cliente servidor, permitiendo así trabajar con grandes conjuntos de datos, y cortos tiempos de respuesta.

Silverlight y SharePoint) utilizando la API disponible. Un SDK se encuentra disponible para Windows Mobile, smartphones y tablets. ArcGIS es un software comercial y el usuario enfrenta costos adicionales con cada actualización disponible. Algunas de sus herramientas (por ejemplo la Clip Geoprocessing Tool) son lentas y producen frecuentemente resultados incorrectos.

2.3.9. AVS/Express

AVS/Express soporta programación orientada a objetos y es mayormente usado con propósitos de visualización, por programadores y no programadores. Disponible para varias plataformas, soporta técnicas de visualización en dos y tres dimensiones, utilizando conjuntos de datos de en el orden de los gigabytes. AVS/Express contiene una librería de 900 módulos para realizar diversas visualizaciones y tareas de manipulación de datos, y admite el procesamiento distribuido a través de memoria compartida, lo cual resulta en tiempos rápidos de procesamiento. AVS/Express es fácilmente integrable con lenguajes de programación tales como C, C++ y Fortran, escalando a grandes y complejos conjuntos de datos.

2.3.10. GrADS

Grid Analysis and Display System (GrADS) es una herramienta de visualización utilizada para manipular y visualizar datos científicos de la tierra (Earth Science Data) en un espacio de 5-dimensional. GrADS utiliza expresiones de línea de comando basadas en la sintaxis FORTRAN para ejecutar operaciones, y provee una amplia gama de funciones embebidas. Adicionalmente, los usuarios pueden desarrollar funciones externas en cualquier lenguaje de programación. La herramienta es open source, y soporta múltiples formatos de datos, aunque presenta una curva de aprendizaje, dado que carece de interfaz gráfica.

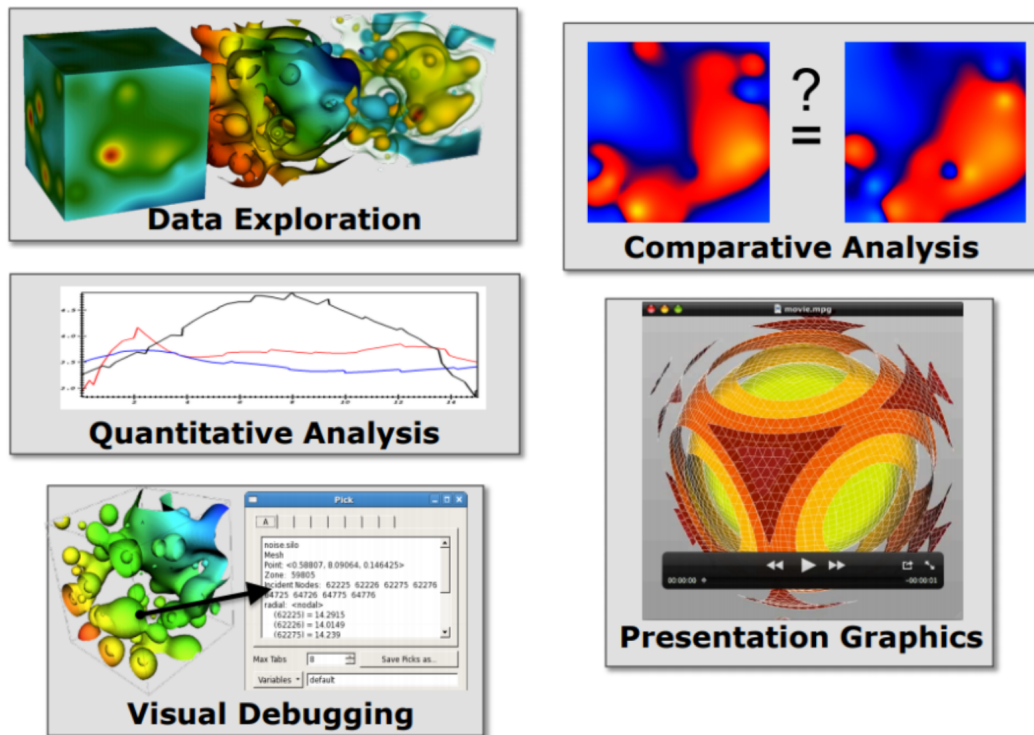


Fig. 2.5: Visit permite al usuario trabajar con un amplio rango de casos de uso.

2.3.11. Integrated Data Viewer (IDV)

IDV es un framework escrito en Java, que permite el análisis y visualización de datos geocientíficos. Java y Java3D son necesarios para ejecutar el framework IDV. IDV es gratuito y permite visualizaciones 3D de alta calidad, así como graficar a partir de datos en servers remotos. Soporta múltiples tipos de datos, pero hace uso de grandes cantidades de memoria RAM, lo cual representa un problema al trabajar con conjuntos de datos excesivamente grandes.

2.3.12. VisTrails

El objetivo fundamental de VisTrails es permitir una comparación visual de resultados de distintos flujos de trabajo. Dichos resultados pueden ser visualizados en un amplio rango de tamaños de pantallas, que va desde pantallas de pared hasta teléfonos celulares. VisTrails está escrito en Python y puede ser extendido por medio de plugins. Existen plugins especializados que permiten la integración con otras herramientas como VisIt y ParaView. Un puntos fuerte de VisTrails es la presentación de una interfaz simple y sencilla de usar, con buenas comparaciones de visualizaciones. Además, VisTrails cuenta con una comunidad de usuarios importante. A pesar de sus buenas características, VisTrails también sufre algunas falencias como capacidades para procesamiento paralelo limitadas. Además, es frecuente que el programa se detenga mientras actualiza grandes cantidades de datos desde un servidor remoto.

Nombre	University / Organization	Open source	Framework / Toolkit	Comparative Visualization?	Uncertainty / Ensembles	Interfaz Web	Scalability
Visus	University of Utah	No	Toolkit / Framework	No	Via UV-CDAT	Yes	Yes
UV-CDAT	University of Utah and the Polytechnic Institute of New York University	Yes	Framework	Yes	Yes	No	Yes
ParaView	Kitware Inc. and Los Alamos National Laboratory	Yes	Toolkit / Framework	Yes	Yes	Via Para-viewWeb	Yes
Visit	Lawrence Livermore National Laboratory	Yes	Toolkit / Framework	Yes	Yes	No	Yes
COVISE	High Performance Computing Center Stuttgart (HLRS), University of Stuttgart	Yes	Framework	No	No	No	Yes
Ensemble-Vis	Sandia National Laboratories, Albuquerque, New Mexico, Lawrence Livermore National Laboratory, Livermore, California	No	Framework	Yes	Yes	No	Yes
VTK	Kitware	Yes	Framework	No	Yes	Via Vtk.web	Yes
ArcGIS	Environmental Systems Research Institute (ESRI)	No	Framework / Toolkit	No	No	Yes	No
AVS / Express	Advanced Visual Systems Inc.	No	Framework	Yes	No	No	Yes
GRADS	Center for Ocean-Land-Atmosphere Studies, Institute of Global Environment and Society, George Mason University	Yes	Toolkit	Yes	Yes	No	No
Integrated Data Viewer (IDV)	University Corporation for Atmospheric Research, Boulder, Colorado	Yes	Framework	No	Yes	No	No
VisTrails	Cientific Computing and Imaging Institute at the University of Utah	Yes	Framework	Yes	No	No	No

Tab. 2.1: Herramientas de visualización utilizadas para el análisis de pronósticos y datos climatológicos.

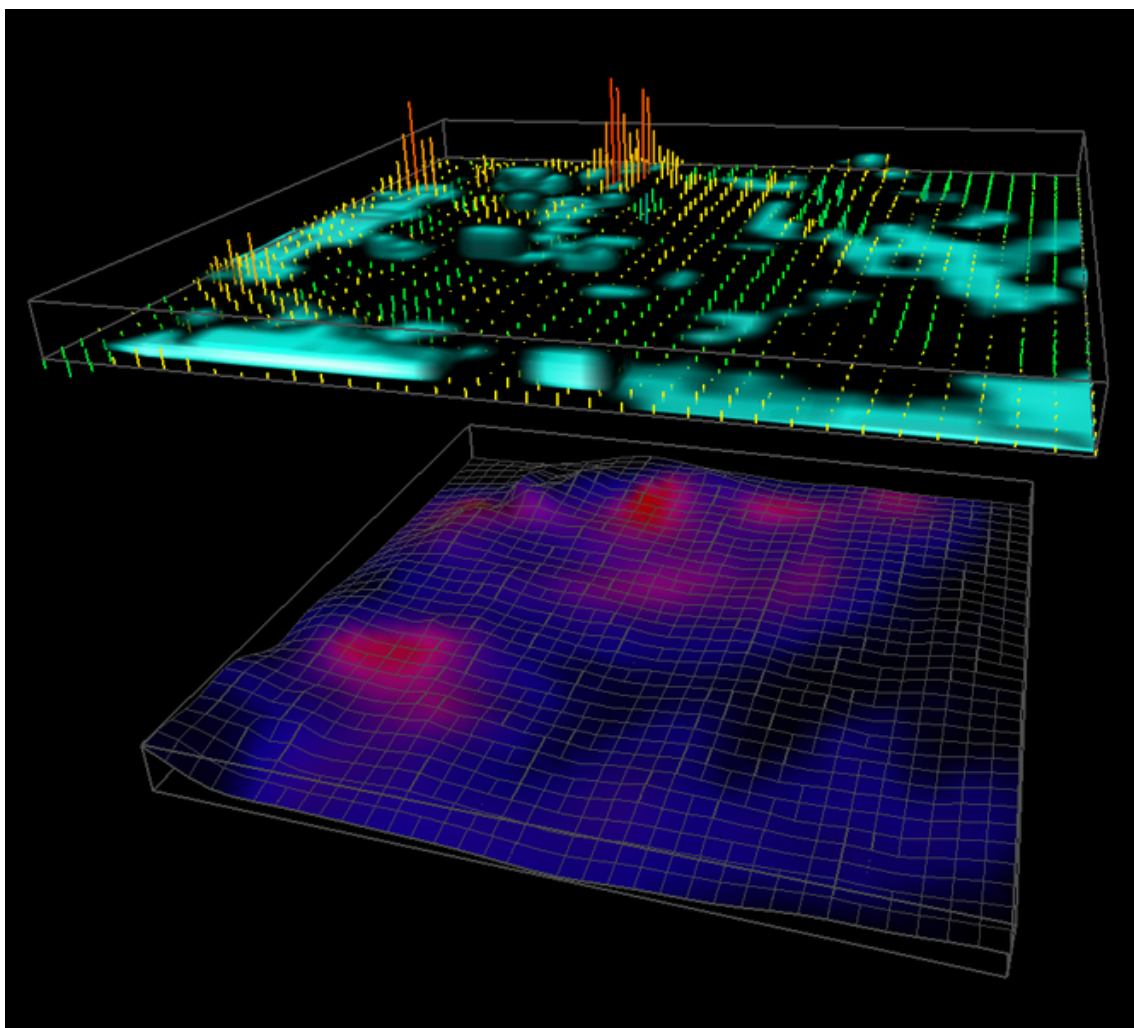


Fig. 2.6: Visualización interactiva en COVISE de la interacción entre la supercificie terrestre y las precipitaciones en el Oeste de Africa.

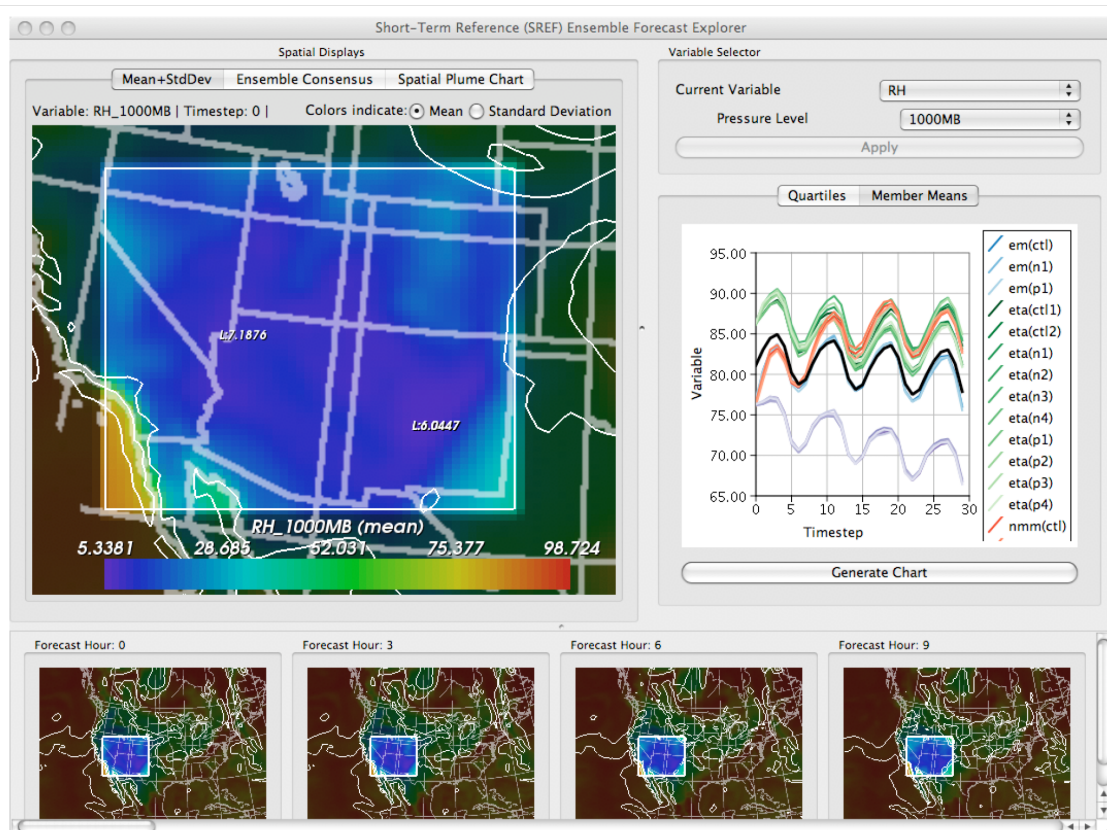


Fig. 2.7: Ensemble-Vis provee una plataforma de visualización y análisis de datos a través de la combinación de múltiples técnicas de visualización estadística y un alto nivel de interacción con el usuario.

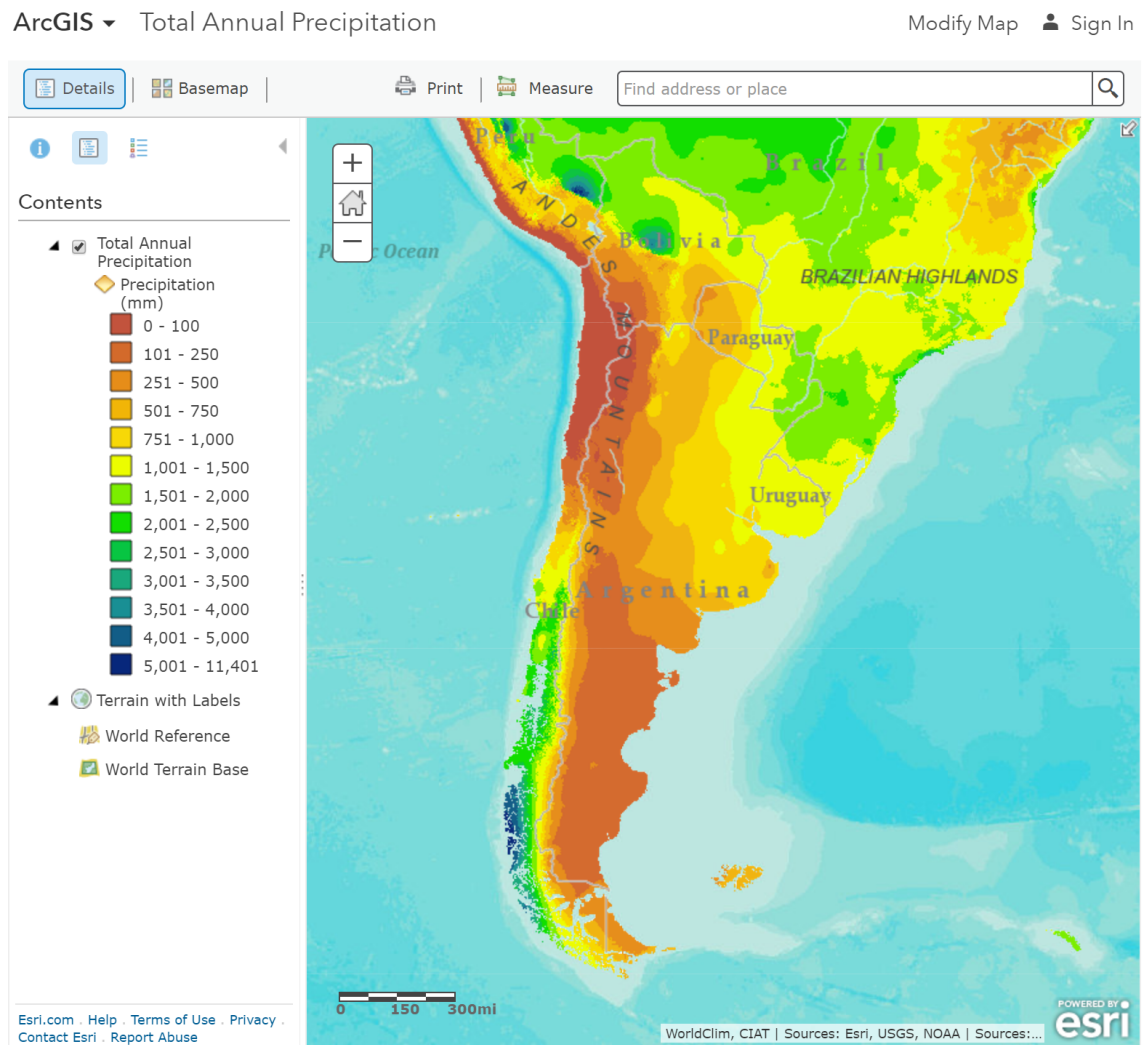


Fig. 2.8: ArcGIS mostrando la precipitación anual observada en 2014, para la región de la República Argentina.

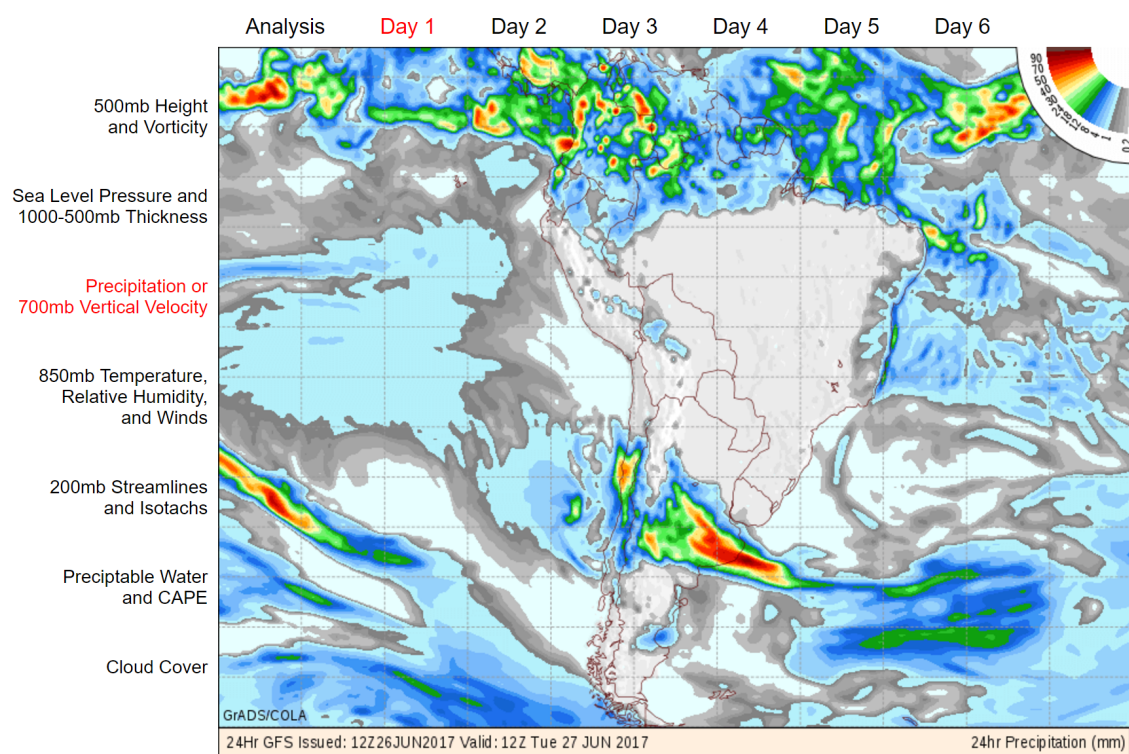


Fig. 2.9: GRADS. Pronóstico diario de precipitación para la región de Sudamérica.

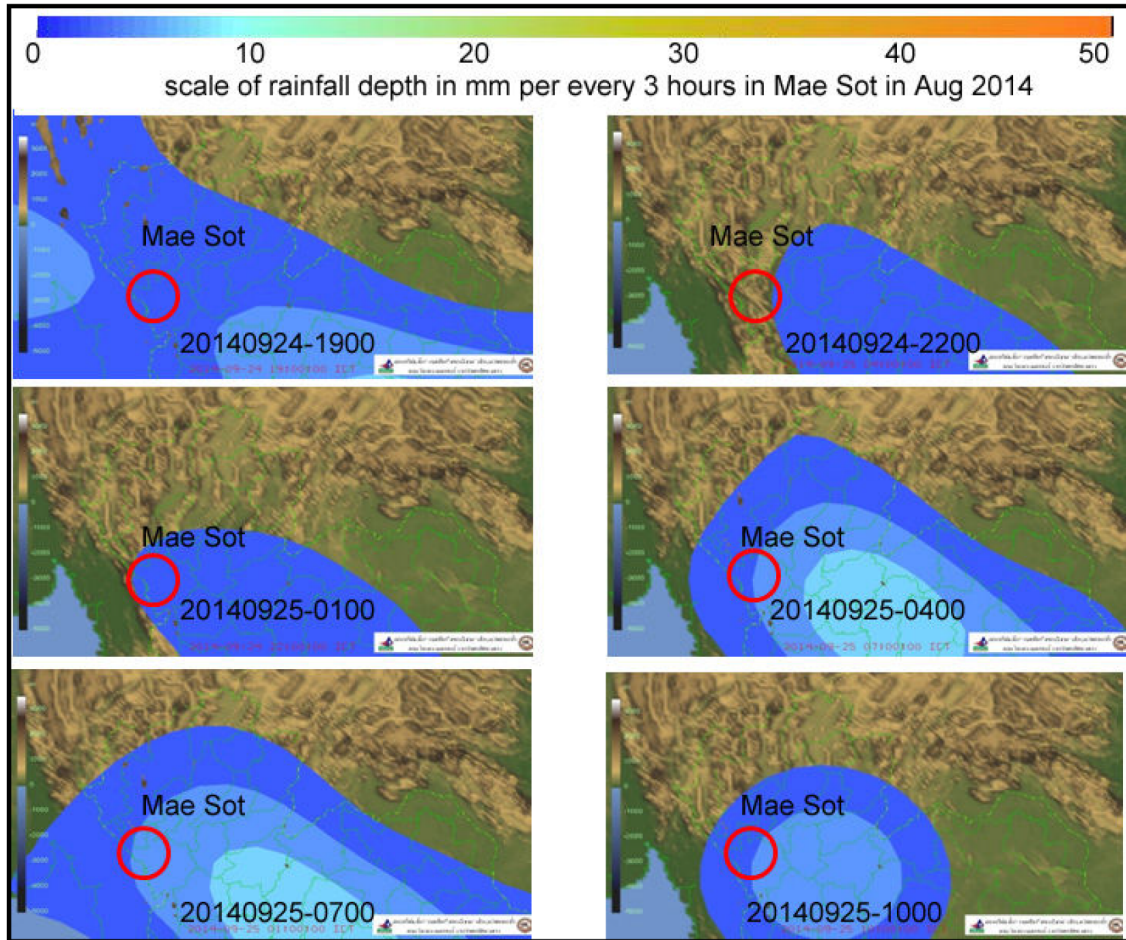


Fig. 2.10: IDV. Mapas de precipitación simulada cada 3 horas, sobre el distrito de Mae Sot, Tailandia, durante 24/09/2014:19.00 a 25/09/2014:10.00.

3. REFORECAST ANALOG REGRESSION (RAR)

3.1. Introducción

Esta técnica de corrección estadística por pronósticos análogos, descrita por Hamill y Whitaker [9], consiste en un proceso de tres pasos que se introdujo como una manera de generar pronósticos probabilísticos, con el objetivo de superar algunas limitaciones de los pronósticos determinísticos y por ensamble. En un primer paso se generan pronósticos para los días siguientes a una fecha dada, utilizando un modelo numérico específico. El modelo numérico utilizado debe coincidir con aquel utilizado para generar los pronósticos de reforecast.

El segundo paso consiste en comparar el pronóstico actual con todos los pronósticos pasados disponibles, en una misma época del año y región local. Segundo, utilizando las fechas de las coincidencias más próximas, se genera un ensamble de observaciones, a partir de los cuales se construye un pronóstico probabilístico utilizando la frecuencia de ocurrencia de un evento en particular. Combinando los pronósticos probabilísticos para cada región local, un pronóstico global es obtenido. Se ha demostrado que esta técnica es capaz de corregir el sesgo del pronóstico y las deficiencias del ensamble [9]. Para este trabajo, un tercer paso ha sido implementado. Dado que los pronósticos probabilísticos son calculados por regiones, y para cada región las fechas de los análogos elegidos pueden resultar muy diferentes, un proceso de smoothing es necesario al combinar los pronósticos obtenidos con el fin de evitar discontinuidades entre los mismos. Estos pasos pueden verse ejemplificados en la Figura 3.1.

3.2. Descripción del método

A continuación se describen en detalle los pasos seguidos por el método de Reforecast Analog Regresion, como han sido implementados para este trabajo.

3.2.1. Paso 1: Selección de análogos

Dada una región y una fecha para la cual desea obtenerse un pronóstico probabilístico, el proceso comienza obteniendo el pronóstico numérico asociado, el cual ha sido generado por medio del mismo modelo numérico utilizado para obtener los pronósticos de reforecast históricos. A continuación, la región es dividida en subregiones de $1^\circ \times 1^\circ$ (Figura 3.2) para las cuales se calculará un pronóstico local. Los pronósticos locales serán luego unidos de forma tal de obtener un pronóstico probabilístico para toda la región.

Para cada subregión obtenida se considera entonces el área de $3^\circ \times 3^\circ$ que la contiene, y se procede a comparar el pronóstico numérico en dicha región con cada uno de los pronósticos de la base de datos histórica, con el fin de elegir aquellos más cercanos a este. En el contexto de este trabajo, solo la precipitación acumulada se ha elegido como variable a tener en cuenta en la comparación, pero el método puede extenderse de forma tal de considerar otras variables, como temperatura o presión atmosférica. Durante la búsqueda de pronósticos análogos, solo se consideran aquellos pronósticos de reforecast contenidos en una ventana temporal de 45 días alrededor de la fecha dada. En el caso

de la Figura 3.1, que corresponde al primero de Abril de 2013, la ventana temporal está compuesta por las fechas entre el 15 de Febrero y el 15 de Mayo de 2013. Es importante limitar temporalmente la búsqueda debido a que los errores sistemáticos del modelo pueden diferir considerablemente dependiendo de la época del año.

El proceso de selección se realiza comparando los valores de precipitación acumulada en cada uno de los 16 puntos de la grilla (los círculos negros en la figura), entre el pronóstico numérico actual y cada uno de los reforecasts. Esta comparación se lleva a cabo utilizando la diferencia cuadrática media (RMS por sus siglas en inglés). Formalmente, si f_i ($1 \leq i \leq 16$) son los 16 puntos del pronóstico actual, y r_i ($1 \leq i \leq 16$) los 16 puntos del reforecast, entonces el RMS es calculado como:

$$\sqrt{\frac{\sum_{i=1}^{16} (f_i - r_i)^2}{16}}$$

Las n fechas históricas de menor RMS son elegidas como las fechas de los análogos. Los cuatro pronósticos más cercanos se muestran en la segunda fila de la Figura 3.1. En el marco de esta presentación, solo esta estrategia de selección ha sido implementada. Es parte de un trabajo futuro permitir al usuario elegir estrategias de selección alternativas.

3.2.2. Paso 2. Construcción del pronóstico probabilístico

El siguiente paso comienza con la obtención de la precipitación acumulada observada en las fechas de los análogos elegidos. En la tercera fila del ejemplo de la Figura 3.1 pueden verse las observaciones asociadas a los cuatro primeros reforecast seleccionados. El lector podrá observar una situación particular en las observaciones del día 17 de Abril de 2003, en la cual el archivo CMORPH provisto por la NOAA expone datos faltantes para ciertas coordenadas. Bajo estas circunstancias, el algoritmo interpola bilinealmente utilizando los datos vecinos existentes al momento de armar el pronóstico probabilístico.

Dado un evento en particular, por ejemplo precipitación mayor a 15 mm, el pronóstico probabilístico se genera en función de la frecuencia relativa del evento en las observaciones de los análogos elegidos.

Esta estrategia se ejemplifica en la Figura 3.3, en la cual cada punto representa una condición, utilizando el color verde en caso de que la misma resulte positiva, y rojo en caso contrario. El ejemplo muestra puntos distribuidos cada 0.25° , representando la resolución de los datos de precipitación acumulada CMORPH. Dada una latitud y longitud, si la condición es válida en un único análogo, entonces la probabilidad de evento en dicho punto se establece en 50 %. Si el evento resulta positivo en los dos análogos, se establece en 100 %, y en 0 % de no cumplirse en ninguno.

Siguiendo esta estrategia, se evalúa para cada punto de la subregión de $1^\circ \times 1^\circ$ la precipitación acumulada observada en las fechas de los análogos elegidos, obteniendo la probabilidad del evento para toda la subregión.

Dado que las probabilidades se calculan únicamente en la región delimitada por el cuadro de líneas punteadas, las probabilidades en las regiones adyacentes se calculan desplazando la región de búsqueda a una celda lindante de la grilla. El pronóstico a nivel global es construido uniendo los pronósticos individuales generados para cada una de las subregiones.

3.2.3. Paso 3. Smoothing

La estrategia utilizada calcula probabilidades para los puntos contenidos en una subregión de $1^\circ \times 1^\circ$, utilizando análogos obtenidos a partir del pronóstico numérico de la región de $3^\circ \times 3^\circ$ que la contiene. Para producir un mapa nacional de las probabilidades, el proceso se repite para otras regiones y el mapa final se construye uniendo dichos resultados.

Lamentablemente, las fechas de los análogos para las distintas regiones pueden ser muy diferentes en comparación con el de las regiones adyacentes. Esto puede resultar en una discontinuidad de las probabilidades en los límites entre los parches; por lo tanto, un algoritmo de smoothing es implementando con el fin de suavizar estas diferencias.

La Figura 3.5 proporciona un ejemplo del algoritmo de smoothing. En el ejemplo se busca estimar las probabilidades para los puntos de color naranja. Los análogos son determinados por la similitud del pronóstico actual en los puntos de la cuadrícula a gran escala (grandes puntos negros) a los pronósticos anteriores a estos mismos puntos. Para los puntos naranjas, hay en realidad nueve regiones donde los análogos y las subsiguientes probabilidades se calculan, expuestas en los nueve paneles de la Figura 3.5. El algoritmo básico de RAR descarta 8 de las probabilidades para la región de puntos naranja, y solo las probabilidades calculadas utilizando los análogos de la región central son utilizados, causando una clara discontinuidad entre parches contiguos. El algoritmo de suavizado implementado calcula y utiliza todas las probabilidades calculadas, aplicando distintos pesos según la distancia al centro de la región. La Figura 3.4 ilustra el nuevo cálculo de probabilidades por región, en la cual se calculan probabilidades para toda el área, y no solo para la subregión central de la misma.

Sean $w_{ul}, w_{um}, w_{ur}, w_{ml}, w_m, w_{mr}, w_{ll}, w_{lm},$ y w_{lr} , los pesos aplicados a las probabilidades utilizando la región superior derecha, la región superior central, y así sucesivamente. Sean $d_{ul}, d_{um}, d_{ur}, d_{ml}, d_m, d_{mr}, d_{ll}, d_{lm},$ y d_{lr} la distancia entre el centro de cada región (la cruz roja) y el punto para el cual se está estimando la probabilidad (el rectángulo verde), y sean $w'_{ul}, w'_{um}, w'_{ur}, w'_{ml}, w'_m, w'_{mr}, w'_{ll}, w'_{lm},$ y w'_{lr} los pesos desnormalizados. Se define D como la distancia máxima entre los puntos azules y el centro de la región. Primero, un peso desnormalizado es calculado como

$$w'_{xx} = \begin{cases} \frac{D-d_{xx}}{D+d_{xx}} & \text{if } d_{xx} < D \\ 0 & \text{if } d_{xx} \geq D \end{cases}$$

Donde $_{xx}$ es una de las nueve regiones. Una vez calculados los pesos desnormalizados, se normalizan de forma tal que la suma de los pesos sea 1.

$$w_{xx} = \frac{w'_{xx}}{w'_{ul} + w'_{um} + w'_{ur} + w'_{ml} + w'_m + w'_{mr} + w'_{ll} + w'_{lm} + w'_{lr}}$$

Luego la probabilidad en un punto dado se calcula como la suma de las probabilidades en las distintas regiones, multiplicados por sus respectivos pesos.

En la cuarta fila de la Figura 3.1 se muestran los pronósticos probabilísticos generados a partir de las observaciones de precipitación acumulada para los análogos elegidos.

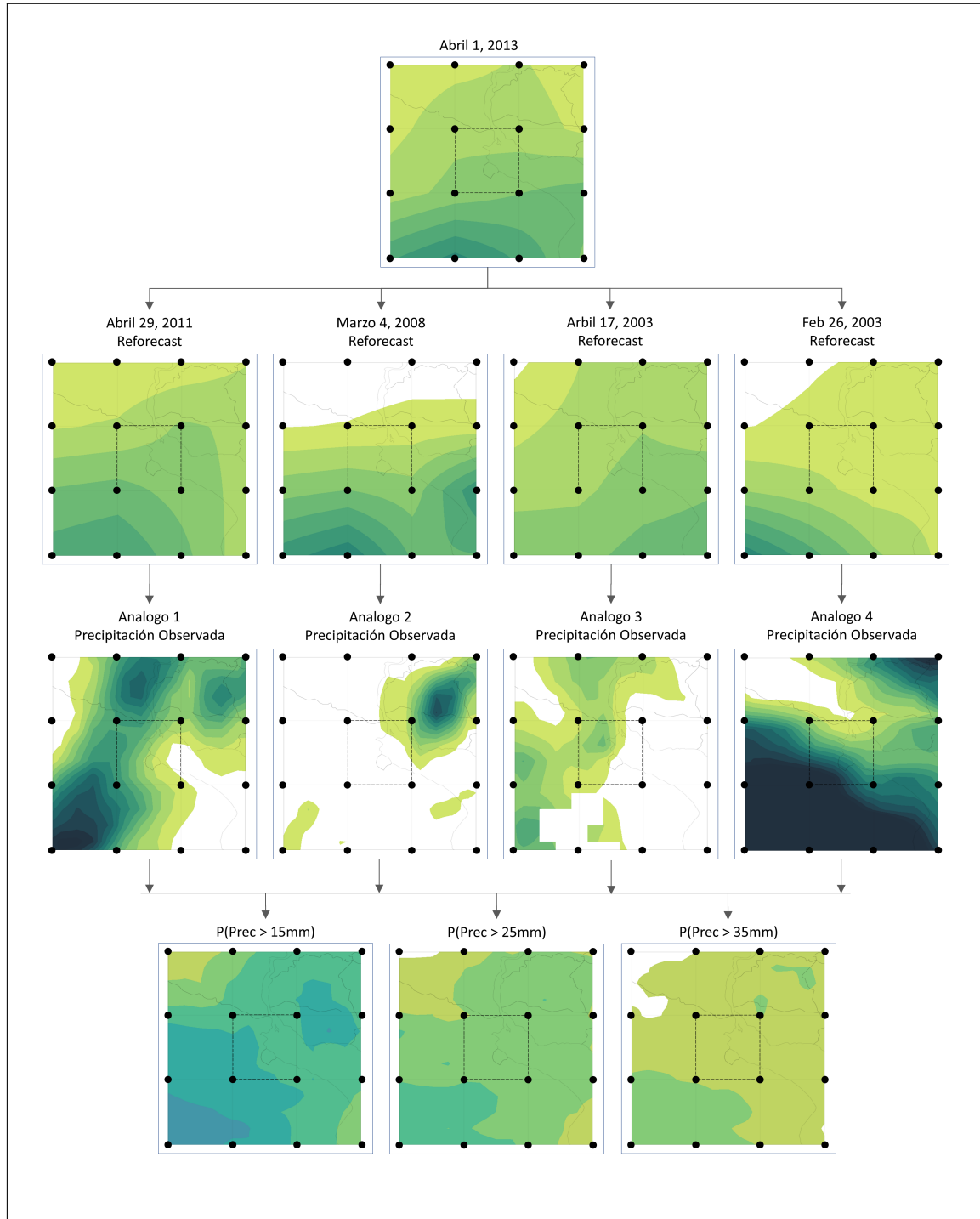


Fig. 3.1: Ilustración de la técnica RAR para un pronóstico de 24 hs. El pronóstico numérico se muestra en la primera fila, definido por los 16 puntos negros. El área central, delimitada por líneas punteadas, define la región para la cual se desea obtener el pronóstico probabilístico. Los cuatro pronósticos análogos mas similares se muestran en la segunda fila. Las precipitación acumulada observada para los análogos elegidos se presentan en la tercer fila. Finalmente, tres pronósticos probabilísticos, para umbrales de 15 mm, 25 mm y 35 mm se muestran en la última fila.

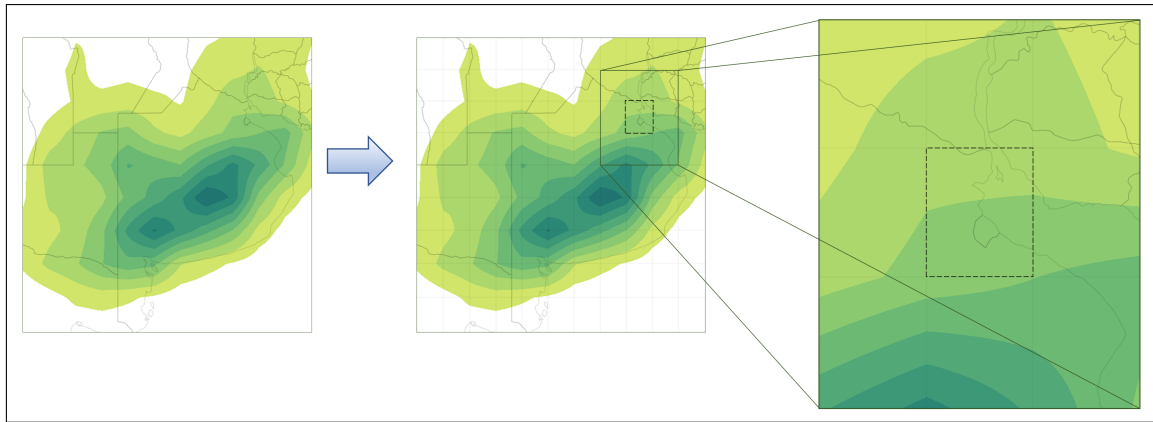


Fig. 3.2: El área es dividida en subregiones de $1^\circ \times 1^\circ$, para los cuales se calcularán pronósticos probabilísticos locales. Posteriormente, los pronósticos serán unidos para conseguir un pronóstico global.

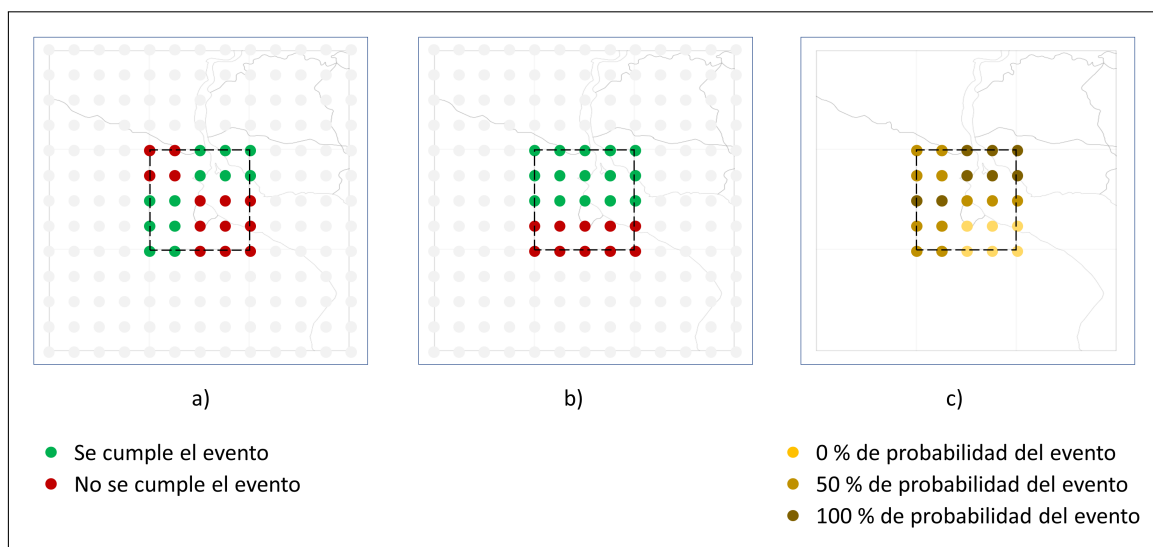


Fig. 3.3: Estrategia de armado de pronóstico probabilístico en función de la validez de un evento dado en los distintos puntos de la grilla. En el algoritmo básico de RAR solo se considera la región central de $1^\circ \times 1^\circ$.

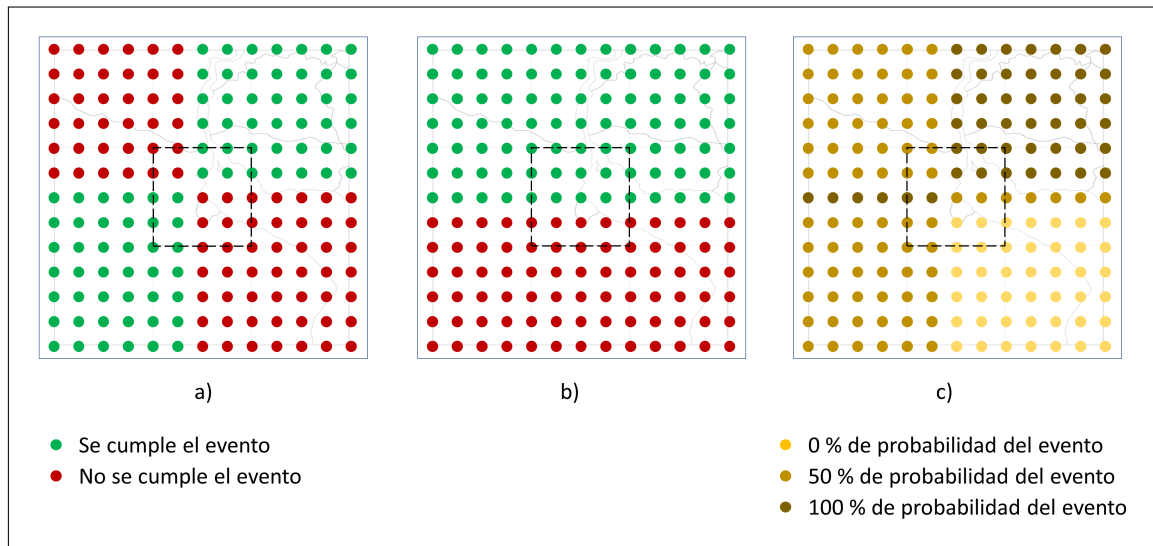


Fig. 3.4: Estrategia de armado de pronóstico probabilístico en función de la validez de un evento dado en los distintos puntos de la grilla, utilizada para el método de suavizado, en la cual se calculan probabilidades para toda el área de $3^\circ \times 3^\circ$.

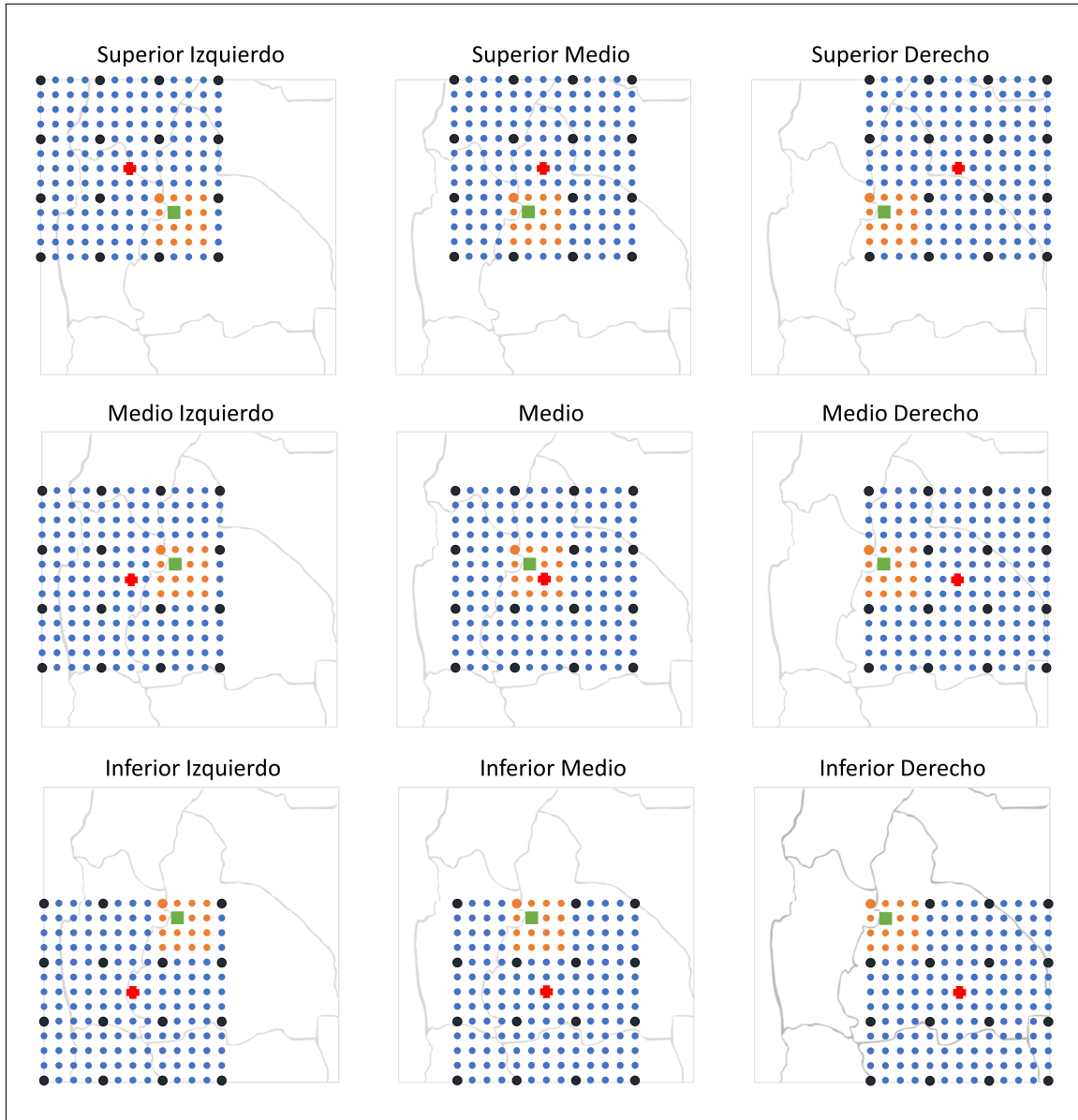


Fig. 3.5: Ilustración del algoritmo de smoothing. En este caso se buscan las probabilidades en los puntos naranjas. Los nueve paneles son las nueve regiones para las cuales análogos han sido calculados, que se superponen con la grilla de puntos naranjas. Los puntos restantes se denotan con color azul. Los análogos son calculados por similitud en los puntos negros de mayor tamaño; el centro de cada región análoga se denota con una cruz roja. Para el punto denotado por el rectángulo verde, la probabilidad final es calculada como la suma ponderada de los nueve estimados.

4. ALBERO

4.1. Descripción general

Albero es una aplicación de análisis visual que enriquece el método RAR mediante técnicas de visualización para analizar los resultados, así como explorar los aspectos internos del método. Es el resultado de un trabajo colaborativo con los meteorólogos investigadores y pronosticadores del Centro de Investigaciones del Mar y la Atmósfera y el Servicio Meteorológico Nacional de Argentina. En su flujo de trabajo actual, se ven obligados a ejecutar la técnica RAR repetidamente para cada umbral, plazo temporal, y configuración si necesitan realizar análisis adicionales. Actualmente utilizan el sistema de visualización GrADS y herramientas tales como Octave para trazar el proceso y sus resultados. Un ejemplo de su visualización actual se muestra en la (Figura 4.1).

Albero integra técnicas de visualización, como small multiples, mapas, gráficos e histogramas, combinadas y coordinadas para facilitar el trabajo del meteorólogo (Figura 4.2). Nuestra herramienta presenta un nuevo flujo de trabajo interactivo (véase Figura 1.3) que permite el análisis de nuevos aspectos de la técnica RAR y los datos asociados, tales como el análisis de múltiples pronósticos probabilísticos, análogos y observaciones, y resúmenes estadísticos.

Este trabajo apunta a dos tipos de usuarios: los meteorólogos que trabajan en predicción meteorológica operacional (pronosticadores) y los meteorólogos que trabajan en el desarrollo y evaluación de nuevas técnicas de pronóstico (investigadores). Nuestro enfoque ayuda a los pronosticadores en la detección anticipada de eventos extremos tales como inundaciones, fuertes precipitaciones, etc., e investigadores en la detección de errores sistemáticos para mejorar su modelo. Además, facilita el estudio del funcionamiento de la técnica RAR en si misma, permitiendo detectar posibles aspectos en los cuales la técnica puede ser mejorada (por ejemplo la estrategia de selección de análogos).

4.2. Fuentes de datos

Utilizamos los pronósticos numéricos retrospectivos (reforecasts) generados por el GFS (Global Forecast System) [10], de los Estados Unidos. Los reforecasts son pronósticos globales que estiman la evolución de la atmósfera durante un plazo máximo de 16 días. Estos pronósticos incluyen variables meteorológicas tales como temperatura, presión, vientos, humedad y precipitación, entre otros. Abarcan todo el globo con una resolución horizontal de 40 kilómetros aproximadamente entre los puntos de la cuadrícula. La resolución horizontal disminuye a 54 kilómetros entre los puntos de la cuadrícula para pronósticos entre una y dos semanas.

El ensamble de los pronósticos se compone de 11 miembros. En este trabajo utilizamos la media del ensamble para buscar análogos y nos centramos en el estudio de la variable de precipitación acumulada. La base de datos de pronósticos retrospectivos fue generada utilizando una misma versión del modelo GFS, lo cual permite suponer que los errores sistemáticos del modelo son los mismos para todos los pronósticos que componen dicho conjunto de datos. Esto permite una validación objetiva de la técnica, un estudio detallado de los sesgos sistemáticos del modelo, y una cuantificación mas precisa de la incertidum-

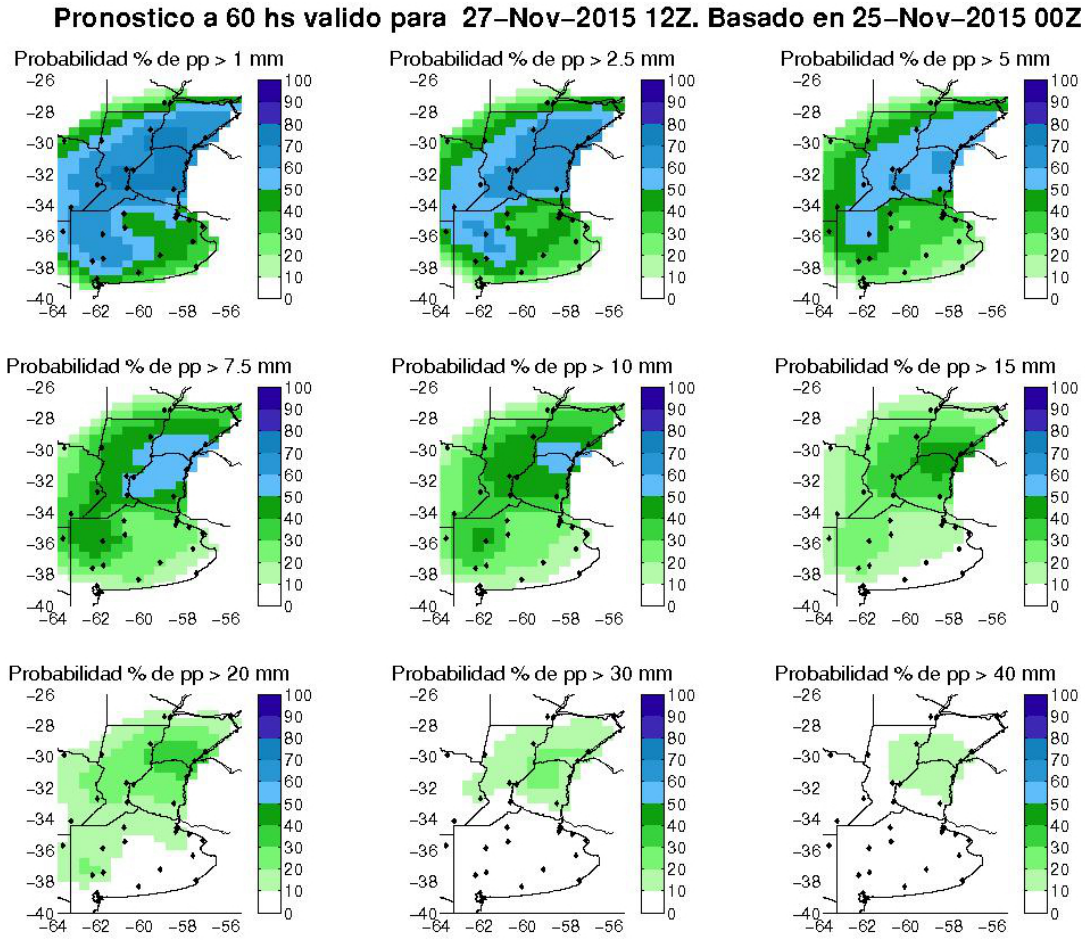


Fig. 4.1: Visualización de pronósticos de precipitación probabilísticos del Servicio Meteorológico Nacional. Creados usando GrADS y Octave [?].

bre de los pronósticos. La principal motivación de los conjuntos de reforecast es el de desarrollar una sólida base de datos históricos para cuantificar los errores sistemáticos y la incertidumbre del modelo y utilizar dicha información para mejorar la calidad de los pronósticos.

Se obtuvieron los pronósticos de precipitación acumulada sobre el área de la (Figura 4.3 para el período 2002-2016, es decir se obtuvieron los datos pronosticados de precipitación acumulada en la región comprendida entre 76W-52W y 56S-20S para diferentes plazos de interés. Los plazos de pronóstico utilizados son: 0, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72, 78, 84, 90, 96, 102, 108, 114, 120, 126, 132, 138, 144, 150, 156, 162, 168, 174, 180, 186, y 192.

Otro aspecto fundamental del método RAR son los datos observacionales. Actualmente, utilizamos datos de precipitación de observación del Climate Prediction Center (CMORPH - NOAA) [11]. Estos datos consisten en un análisis de precipitación global con una muy alta resolución espacial y temporal, de 0.25 grados y 30 minutos, respectivamente. Están basados en la información obtenida mediante la radiación emitida desde la tierra y la atmósfera en los rangos de frecuencia de microondas para estimar el contenido de

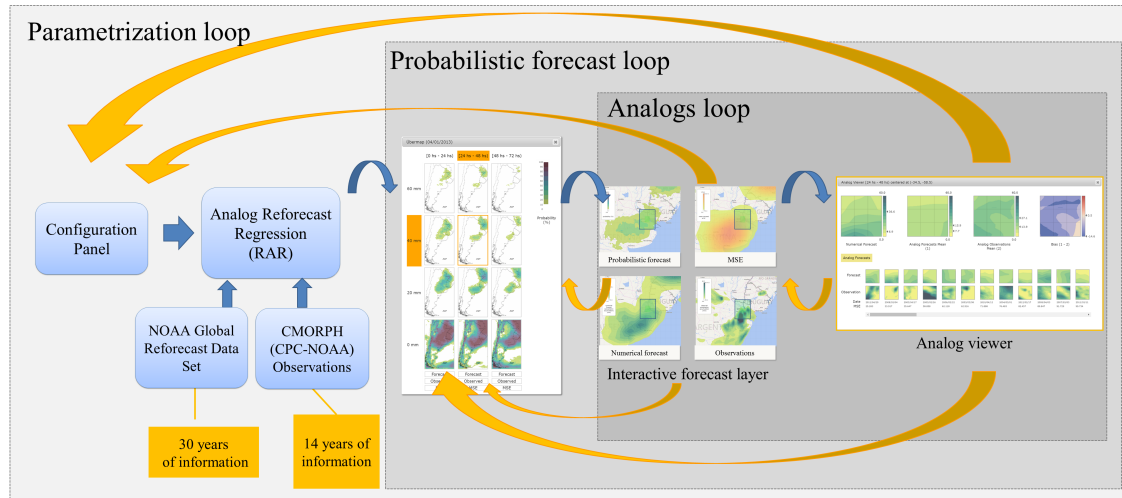


Fig. 4.2: El diagrama muestra cómo Albero se integra en el nuevo workflow interactivo. Distintas vistas vinculados son coordinados en un fluido diálogo entre el usuario y la aplicación, para ayudar al usuario en el proceso de toma de decisiones y la optimización de la técnica RAR

hielo y agua líquida en las nubes. La precipitación de superficie es inferida utilizando estos datos. Salio et al. [31] verificó que estos datos son válidos para la región de interés. Estos son rutinariamente utilizados para verificar simulaciones numéricas de modelos NWP y para el estudio de la variabilidad climática de las lluvias en distintas regiones del planeta.

Cabe destacar que si bien los datos de reforecast se encuentran disponibles desde el año 1984, este trabajo se limita a años posteriores a 2002, limitado por los datos observacionales provistos por la NOAA.

4.3. Workflow de tres ciclos

Albero está compuesto de múltiples vistas coordinados y vinculados en un diseño eficiente que permite al usuario una fluida interacción con la aplicación. Se compone de un novedoso workflow interactivo de tres ciclos anidados, el cual permite a los meteorólogos acceso a los componentes internos y los datos involucrados en la técnica RAR.

4.3.1. Ciclo de Parametrización

Nuestro workflow proporciona a los meteorólogos un panel de configuración que permite ajustar de forma interactiva parámetros de la técnica RAR. Algunos parámetros tienen efectos directos sobre la predicción meteorológica operacional y otros deberían ser establecidos por un experto, realizando previamente una validación objetiva de la nueva configuración.

El proceso de análisis se inicia cuando el usuario selecciona una región en el mapa donde quiere realizar el estudio. Por ejemplo, si un evento público es planificado, se necesita información acerca de las condiciones meteorológicas más probables para tomar decisiones respecto de dicho evento. Los pronósticos probabilísticos pueden darle al usuario una medida más exacta de la incertidumbre de la predicción. Si el usuario es un pronosticador,

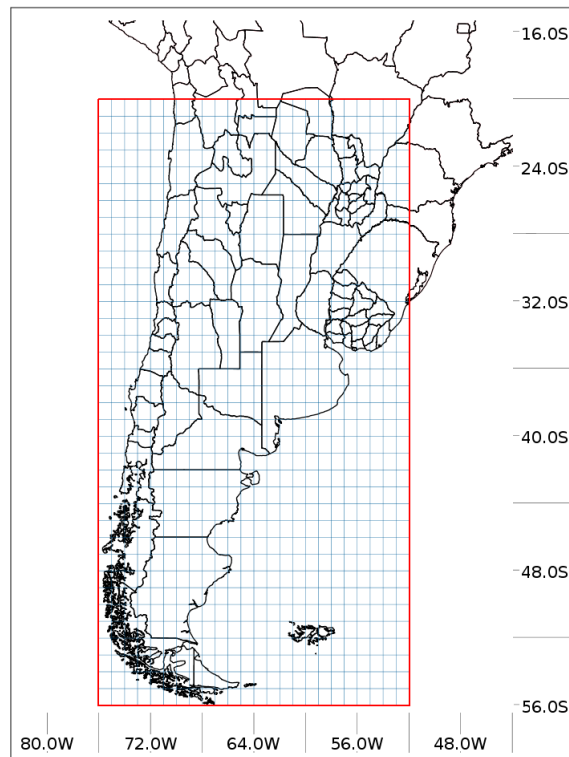


Fig. 4.3: Dominio para el cual se obtuvo el set de datos de pronósticos retrospectivos en el período 2002-2016, definido por 76°W - 52°W y 56°S - 20°S .

podrá limitar el análisis a un determinado ajuste temporal, caracterizado por la fecha de pronóstico, plazo, rango de acumulación, intervalo de búsqueda, entre otros. Un investigador puede, por ejemplo, analizar el impacto de aumentar o disminuir la cantidad de análogos que se usan en el cálculo de los pronósticos probabilísticos. Para esa tarea, el investigador selecciona un conjunto de casos donde se produjeron precipitaciones intensas en el pasado. Para cada uno de esos casos, evalúa los cambios del pronóstico probabilístico al modificar la cantidad de análogos utilizados, y calcula los errores usando los datos observados. De esta manera, determina si los pronósticos probabilísticos están funcionando correctamente y que impacto tiene incrementar o disminuir la cantidad de análogos. Si bien un análisis de estas características es subjetivo, Alberio permite especificar el rango de datos históricos, el número de análogos utilizados para calcular los pronósticos probabilísticos y otros atributos específicos de la técnica. Como trabajo futuro, se incluirán métricas objetivas dentro de la herramienta para incrementar su utilidad con fines de investigación y desarrollo.

Una vez establecidas la configuración, el usuario puede continuar con el ciclo de pronóstico probabilístico, o permanecer en el ciclo de parametrización para afinar el análisis.

4.3.2. Ciclo de Pronóstico Probabilístico

En el ámbito de la predicción meteorológica operacional, existen diversas situaciones que requieren la evaluación de las probabilidades de ocurrencia de diferentes fenómenos

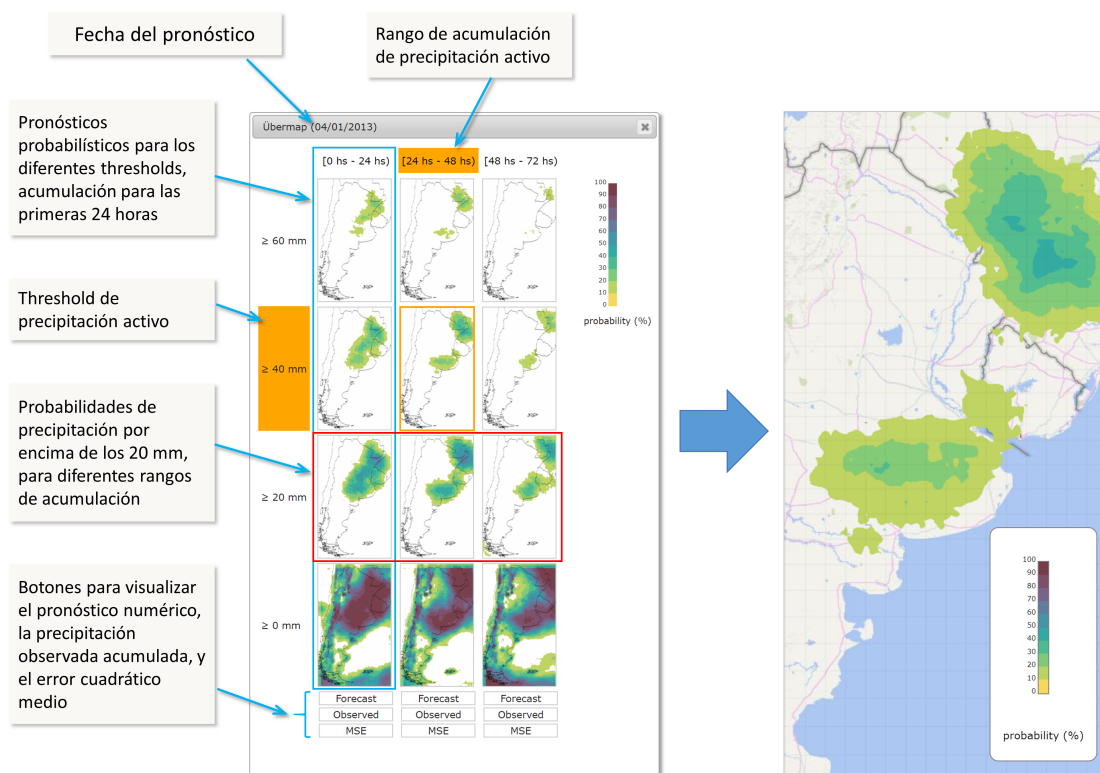


Fig. 4.4: Übermap. La visualización global muestra los pronósticos probabilísticos para los diversos umbrales de precipitación y rangos de acumulación. También permite visualizar el pronóstico determinístico, las observaciones y el error cuadrático medio (MSE), utilizando los botones correspondientes.

para distintas regiones y plazos temporales. Por ejemplo, la evaluación de las condiciones de sequía requiere el análisis de la precipitación acumulada a lo largo de una o dos semanas, mientras que la evaluación de riesgo de lluvia durante un evento público podría requerir el análisis de probabilidades de lluvia en cortos períodos de acumulación (por ejemplo, 12 horas).

Proponemos un componente de vista general, denominado “Übermap” (véase Figura 4.4). Esta herramienta visual se compone de una grilla de small-multiples de pronósticos probabilísticos, organizados en una cuadrícula por tiempo y magnitud. El tiempo se muestra en el eje horizontal, dividido en pasos definidos por el rango de acumulación y plazo del pronóstico. El eje vertical muestra diferentes umbrales definidos por la magnitud de la acumulación.

Este componente permite a los pronosticadores analizar simultáneamente la distribución de probabilidad como una función del tiempo, espacio y magnitud. El Übermap provee a los meteorólogos un segundo nivel de análisis: el ciclo de pronóstico probabilístico. En este bucle el usuario analiza el conjunto resultado de pronósticos probabilísticos asociadas con los diferentes umbrales. El Übermap provee una visión general de la probabilidad de un determinado fenómeno meteorológico para los próximos días, permitiendo al usuario analizar un conjunto de pronósticos probabilísticos para una determinada variable, resumir información sobre observaciones y filtrar los pronósticos por umbral de probabilidad.

El operador puede mostrar cualquiera de los pronósticos probabilísticos en el mapview, pudiendo luego desplazar el mapa, así como realizar operaciones de zoom, o seleccionar una sub-región del mismo. El mapview se divide en una cuadrícula de sub-regiones, utilizadas por el algoritmo de Hamill y Whitaker para calcular los análogos.

Es posible elegir un determinado umbral y rango de acumulación y visualizar la distribución de probabilidades en el mapview. Hemos diseñado un componente denominado Interactive Forecast Layer (véase Figura 4.5), el cual presenta una cuadrícula regular que divide el mapa en sub-regiones utilizadas por el algoritmo Hamill y Whitaker, compuesto por áreas de $3^\circ \times 3^\circ$ (como trabajo futuro se permitirá configurar estas dimensiones). Nuestra herramienta permite la exploración de las sub-regiones, así como sus correspondientes errores cuadráticos medios (MSE). El Interactive Forecast Layer permite visualizar los pronósticos probabilísticos (Figura 4.5a), los pronósticos numéricos (Figura 4.5b), observaciones (Figura 4.5c), y distribución espacial del error (Figura 4.5d). Los errores espaciales dependen de cómo el algoritmo se comporta en una determinada región del mapa. Analizando el MSE, el usuario puede evaluar el rendimiento esperado del método cuando se aplica a una situación concreta y a una determinada parametrización.

4.3.3. Ciclo de Análogos

En este ciclo, el meteorólogo puede analizar el conjunto de análogos y observaciones utilizado en la construcción del pronóstico probabilístico para una determinada sub-región. Haciendo clic en una sub-región, el usuario tiene acceso a un nuevo componente, llamado Analog Viewer (véase Figura 4.6). Este componente proporciona una visión desglosada de los componentes internos de la técnica. Aporta nuevas capacidades de análisis visual más allá de los pronósticos probabilísticos, tales como una visualización de los análogos y sus correspondientes observaciones, su media y desviación estándar, los MSE asociados y otras propiedades estadísticas. Esta vista detallada facilita el análisis de fenómenos concretos, por ejemplo, posibles precipitaciones extremas sobre una zona densamente poblada.

Cuando el análisis se haya completado, el usuario puede regresar al ciclo de pronóstico probabilístico para analizar otros pronósticos o al ciclo de parametrización.

4.4. Proceso de Diseño

En esta sección describiremos el proceso de diseño participativo que hemos seguido en el desarrollo de este trabajo. Se llevaron a cabo una serie de sesiones con siete expertos del dominio, meteorólogos con diferentes orígenes. Dos de ellos trabajan en el Centro de Investigación De La Atmósfera (CIMA), siendo uno de ellos un investigador senior que participó desde el principio del proyecto, y el otro posee conocimientos de ciencias de la computación. Los otros cinco meteorólogos son investigadores que trabajan en el Servicios Meteorológico Nacional. Uno de ellos siendo el jefe del departamento de investigación y desarrollo. Este trabajo requirió extensos esfuerzos interdisciplinarios entre los meteorólogos y los científicos en computación. Se describen a continuación las distintas fases del proceso y los resultados de nuestras reuniones, así como hitos y prototipos construidos.

El trabajo fue estructurado en tres fases: inepción, diseño y evaluación. El objetivo de la primera fase, inepción, fue el de generar ideas para el contenido y el diseño, lograr una mejor comprensión del público apuntado, y obtener información sobre los requisitos de nuestros expertos del dominio. El objetivo de la fase de diseño fue el de reunir los requisitos detallados de los expertos, discutir los prototipos, y elaborar y validar las opciones de

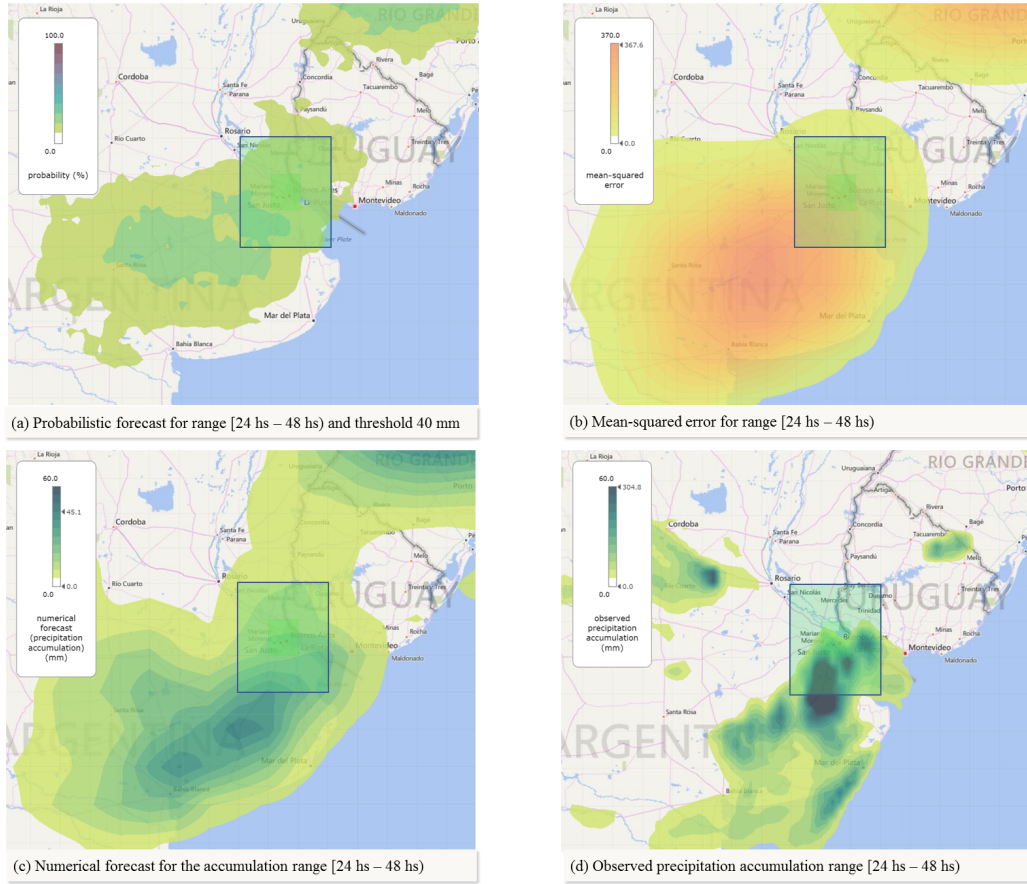


Fig. 4.5: Interactive Forecast Layer, mostrando un primer plano de: (a) pronóstico probabilístico mostrando una grilla interactivo dividida en sub-regiones utilizadas por el RAR, (b) distribución espacial del MSE (c) pronóstico numérico, (d) observaciones. Haciendo clic en una de las sub-regiones, el usuario tiene acceso a información detallada del ensamble de análogos y observaciones utilizadas en el cálculo del pronóstico para la sub-región.

diseño con ellos. Esta fase se realizó siguiendo un proceso iterativo e incremental. Por último, el objetivo de la fase de evaluación fue el de presentar Albero, junto con un conjunto de casos de estudio, y evaluarlos con los expertos del dominio, mostrando las capacidades de nuestra solución.

4.4.1. Fase de Incepción

Durante la fase inicial hemos discutido con nuestros expertos de dominio el estado del arte de técnicas de visualización y prácticas, tales como information visualization seeking-mantra, técnicas de focus+context, linking and brushing y lentes interactivos, entre otros métodos. Los expertos utilizan Octave y GrADS para procesar sus datos y generar gráficas 2D.

Elegimos la variable de acumulación de precipitación dado que se trata de una de las variables de mayor impacto en las actividades humanas. Afecta a distintos aspectos de la vida, tales como la planificación, toma de decisiones, y la productividad. Las inundaciones

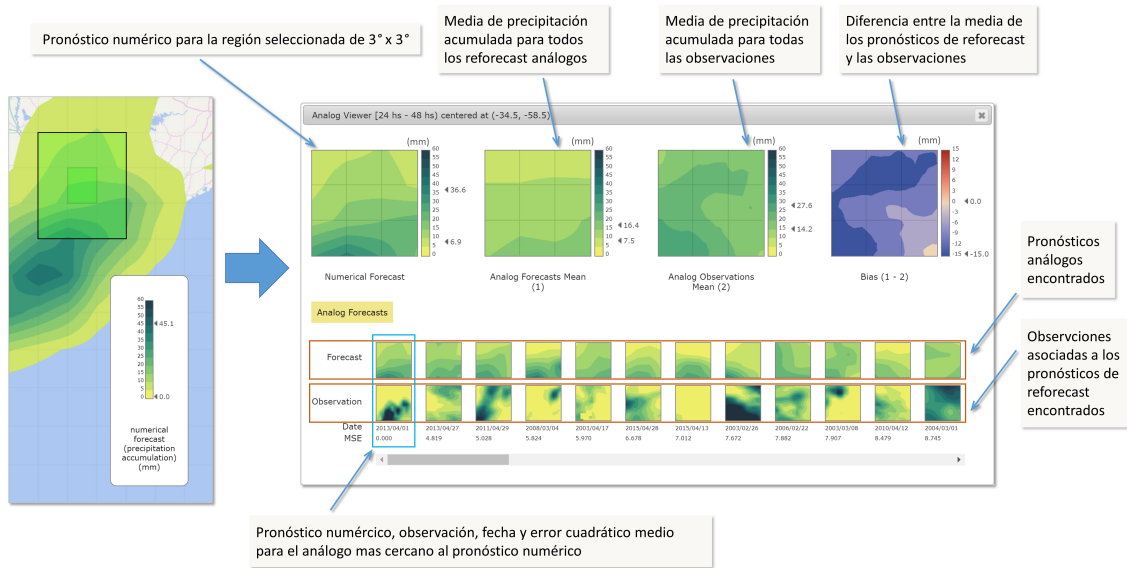


Fig. 4.6: Analog Viewer para la región seleccionada, en el segundo rango de acumulación [24 hs a 48 hs). Muestra la información utilizada internamente en el método RAR: la lista de análogos elegidos y sus observaciones, los valores medios y el sesgo, vinculada al Interactive Forecast Layer.

y tormentas severas pueden causar desastres y son un gran problema en nuestro país.

En el workflow automático, los expertos del dominio no tienen grado de interactividad alguno. Cada vez que se requería evaluar un nuevo umbral o intervalo de tiempo, se debía volver a ejecutar el proceso desde el principio, obteniendo como resultado una única imagen. Por ejemplo, si un pronosticador necesitaba analizar siete días de acumulación de precipitación, debía ejecutar un script y combinar las imágenes de salida, haciendo la comparación engorrosa. El resultado de esta fase es la realización de que las herramientas de visualización existentes fallaban en satisfacer las necesidades de los expertos del dominio.

4.4.2. Fase de Diseño

En la fase de diseño, hemos redactado un primer prototipo para ser revisado junto con nuestro principal stakeholder, quien es uno de los expertos involucrados en este trabajo. Presentamos una visualización basada en mapas de los pronósticos probabilísticos, y varias vistas vinculadas, tales como gráficos de líneas e histogramas. Esto dio a luz al primer bucle de flujo de trabajo, el bucle de parametrización.

Seguimos este proceso con una serie de sesiones con todos los expertos, debatiendo nuevos prototipos, lo que nos permitió diseñar el flujo de trabajo del segundo bucle, el bucle de pronóstico probabilístico. Estos prototipos evolucionaron, incluyendo varias vistas vinculadas y coordinadas, entre ellas, una vista de ensambles de pronósticos probabilísticos para diferentes umbrales. Esto dio nacimiento al Übermap. Los meteorólogos sugirieron la adición de resúmenes estadísticos. Esto les ayudaría a identificar las precipitaciones extremas y proporcionaría información adicional. Por ejemplo, el meteorólogo puede analizar la probabilidad de ocurrencia de diferentes umbrales e identificar las regiones donde la precipitación supere cierto umbral de interés.

Se propuso además representar el pronóstico determinista asociado, y un mapa de la distribución espacial del error cuadrático medio de la técnica RAR. De esta manera, es posible recopilar información valiosa para el análisis del comportamiento del algoritmo para un evento concreto.

Una de las principales preocupaciones discutidas en nuestras reuniones fue la selección de las escalas de colores. Hemos reconocido que las opciones elegidas no eran lo suficientemente útiles para detectar las probabilidades en el mapa y, en general, las variaciones eran demasiado suaves para identificar niveles contiguos de una variable, incluso utilizando escalas cuantificadas. Aunque la combinación de colores del arco iris puede conducir a una interpretación errónea, los meteorólogos consideran que es útil identificar rápidamente las variaciones de los diferentes umbrales y detectar así un acontecimiento grave. La elección adecuada de las paletas de colores puede incrementar considerablemente la usabilidad de visualizaciones. Las variables cuantitativas se visualizan mejor mediante escalas de colores de un solo matiz (single-hued), usando diferentes niveles de luminancia para representar los distintos niveles de la variable. La Sociedad Meteorológica Americana (IIPS por sus siglas en inglés) ofrece recomendaciones sobre el uso de colores en funciones comúnmente utilizadas en meteorología, siendo una de ellas la utilización de escalas de color verde para indicar precipitación.

Stauffer et.al [32] Destaca el uso del modelo de color HCL para la meteorología, ya que ofrece una mayor legibilidad, y accesibilidad para personas con impedimentos visual. En el diseño de los esquemas de color hemos definido una escala de dos matices, lineal y cuantificada, utilizando el modelo de color HCL para evitar problemas debido a la no-linealidad perceptual. La Figura 4.7 muestra una comparación de diferentes escalas analizadas durante la fase de diseño. El primer mapa de la izquierda (véase Figura 4.7a) muestra el esquema de colores usado anteriormente por los meteorólogos. Se les sugirió a los expertos del dominio otras dos escalas de colores de dos matices. La escala del medio varía de verde a rojo, y el de la derecha de verde al violeta (véase Figura 4.7b). Estas escalas muestran la combinación de colores finalmente seleccionadas por nuestros expertos del dominio (véase Figura 4.7c).

Hemos propuesto, además, visualizar cada uno de los componentes utilizados en los pasos del algoritmo RAR. Para ello hemos diseñado el Analog Viewer, que expone análogos, sus errores, y observaciones involucradas en el cómputo de cada sub-región. Los expertos pidieron la inclusión de resúmenes estadísticos para los análogos, observaciones y el sesgo entre ellos. Este nuevo ciclo de análisis visual, denominado ciclo de análogos, permite identificar en qué sub-regiones la técnica seleccionó análogos más similares (o más dispares), y analizar el rendimiento del modelo NWP comparando los análogos y sus observaciones.

Hemos concluido este proceso cuando nuestras opciones de diseño satisficieron las exigencias de nuestros colaboradores. Como resultado de ello hemos construido Albero, teniendo en cuenta eficiencia y escalabilidad. La solución admite un análisis altamente interactivo de grandes volúmenes de datos de información meteorológica, utilizando técnicas de computación de alto rendimiento para funcionalidad interna.

4.5. Implementación

Albero es un módulo completamente nuevo para la predicción meteorológica probabilista, construido mediante una arquitectura flexible basada en el sistema VIDa [6]. VIDa es una solución de análisis visual que proporciona soporte para la predicción meteorológi-

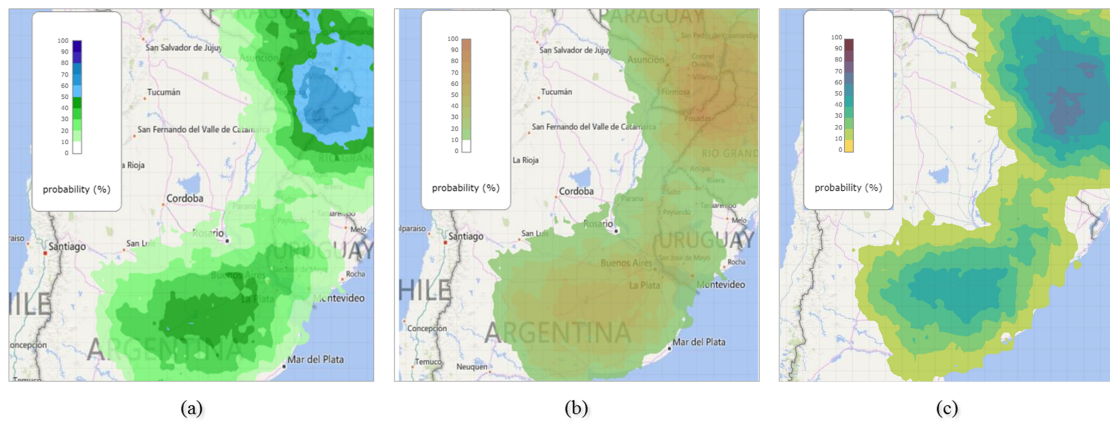


Fig. 4.7: Diferentes mapas de colores evaluados por nuestros expertos del dominio durante el proceso iterativo de refinamiento de nuestras opciones de diseño. Para representar la probabilidad de precipitación acumulada, prefirieron el mapa de colores HCL de dos matices (C) , que va desde el transparente a verde a violeta.

ca numérica a corto plazo. La solución propuesta se basa en el modelo cliente/servidor de VIDA. El cliente, implementado como una aplicación web, ofrece una interfaz visual interactiva construida utilizando HTML, JavaScript y JQuery. BingMaps AJAX Control 8.0 se utiliza para mostrar el componente mapview. El servidor, implementado en C++, utiliza el procesamiento GPGPU para lograr un rendimiento interactivo. Una captura de nuestra solución se muestra en la Figura 4.8b.

VIDa ha sido construido como un sistema multicapas, con una capa de presentación que expone un front-end basado en tecnologías web, una capa de negocios que implementa los algoritmos meteorológicos, y una capa de persistencia basado en bases de datos geospaciales. VIDa facilita el desarrollo de módulos de capa de negocios y visualización, apoyados en la estructura multithread y de generación de tiles que VIDa provee. La arquitectura de VIDa puede apreciarse en la figura Figura 4.9, en la cual Albero se apoya, haciendo uso de su servidor. Múltiples módulos, como Semillón y Hornero se encuentran actualmente en desarrollo.

El servidor multithread de VIDA es el encargado de atender pedidos realizados desde la interfaz web, por medio de un proxy con el cual se comunica mediante sockets. Los pedidos son entonces decodificados y entregados al Albero, donde el método RAR es implementado. Una vez procesado el pedido, y habiendo obtenido un resultado, el servidor genera una salida (un mensaje, una imagen, un tile de Bing Maps, etc), la cual es regresada a la interfaz web. La implementación modular de VIDA lo hace tentador al momento de implementar una nueva herramienta meteorológica. En la Figura 4.10 se muestra una esquema resumido de las partes involucradas en la atención de pedidos. Agregar funcionalidad adicional se reduce a implementar nuevos comandos, permitiendo un incremento modular sin afectar el comportamiento existente.

4.5.1. Arquitectura

Los comandos aceptados por el módulo Albero montado sobre VIDA, son los siguientes:

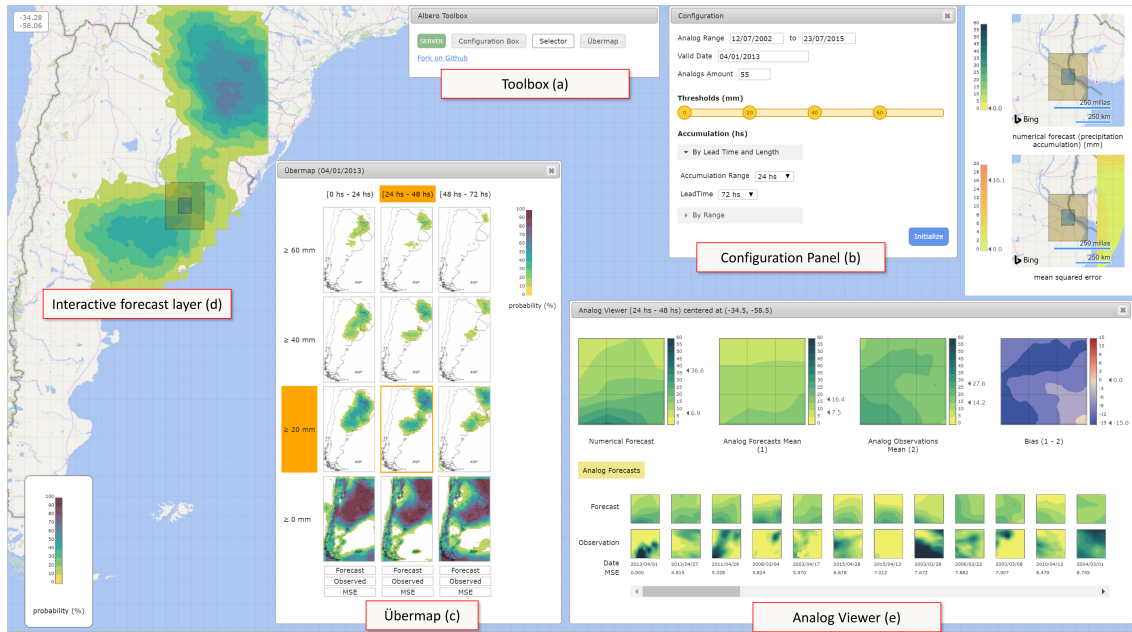


Fig. 4.8: Una captura de Albero mostrando algunos resultados de la técnica RAR, para un conjunto dado de parámetros de configuración.

4.5.1.1. GETAVAILABLEDAYS

Retorna un arreglo de fechas de reforecast disponibles en la base de datos actual.
Ejemplo:

```
["2002120700", "2002120800", "2002120900", "2002121000", ...]
```

4.5.1.2. INITIALIZE

Inicializa el servidor utilizando los parámetros indicados. Esta tarea lleva unos pocos segundos y almacena los resultados en una caché, permitiendo tiempos de respuesta prácticamente inmediatos en los sucesivos pedidos.

Este comando acepta los siguientes parámetros:

- ***analog-range-from***: Fecha desde la cual se contemplarán pronósticos análogos.
- ***analog-range-to***: Fecha hasta la cual se contemplarán pronósticos análogos.
- ***date***: Fecha de estudio.
- ***accumulation-ranges***: Periodos de acumulación en rangos de 6 horas. Por ejemplo $[[0,4],[4,8]]$ define los rangos [0 hs, 24 hs] y [24 hs, 48 hs].
- ***analog-amount***: Cantidad de análogos a considerar (75 por defecto).
- ***threshold-ranges***: Umbrales de acumulación (Ejemplo: 0 mm, 15 mm y 30 mm).

El siguiente es un ejemplo de pedido de inicialización.

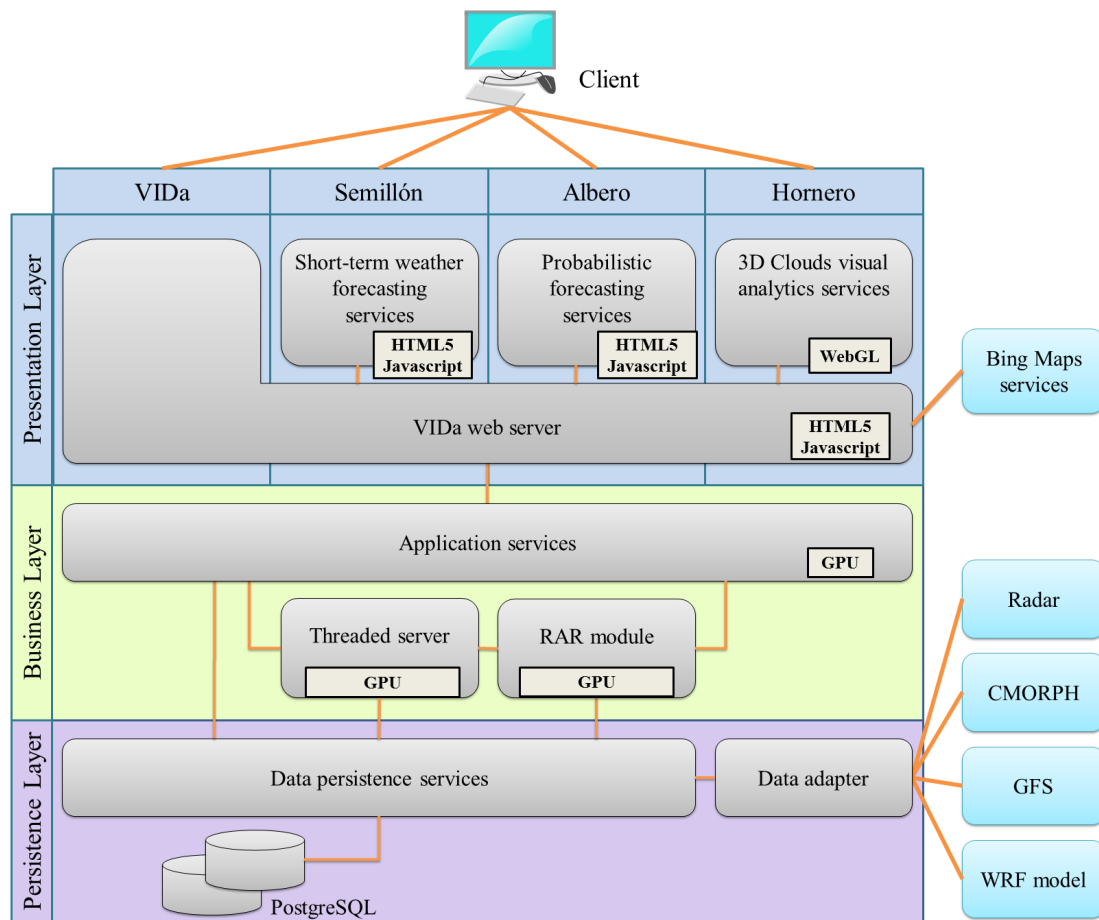


Fig. 4.9: Diagrama de arquitectura de VIDa (Visual Interactive Dashboard)

```
{
  analog-range-from:"12/07/2002",
  analog-range-to:"23/07/2015",
  date:"20172301",
  analogs-amount:55,
  action:"initialize",
  threshold-ranges:[0,20,40,60],
  accumulation-ranges:[[0,4],[4,8],[8,12]]
}
```

El comando retorna estadísticas del proceso de análisis, compuestos por:

- ***min_numerical_forecast***: Mínimo valor de precipitación acumulada pronosticada (en mm) para todos los pronósticos numéricos análogos.
- ***max_numerical_forecast***: Máximo valor de precipitación acumulada pronosticada (en mm) para todos los pronósticos numéricos análogos.

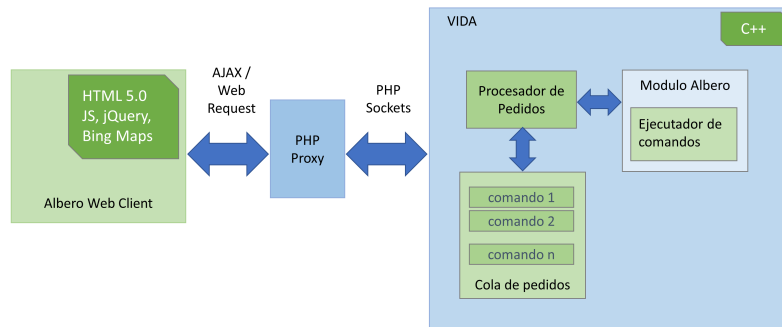


Fig. 4.10: Abstracción de las partes involucradas en la atención de pedidos. La estructura permite incrementos de funcionalidad modulares.

- ***min_observation***: Mínimo valor de precipitación acumulada observada (en mm) en las fechas de los pronósticos análogos.
- ***max_observation***: Máximo valor de precipitación acumulada observada (en mm) en las fechas de los pronósticos análogos.
- ***min_probabilistic_forecast***: Mínimo valor (en %) de probabilidad encontrada para todos los rangos de acumulación.
- ***max_probabilistic_forecast***: Máximo valor (en %) de probabilidad encontrada para todos los rangos de acumulación.
- ***min_mse***: Menor valor de MSE encontrado en la búsqueda de análogos.
- ***max_mse***: Mayor valor de MSE encontrado en la búsqueda de análogos.

El siguiente es un ejemplo de respuesta, en formato JSON.

```
{
  min_numerical_forecast: "0.000000",
  max_numerical_forecast : "42.360001",
  min_observation: "0.000000",
  max_observation: "151.383331",
  min_probabilistic_forecast: "0.000000",
  max_probabilistic_forecast: "29.696970",
  min_mse: "0.000000",
  max_mse: "17.440235"
}
```

4.5.1.3. PING

Este comando no acepta ningún parámetro, y es utilizado por el cliente web para conocer si el servidor se encuentra activo. Retorna el string “pong”.

4.5.1.4. GEOOBSERVATION

Retorna una imagen de la precipitación observada para la región de 3°x 3° alrededor de las coordenadas dadas, para un nivel de zoom en particular. Este comando es utilizado por BingMaps para visualizar la precipitación observada sobre el mapa, por medio de un custom Tile Layer.

Este comando acepta los siguientes parámetros:

- **lat**: Latitud del centro de la región.
- **lon**: Longitud del centro de la región.
- **z**: Nivel de zoom deseado.
- **range_index**: Índice del rango de acumulación deseado.
- **date**: Fecha para la cual se desea obtener la precipitación observada.

4.5.1.5. GEOPROBABILISTICFORECAST

Retorna una imagen del pronóstico probabilístico para la región de 3°x 3° alrededor de las coordenadas dadas, para un nivel de zoom en particular. Este comando es utilizado por BingMaps.

Este comando acepta los siguientes parámetros:

- **lat**: Latitud del centro de la región.
- **lon**: Longitud del centro de la región.
- **z**: Nivel de zoom deseado.
- **range_index**: Índice del rango de acumulación deseado.
- **date**: Fecha para la cual se desea obtener el pronóstico probabilístico.
- **threshold_index**: Índice del umbral de acumulación para el cual se desea obtener el pronóstico probabilístico.

4.5.1.6. GETNUMERICALFORECAST

Retorna una imagen del pronóstico numérico para la región de 3°x 3° alrededor de las coordenadas dadas, para un nivel de zoom en particular. Este comando es utilizado por BingMaps.

Este comando acepta los siguientes parámetros:

- **lat**: Latitud del centro de la región.
- **lon**: Longitud del centro de la región.
- **z**: Nivel de zoom deseado.
- **range_index**: Índice del rango de acumulación deseado.
- **date**: Fecha para la cual se desea obtener el pronóstico numérico.

4.5.1.7. GETMSE

Retorna una imagen del error cuadrático medio (MSE) para la región de 3°x 3° alrededor de las coordenadas dadas, para un nivel de zoom en particular. Este comando es utilizado por BingMaps.

Este comando acepta los siguientes parámetros:

- **lat**: Latitud del centro de la región.
- **lon**: Longitud del centro de la región.
- **z**: Nivel de zoom deseado.
- **range_index**: Índice del rango de acumulación deseado.
- **date**: Fecha para la cual se desea obtener el MSE.

4.5.1.8. GETANALOGVIEWER

Obtiene la información necesaria para visualizar el Analog Viewer en el cliente Web. Este comando acepta los siguientes parámetros:

- **lat**: Latitud del centro de la región.
- **lon**: Longitud del centro de la región.
- **z**: Nivel de zoom deseado.
- **range_index**: Índice del rango de acumulación deseado.

El comando retorna una estructura con la siguiente información, en formato JSON.

- **min_forecast_value**: Mínimo valor de precipitación acumulada pronosticada (en mm) para todos los pronósticos numéricos análogos.
- **max_forecast_value**: Máximo valor de precipitación acumulada pronosticada (en mm) para todos los pronósticos numéricos análogos.
- **min_observation_value**: Mínimo valor de precipitación acumulada observada (en mm) en las fechas de los pronósticos análogos.
- **max_observation_value**: Máximo valor de precipitación acumulada observada (en mm) en las fechas de los pronósticos análogos.
- **min_bias_value**: Mínimo valor de bias (en mm).
- **max_bias_value**: Máximo valor de bias (en mm). El bias se define como la diferencia entre el valor de precipitación pronosticado y el observado.
- **min_mean_forecast_value**: Menor valor (en mm) de precipitación pronosticada media.
- **max_mean_forecast_value**: Mayor valor (en mm) de precipitación pronosticada media.

- ***min_mean_observation_value***: Menor valor (en mm) de precipitación observada media.
- ***max_mean_observation_value***: Mayor valor (en mm) de precipitación observada media.
- ***analog_images***: Listado de Análogos con sus respectivas fechas, MSE, y ruta a las imágenes de pronóstico numérico y observaciones.

El siguiente es un ejemplo de respuesta del comando, en formato JSON.

```
{
  min_forecast_value : "0.093750",
  max_forecast_value: "58.023438",

  min_observation_value : "0.000000",
  max_observation_value : "268.620392",

  min_bias_value : "-30.000000",
  max_bias_value : "0.000000",

  min_mean_forecast_value : "11.971157",
  max_mean_forecast_value: "20.341810",

  min_mean_observation_value : "16.509752",
  max_mean_observation_value : "47.431324",

  images: [
    {
      date: "2003042600",
      filename: "analog_region_50_2003042600.png",
      mse: "16.203133"
    },
    {
      date: "2003042700",
      filename: "analog_region_51_2003042700.png",
      mse: "15.094675"
    },
    ...
  ]
}
```

4.5.2. Detalles de Implementación

A continuación se explicarán algunos detalles internos de la implementación de Albero. Esta sección resultará útil a todo aquel que desee involucrarse en el desarrollo, o realizar modificaciones a través de un fork. El código y los archivos necesarios para la instalación, así como esta guía en inglés, se encuentra disponible en

<https://github.com/leandropelorusso/albero>.

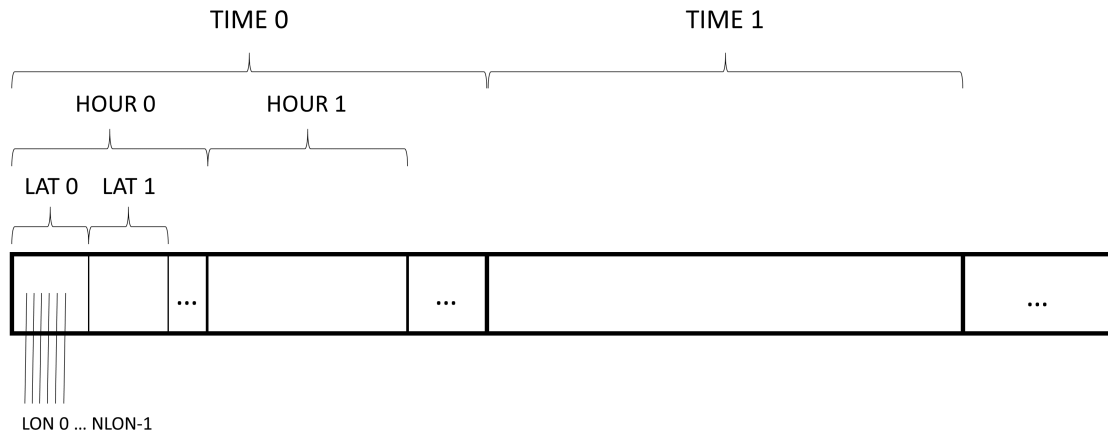
El algoritmo implementado por Albero se compone de tres secciones claramente definidas: Read Historic Forecast (lectura de datos de reforecast), Calculate Analogs (cálculo de análogos), y Calculate Probabilistic Forecast (cálculo de pronóstico probababilístico).

4.5.2.1. LECTURA DE DATOS DE REFORECAST (READHISTORICFORECAST)

El método ***ReadHistoricForecast()*** invoca la lectura de datos históricos de reforecast, así como el reforecast actual. Llamaremos reforecast actual (o Current Forecast) a aquel para el cual deseamos calcular el pronóstico probabilístico. Los archivos de reforecast contienen datos de precipitación acumulada ordenados por Latitud, Longitud, Tiempo (date) y Hora. Los mismos son almacenados en formato NetCDF, como se muestra a continuación:

$NTIME \times NFHOUR \times NLAT \times NLON$

Donde *NTIME* representa la cantidad de fechas para las cuales se poseen datos de reforecast, *NFHOUR* la cantidad de horas de pronóstico, y *NLAT* y *NLON* la cantidad de latitudes y longitudes respectivamente.



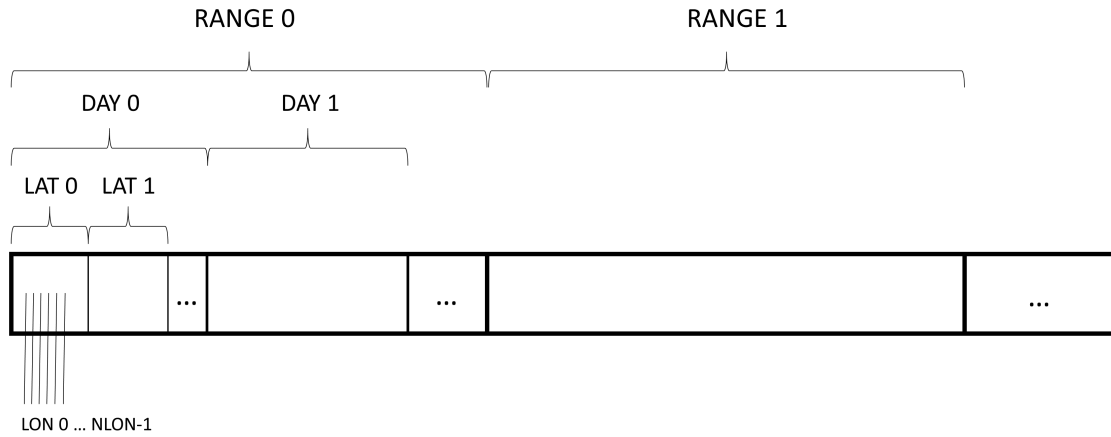
Los datos de precipitación del pronóstico actual (Current Forecast) son obtenidos localizando el comienzo de los datos para el atributo *Time* correspondiente.

El algoritmo requiere reforecasts históricos únicamente para fechas alrededor del pronóstico actual en años anteriores, por lo que una ventana temporal (configurada en 90 days) es extraída y almacenada en memoria para cada año disponible. Los datos leídos son agrupados por rango de acumulación, en una estructura de memoria contigua, de la siguiente manera:

$NRANGE \times NDAY \times NLAT \times NLON$

Donde *NRANGES* representa la cantidad de rangos de acumulación configurados por el usuario, *NDAY* la cantidad de días dentro de las ventanas temporales extraídas, y *NLAT* y *NLON* la cantidad de latitudes y longitudes respectivamente.

Los datos de precipitación se almacenan en un bloque de memoria contigua, utilizando una estructura adicional a modo de índice, lo cual permite la localización de valores para un rango de acumulación y día en particular de manera eficiente.



4.5.2.2. CÁLCULO DE ANÁLOGOS (CALCULATEANALOGS)

El método `CalculateAnalog()` realiza la búsqueda de fechas para las cuales se ha pronosticado valores similares para los del pronóstico actual, para todos los rangos de acumulación.

Para cada rango de acumulación, el método busca pronósticos análogos para cada región de $1^\circ \times 1^\circ$. Si bien las regiones están compuestas por áreas de $1^\circ \times 1^\circ$, se considera la región de $3^\circ \times 3^\circ$ que la contiene.

Los análogos son computados buscando aquellos reforecast históricos con menor error cuadrático medio (MSE), en los 16 puntos de la grilla (como se explica en el detalle del método). Los análogos son almacenados por rango de acumulación, latitud y longitud, asociados al punto superior izquierdo de la región central (el punto rojo de la imagen Figura 4.11).

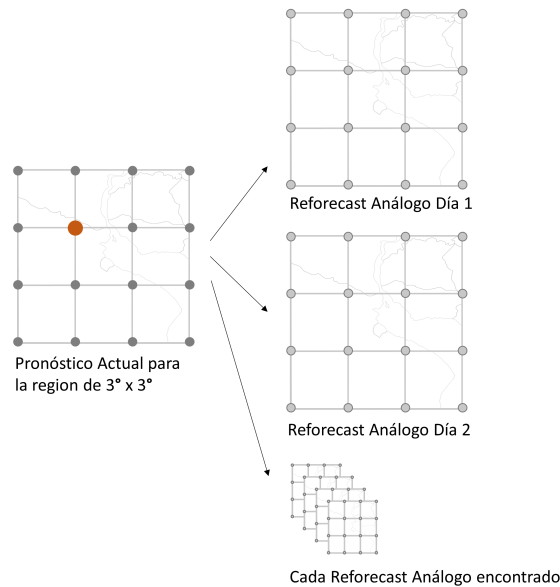


Fig. 4.11: Para obtener los análogos de una región de $1^\circ \times 1^\circ$, la región de $3^\circ \times 3^\circ$ que la contiene es comparada con la misma región en cada uno de los días en la ventana temporal de los años disponibles.

Los análogos obtenidos son almacenados en memoria, ordenados de la siguiente manera:

`NRANGE x NANALOGS x NLAT x NLON`

Donde *NRANGE* representa la cantidad de rangos de acumulación, *NANALOGS* la cantidad de análogos a considerar, y *NLAT* y *NLON* la cantidad de latitudes y longitudes respectivamente.

Cada análogo encontrado es representado por la estructura *AnalogIndex*.

```
struct AnalogIndex{
    float mse; // error cuadrático médio
    int lat; // latitud del análogo
    int lon; // longitud del análogo
    int date; // fecha del análogo
};
```

Información adicional sobre el análogo podría almacenarse en caso de que otras variables fueran consideradas en el proceso de elección.

4.5.2.3. CÁLCULO DE PRONÓSTICOS PROBABILÍSTICOS (CALCULATEPROBABILISTICFORECAST)

Este método calcula el pronóstico probabilístico para cada rango de acumulación de precipitación y threshold elegido por el usuario.

Cada región de $1^\circ \times 1^\circ$ es dividida en 5×5 puntos (*OBSERVATIONS_PER_PANEL*). Un valor de probabilidad de precipitación es entonces calculado para cada punto, comparando las precipitaciones observadas con los umbrales definidos por el usuario en cada fecha análoga encontrada. Este cálculo se realiza, en un primer paso, por región.

Para cada una de las regiones de $1^\circ \times 1^\circ$ se calcula un pronóstico probabilístico para el área de $3^\circ \times 3^\circ$ que la contiene, utilizando los análogos asociados con el punto superior izquierdo de dicha región (el punto rojo de la imagen Figura 4.11). Esto significa que, por cada punto, contaremos con 9 probabilidades distintas. Dado que los análogos para regiones adyacentes pueden diferir significativamente, las 9 probabilidades son combinadas de acuerdo a la distancia al centro de la región. Diremos que cada punto tiene 9 dimensiones, dado que contará con 9 conjuntos de observaciones distintas. Una dimensión por conjunto de observaciones.

Calculamos la probabilidad de precipitación mayor a cierto umbral para *probability_map_height* x *probability_map_width* puntos, dimensiones definidas como:

```
probability_map_height = ((NLAT - 1) * (OBSERVATIONS_PER_PANEL-1)) + 1;
probability_map_width = ((NLON - 1) * (OBSERVATIONS_PER_PANEL-1)) + 1;
```

Donde *OBSERVATIONS_PER_PANEL* representa la cantidad de puntos a considerar dentro de un área de $1^\circ \times 1^\circ$, y se encuentra definido como 5.

Comenzamos el proceso definiendo un arreglo de observaciones para cada punto, esto es, *NANALOGS_PER_LAT_LON + 1* análogos para una de las 9 dimensiones. Este arreglo es llamado *point_dimension_observations*. Se le agrega un +1 a la longitud del arreglo debido a que la posición inicial será utilizada para almacenar la distancia al centro de la región de la dimensión, para utilizar luego en el algoritmo de smoothing.

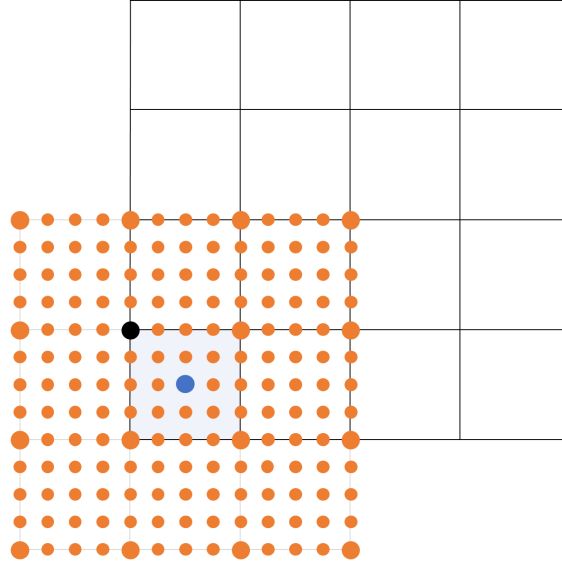


Fig. 4.12: División en 5 x 5 puntos (puntos naranjas) en el primer paso del algoritmo. El punto negro se asocia a los análogos de la región de 3°x 3°compuesta por los puntos naranjas. El punto azul denota el centro de dicha región.

El algoritmo comienza iterando cada una de las regiones de 1°x 1°, desde la esquina inferior izquierda de la grilla.

Los análogos para la región de 3°x 3°están asociados al punto negro de la figura Figura 4.12.

Para cada subregión de un grado se calcula la distancia al centro de la región (el punto azul). Luego, las observaciones para los análogos del área de puntos naranjas son obtenidos y almacenados en la estructura *region_observations*.

El siguiente paso consiste en iterar los puntos naranjas y almacenar sus observaciones en el arreglo *point_dimension_observations*, para las correspondientes dimensiones. Para cada punto, el índice de la dimensión (entre 0 y 8) es calculado como muestra la Figura 4.13.

Como primer elemento del arreglo de observaciones para un punto y dimensión dado, la distancia entre el punto y el centro de la región (el punto azul en la imagen) es almacenado. Notar que los valores en la dimensión 4 (que es la dimensión asociada a los análogos que utilizamos para obtener las observaciones) serán los de menor distancia al centro. De no implementar el algoritmo de smoothing, solo esta dimensión sería considerada.

Este proceso es repetido para cada latitud y longitud del area disponible. El resultado es un conjunto de 9 observaciones y una distancia (al centro de la región) para cada uno de los puntos en la estructura *point_dimension_observations*.

A continuación se procede a calcular finalmente el mapa de probabilidades. Recuerde, el usuario selecciona r rangos de acumulación y t thresholds, así que generaremos $t*r$ mapas de probabilidad. Cada uno de estos tendrá tantos puntos como la estructura *point_dimension_observations*.

Para cada punto en cada uno de los mapas de probabilidad, es necesario combinar las observaciones en cada una de las 9 dimensiones, para suavizar el pronóstico resultante, utilizando las distancias previamente calculadas.

Cada punto tiene asociado:

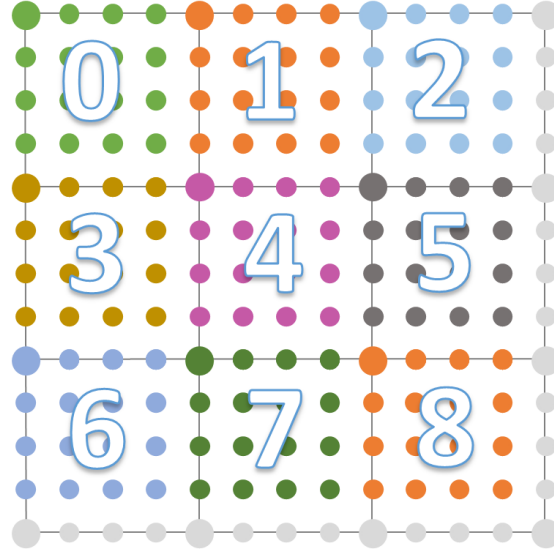


Fig. 4.13: Dimensión para cada uno de los puntos rojos.

$$dimension_0 = [d_0, v_{(0,0)}, v_{(0,1)}, v_{(0,2)}, \dots, v_{(0,n_analogos_per_lat_lon)}]$$

$$dimension_1 = [d_1, v_{(1,0)}, v_{(1,1)}, v_{(1,2)}, \dots, v_{(1,n_analogos_per_lat_lon)}]$$

\vdots

$$dimension_8 = [d_8, v_{(8,0)}, v_{(8,1)}, v_{(8,2)}, \dots, v_{(8,n_analogos_per_lat_lon)}]$$

Donde d_i es la distancia al centro de la región.

El próximo paso consiste en convertir cada distancia a un peso no normalizado, utilizando la siguiente formula:

$$w'_x = \begin{cases} \frac{D-d_x}{D+d_x} & \text{if } d_x < D \\ 0 & \text{if } d_x \geq D \end{cases}$$

Donde $x \in [0, 8]$ y D es la mayor distancia encontrada. Una vez calculados los pesos desnormalizados, se normalizan de forma tal que la suma de los pesos sea 1,0

$$w_x = \frac{w'_x}{w'_0 + w'_1 + w'_2 + w'_3 + w'_4 + w'_5 + w'_6 + w'_7 + w'_8}$$

Luego de normalizar los pesos tenemos, para cada punto:

$$dimension_0 = [w_0, v_{(0,0)}, v_{(0,1)}, v_{(0,2)}, \dots, v_{(0,n_analogos_per_lat_lon)}]$$

$$dimension_1 = [w_1, v_{(1,0)}, v_{(1,1)}, v_{(1,2)}, \dots, v_{(1,n_analogos_per_lat_lon)}]$$

⋮

$$dimension_8 = [w_8, v_{(8,0)}, v_{(8,1)}, v_{(8,2)}, \dots, v_{(8,n_analogs_per_lat_lon)}]$$

A continuación, para cada rango de acumulación r y threshold t , consideramos cada punto y calculamos la probabilidad como:

$$\sum_{i=0}^8 w_i * p_i(t)$$

Donde $p_i(t)$ es la probabilidad de encontrar una observación mayor que t , definida como:

$$\frac{100 * c(t)}{n_analogs_per_lat_lon}$$

Donde $c(t)$ es la cantidad de observaciones mayores a t .

4.5.3. BingMaps Tiling

Albero utiliza Microsoft BingMaps como herramienta de visualización del Map View, por lo que resulta necesario contar con la habilidad de generar y retornar múltiples regiones (tiles) del pronóstico probabilístico de manera rápida y eficiente. VIDa resuelve esta necesidad por medio de un server multithread. El server ofrece su interfaz por medio de Sockets (C++ Linux Sockets), por medio de la cual el cliente realiza pedidos especificando la tarea a realizar (por ejemplo, generar la imagen del pronóstico probabilístico para una latitud y longitud en particular, con cierto nivel de zoom específico). El pedido es encolado, de forma de asegurar la ejecución asíncrona, permitiendo la recepción de pedidos futuros.

El server cuenta con un thread encargado de monitorear la cola de pedidos. Al encontrarse una tarea, un nuevo proceso (thread) es creado, y el trabajo es asignado, de forma de permitir la generación de múltiples tiles de forma simultánea. Cuando el pedido ha sido completado, el controlador retorna la respuesta al cliente, utilizando el socket original, el cual es entonces cerrado. En la figura Figura 4.14 puede verse un diagrama de secuencias, ilustrando la atención de un pedido de generación de tile.

En algunas circunstancias, los mapas de BingMaps utilizados por defecto, los cuales incluyen topografía, nombres, ríos, lagos y otros elementos, pueden interferir con la visualización de las probabilidades u observaciones. Por este motivo, Albero permite la utilización de servidores de tiles adicionales, compatibles con el formato de OpenStreetMap. Se ha diseñado una configuración de CartoCSS [33], la cual es suministrada a través de TileMill [34], que permite visualizar mapas compuestos únicamente de fronteras internacionales y provincias. Esto permite una visualización clara de las escalas de colores, proporcionando un marco espacial por medio del mapa de contornos. Un ejemplo de esta configuración puede apreciarse en la Figura 4.15.

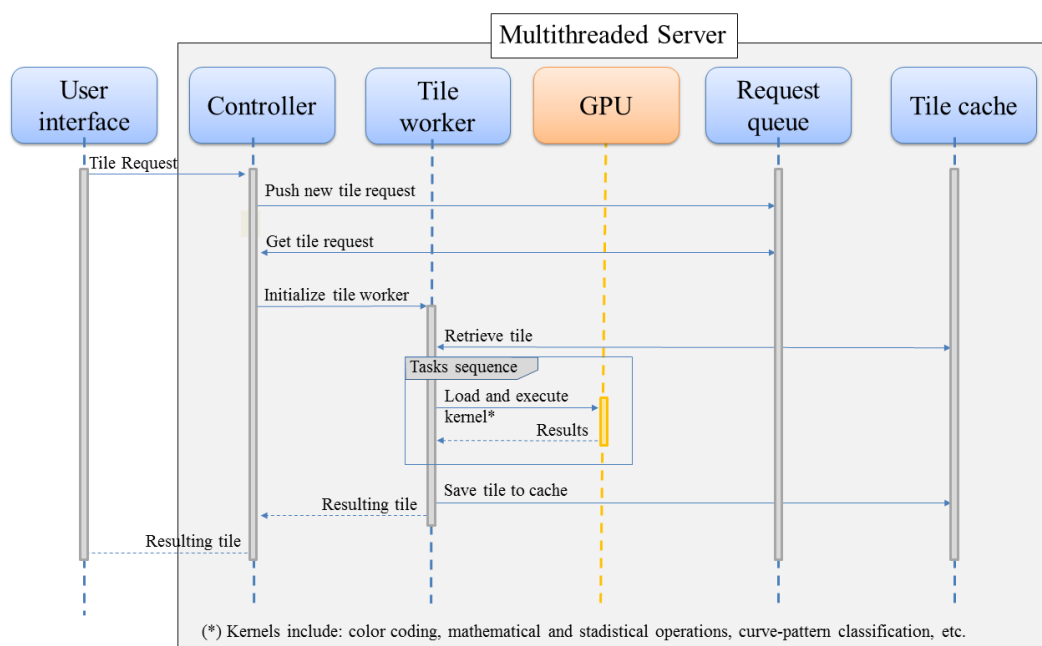


Fig. 4.14: VIDA: Atención de un pedido de generación de tile.

4.5.4. Manipulación de Datos Observacionales

Albero utiliza una base de datos histórica de 30 años de ensambles de GFS, y observaciones de CMORPH. La información de reforecast se encuentra disponible en una grilla de 1° de resolución, como el total de precipitación acumulada en el periodo compuesto por las últimas tres o seis horas. Los datos observacionales CMORPH están disponible como una grilla de precipitación estimada, con una resolución espacial de $0,25^\circ$, cada tres horas.

La resolución del pronóstico probabilístico está dada por la resolución de las observaciones de CMORPH, es decir, $0,25^\circ$. Mediante el proceso denominado Statistical downscaling, nuestra solución visualiza los pronósticos probabilísticos con una resolución de $0,25^\circ$, en lugar de la resolución de 1° provisto por los pronósticos numéricos.

Los datos observacionales CMORPH son proporcionados por la Climate Prediction Center como un archivo por día. El nombre de los archivos siguen el formato:

YYYY/cmorph/3hs-025deg.YYYYMMDD.nc

El proceso de calculo de pronósticos probabilísticos requiere acceder a observaciones de múltiples días distintos. El proceso de apertura de un archivo por día lleva a tiempos de Entrada/Salida extremadamente lentos, incluso en discos de estado sólido (SSD). Para mejorar los tiempos de lectura de datos observacionales, los archivos son procesados y compilados en único archivo binario, agrupando por periodos de acumulación de 6 horas (el mínimo periodo de acumulación permitido por Albero).

En conjunto con el archivo de datos, un índice en formato de texto plano es generado. Este índice especifica rangos de latitud y longitud encontrados en el archivo, así como el comienzo, en bytes, de la región de datos para un día en particular.

El siguiente es un extracto del archivo índice utilizado actualmente:

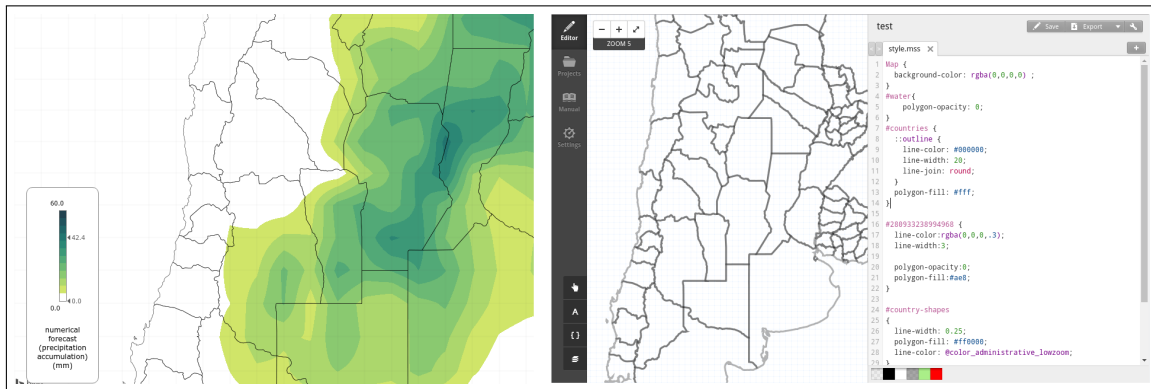


Fig. 4.15: Ejemplo de integración con TileMill. En la izquierda, Albero utilizando el mapa generado por TileMill, configurado para mostrar únicamente fronteras internacionales y provincias. En la derecha, interfaz web de TileMill y el CartoCSS utilizado.

```

init_lat=-57.125000
end_lat=-18.875000
init_lon=282.875000
end_lon=309.125000
20021207=0
20021208=261184
20021209=522368
20021210=783552
...

```

El índice generado es almacenado en memoria para su rápido acceso. Este mecanismo permite abrir el archivo de datos una única vez, lo cual expone mejoras en tiempos de lectura de datos de un 82 %.

5. CASOS DE ESTUDIO

Para la evaluación, hemos realizado dos entrevistas guiadas con siete expertos del dominio de diferentes orígenes, recolectando sus comentarios. Los casos de estudio centran en el análisis de la variable de precipitación acumulada.

Ejemplificaremos las capacidades de Albero por medio de dos escenarios que cubren el pronóstico meteorológico operacional y el análisis retrospectivo, para un evento de precipitaciones extremas en un corto periodo de tiempo. Este evento tuvo lugar del 1 al 3 de abril de 2013, causando inundaciones en diferentes zonas del nordeste de nuestro país. Los mismos siete expertos de dominio evaluaron estos escenarios.

5.1. Albero para pronóstico operacional

En este caso, los expertos de dominio desean utilizar las características de Albero para generar pronósticos probabilísticos. Quieren proporcionar una estimación más precisa de la incertidumbre de los pronósticos, que puede resultar en una mejor toma de decisiones para la comunidad.

En este escenario, un pronosticador establece la fecha válida a partir del 1° de abril de 2013, un plazo de 72 horas, intervalos de 24 horas, y varios umbrales para el que desea crear los pronósticos probabilísticos. Se configuran los umbrales por encima de 20 mm, 40 mm, 60 mm y 80 mm. Otros parámetros, tales como el número de análogos y el rango de temporal de la base de datos históricos, son preestablecidos por un meteorólogo experto en la técnica. A continuación, un resumen de los diferentes umbrales, agrupados por intervalo de tiempo, se muestra en las columnas Übermap (véase Figura 4.4).

El usuario quería detectar las áreas con mayor probabilidad de acumulación de precipitación. El verde y el azul en la segunda fila y segunda columna del Übermap indica que existen altas probabilidades de precipitaciones superiores a 20 mm durante el segundo día para la Pampa y la Mesopotamia. Tonos verdes en la tercera fila de la Übermap indica que es son probables fuertes precipitaciones sobre la Pampa durante las próximas 72 hs. El usuario notó que durante el segundo intervalo de tiempo, correspondiente a la segunda jornada, se pronostica fuertes lluvias. La Figura (véase Figura 5.1) muestra un acercamiento de los pronósticos probabilísticos para el segundo intervalo de tiempo ([24-48 hs), en los umbrales: 20 mm, 40 mm, 60 mm y 80 mm.

El usuario también analizó los pronósticos probabilísticos por encima de 60 mm, y el pronóstico determinístico, detectando probabilidades de precipitación alrededor del 50 % en la Mesopotamia (Figura 5.2). Estas probabilidades son varias veces más grandes que la probabilidad climatológica de precipitación acumulada, indicando que la ocurrencia de un evento de precipitación extrema es mayor que en otros días.

Utilizando el Übermap, el pronosticador puede elegir visualizar la distribución espacial de las MSE para un determinado rango de acumulación. El MSE se calcula como la diferencia entre los pronósticos deterministas y los análogos para cada una de las sub-regiones. Esta es una medida de la exactitud de la técnica para la situación particular que el usuario está evaluando. La Figura 5.1d muestra una visualización de la distribución espacial de MSE. Los errores de mayor magnitud se encuentra en la parte central y noreste del país, que corresponde a la Pampa y la Mesopotamia. Éstas se muestran en tonos rojos

en la figura. Con Albero, el pronosticador fue capaz de identificar rápidamente dos regiones (Este y Noroeste del país), donde la probabilidad de ocurrencia de fuertes precipitaciones es alta, a causa de la precipitación sostenida durante un período de 3 días. El usuario llegó a esta conclusión de manera mucho más rápida que con las herramientas anteriores, con el agregado de una mejor estimación de la incertidumbre.

5.2. Albero para optimización de la técnica

En este escenario el meteorólogo analiza eventos del pasado. El usuario sabía que el evento seleccionado fue subestimado por el modelo numérico, ya que se trató de una situación muy extrema y, en general, la precipitación es una de las variables meteorológicas más difíciles de predecir.

Utilizando el Analog Viewer, analizó el pronóstico numérico, los análogos y las observaciones. La diferencia entre los análogos y las observaciones, el sesgo, aporta información acerca de los errores sistemáticos del modelo asociado con situaciones que son similares a la pronosticada por el modelo determinístico. De esta manera, el analista comparó si el pronóstico numérico subestimó o sobrestimó la acumulación de precipitación a lo largo de los tres días, analizando el sesgo entre las media de los análogos y las observaciones, detectando grandes diferencias entre ellos.

La figura 12 muestra que el pronóstico numérico sistemáticamente subestimó la cantidad de precipitación acumulación en la sub-región seleccionada, dado que la observación media es superior a la media pronosticada por los análogos. Esto también puede observarse en los análogos individuales y sus correspondientes observaciones.

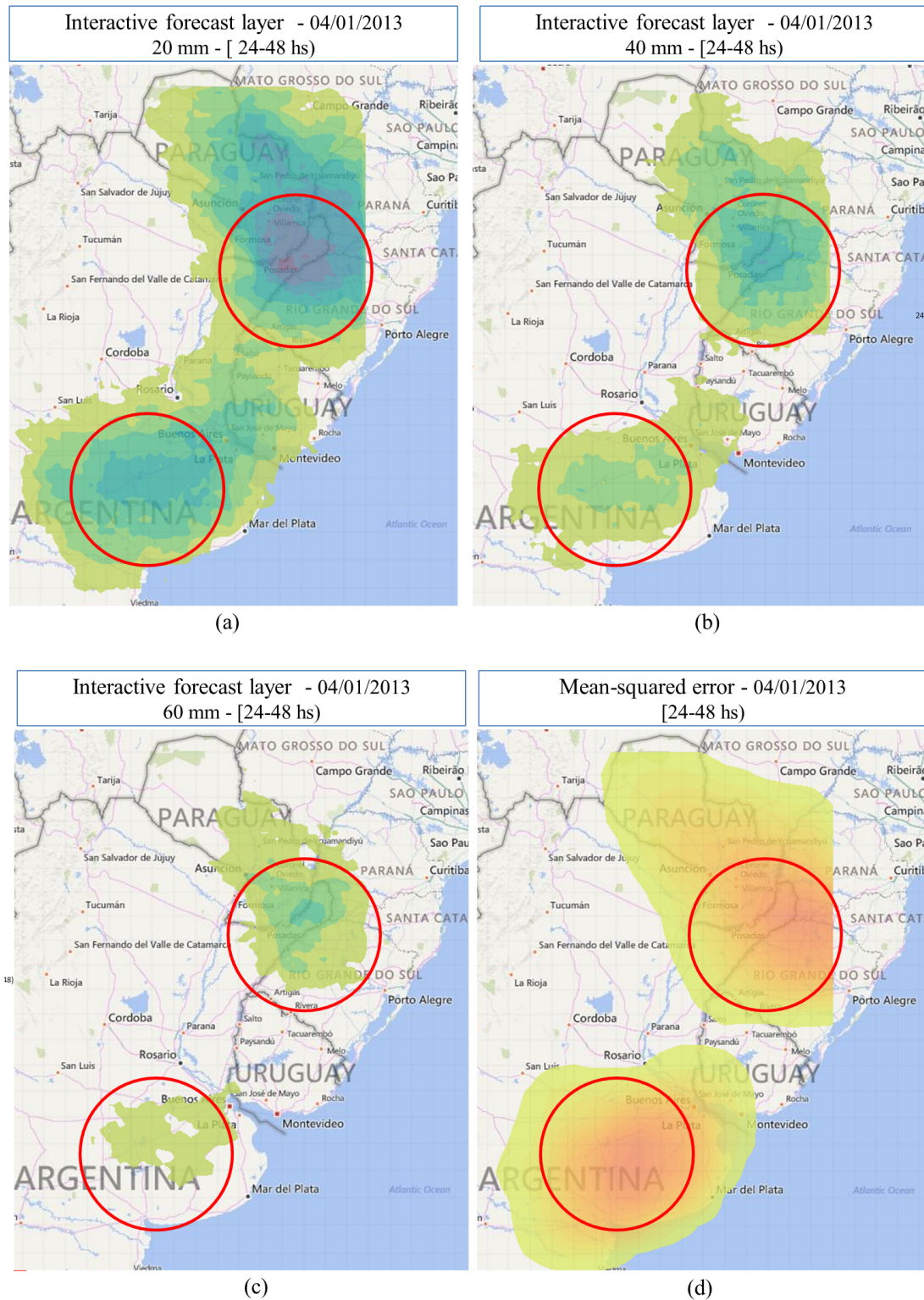


Fig. 5.1: Una captura de Albero mostrando algunos resultados de la técnica RAR, para un conjunto dado de parámetros de configuración.

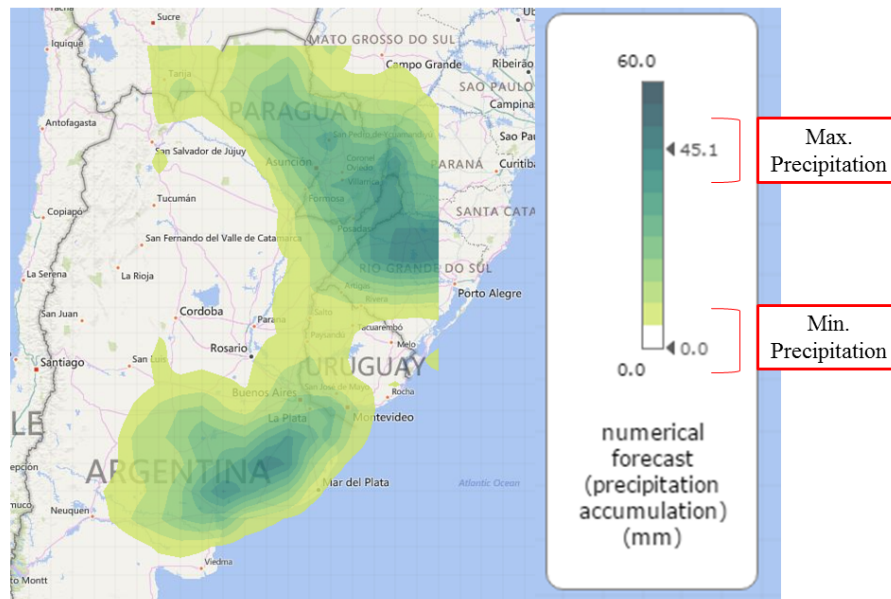


Fig. 5.2: Interactive Layer Forecast mostrando el pronóstico determinístico para el intervalo de [24-48 hs), válido para el mes de Abril, 1st, 2013. El usuario encuentra que el pronóstico numérica subestima la cantidad precipitación acumulada dado que los valores de precipitación más altos son inferiores a 45,1 mm.

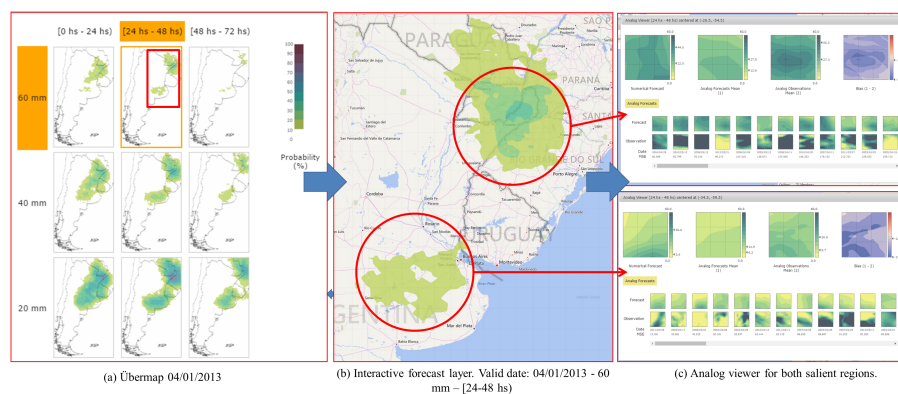


Fig. 5.3: Pantalla de Albero mostrando las precipitaciones extremas ocurridas durante el primer y tercer día del mes de Abril de 2013. (a) El usuario identifica un pronóstico probabilístico con una gran área de precipitación acumulada por encima de 60 mm. (b) El usuario hace zoom-in y selecciona un área de interés. (c) el usuario considera que el pronóstico numérico sistemáticamente subestima la cantidad de precipitación acumulada, dado que la observación media es superior a la media de los análogos.

6. LECCIONES APRENDIDAS Y CONCLUSIONES

6.1. Lecciones Aprendidas

Las lecciones aprendidas cubren el workflow de tareas, su proceso de diseño, y la evaluación por parte del usuario.

6.1.1. Proceso de Diseño

Este trabajo ha sido estructurado en tres fases: inepción, diseño y evaluación. El objetivo de la primer fase fue el de generar ideas de diseño y contenido, obtener un mejor entendimiento de nuestro publico apuntado, y entender y refinar los requerimientos de nuestros expertos del dominio. La fase de diseño consistió en obtener información detallada de los requerimientos por parte de nuestros expertos del dominio, así como discutir prototipos y elaborar y validar las decisiones de diseño de manera conjunta. Esta fase fue llevada a cabo siguiendo un proceso iterativo e incremental.

Durante la fase de inepción trabajamos principalmente con uno de nuestros expertos, considerado el principal stakeholder. En las múltiples reuniones llevadas a cabo, discutimos el estado del arte de las prácticas y técnicas de visualización, tales como seeking mantra, técnicas de focus+context, linking and brushing, y lentes interactivos, entre otras. El resultado de esta fase fue la realización de que las herramientas de visualización existentes no satisfacían completamente las necesidades de los expertos del dominio.

Durante la fase de diseño se llevaron a cabo diversas reuniones con todos los expertos del dominio, no solo con el principal stakeholder. Diseñamos un primer prototipo, el cual fue revisado y probado por nuestro principal stakeholder, quien es co-autor de las publicaciones relacionadas a Albero, y co-director de este trabajo. El prototipo presentaba una primera versión de visualizaciones de pronósticos probabilísticos, así como vistas enlazadas e histogramas. El experto del dominio nos entregó un feedback positivo, y expresó la necesidad de un workflow interactivo, siendo claro en el requerimiento: el workflow debe permitir al meteorólogo una configuración flexible de los parámetros de la técnica RAR, de forma tal de realizar y evaluar distintos escenarios. Estas reuniones llevaron a una mejora constante del workflow y las funcionalidades ofrecidas, conforme se avanzaba con el proceso iterativo e incremental. Respondimos a los requerimientos permitiendo una parametrización flexible de la técnica RAR, la cual permite al usuario configurar parámetros dependientes del tipo de análisis que se desea realizar, ya sea con fines de pronóstico operativo, o de optimización de la técnica. Esto dió lugar al primer ciclo de trabajo, el ciclo de parametrización.

Una vez diseñado el ciclo de parametrización, todos nuestros expertos del dominio fueron involucrados en el proceso de diseño. Continuamos con una serie de sesiones con todos ellos, discutiendo nuevos prototipos, lo que nos permitió diseñar el segundo ciclo de trabajo, el ciclo de pronóstico probabilístico. Estos prototipos evolucionaron para incluir vistas enlazadas y coordinadas, entre ellas, una vista combinada de múltiples pronósticos probabilísticos, para diversos rangos temporales y umbrales de precipitación. Sugerimos a los expertos del dominio la posibilidad de realizar un análisis mas detallado de la técnica RAR, proveyendo una vista de los principales análogos, utilizando por ejemplo el MSE para elegir los análogos mas significativos. Los expertos del dominio prefirieron la opción de

visualizar todos los análogos, junto con información estadística relacionada, para permitir un mejor análisis de la performance del método. De esta forma, ellos podrían recolectar información que pudiera ser utilizada en la optimización del algoritmo. Propusieron además visualizar un mapa de la distribución espacial del MSE de un ensamble de análogos, argumentando que ayudaría a la evaluación del método en situaciones particulares. Las decisiones de diseño fueron finalmente validadas y bien aceptadas. Esto llevó al último ciclo de trabajo, el ciclo de análogos. Por medio del visor de análogos, los expertos del dominio pueden visualizar y obtener información estadística de los análogos seleccionados, permitiendo así, por ejemplo, detectar sesgos sistemáticos en el modelo numérico, comparando pronósticos y observaciones para las fechas análogas escogidas. Concluimos nuestras reuniones cuando el diseño propuesto cumplió en satisfacer los requerimientos de nuestros colaboradores. Como resultado de este proceso, construimos Albero, con eficiencia y escalabilidad en mente. La solución soporta un análisis interactivo de grandes volúmenes de información, proveyendo resultados en tiempo real.

Los meteorólogos destacaron el potencial de nuestra herramienta. Uno de ellos expresó: “Hay nueva información disponible. Anteriormente sólo se podía procesar un pronóstico probabilístico por tarea, e intentar asociarlo con gráficas de otras herramientas como GrADS”. Otro experto dijo: “Ahora, con Albero tenemos la posibilidad de seleccionar diferentes tiempos y rangos de acumulación, lo cual no era posible con nuestras herramientas anteriores, de forma interactiva y sin recurrir a la programación o sistemas de scripting”. Los expertos de dominio aprendieron, por medio de la visualización e interacción con el ensamble de análogos y las demás vistas vinculadas, que para algunas regiones y eventos específicos los análogos pueden tener marcadas diferencias. Estas son observaciones subjetivas y necesitan una validación objetiva. También sugirieron ideas para futuros trabajos: validación objetiva de la técnica mediante puntajes (skill score), estadísticas y algoritmos de machine learning para la optimización de los parámetros de la técnica. Estamos actualmente implementando un entorno cooperativo con Albero. Además, la solución estará disponible públicamente. Esto nos permitirá realizar una prueba futura exhaustiva de nuestra solución.

6.2. Conclusiones

Este trabajo de tesis presenta un nuevo enfoque de análisis visual para el análisis de pronósticos probabilísticos en el contexto de la técnica de RAR que expone los componentes internos del algoritmo de Hamill y Whitaker y permite explicar el rendimiento del algoritmo basado en los resultados intermedios. Hemos diseñado nuestra solución tras un proceso iterativo y participativo para satisfacer las necesidades de nuestros expertos del dominio en cada uno de los bucles del flujo de trabajo propuesto. Hemos validado el diseño utilizando dos escenarios diferentes, preparado en colaboración con nuestros expertos del dominio. Los meteorólogos evaluaron nuestra solución y nos dieron una respuesta positiva. Al mejor de nuestro conocimiento, este es el primer sistema de análisis visual para el análisis de pronósticos probabilísticos basadas en la técnica de RAR. Nuestro sistema es completamente open-source (<https://github.com/leandropelorosso/albero>), en la actualidad está implementado e instalado en el CIMA, en una fase de testeo pre productiva. La plataforma de desarrollo es completamente extensible permitiendo su extensión en futuros trabajos de investigación y desarrollo.

Futuros trabajos incluyen el diseño de un nuevo módulo para la validación técnica

RAR que será capaz de calcular medidas de calidad, conocida como la “habilidad”, utilizando diferentes puntajes de verificación. Se considera además la inclusión de distintas alternativas de la función de similitud utilizada por la técnica RAR, la adición de variables meteorológicas adicionales, y nuevas visualizaciones tales como histogramas, y resúmenes de los valores extremos. Además, un nuevo módulo será desarrollado para la optimización de parámetros, por ejemplo, utilizando algoritmos genéticos o redes neuronales, para optimizar la parametrización para una determinada región y situación. Esto reducirá el tiempo destinado por los investigadores para optimizar la técnica RAR en dicha situación. En la actualidad, los niveles de datos disponibles a los investigadores están aumentando en volumen, velocidad de recolección, variedad y veracidad. Albero es un aporte para el análisis visual de datos meteorológicos altamente complejos, provenientes de simulaciones y observaciones, para ayudar a la toma de decisiones bajo incertidumbre.

Índice de figuras

1.1. Diagrama de proceso de Visual Analytics. Adaptado de <i>Visual Analytics: Scope and Challenges</i> . [2]	2
1.2. Modelo de generación de conocimiento. Adaptado de <i>Knowledge Generation Model for Visual Analytics</i> . [3]	4
1.3. Flujo de trabajo automático actual. Proporciona una flexibilidad reducida, y la necesidad de generar múltiples imágenes de manera individual.	6
1.4. Nuevo flujo de trabajo interactivo. Proporciona tres ciclos de análisis anidados para transmitir conocimientos sobre los pronósticos probabilísticos y la técnica RAR, para permitir una mejor toma de decisiones y la optimización de la técnica.	6
2.1. Visualización del clima con ViSUS. Izquierda: el framework provee una visualización del ensamble de simulación de cambio de temperatura. Centro: dos conjuntos de datos de diferente resolución espacial mostrando la precipitación a nivel global. Derecha: simulación del clima mostrando las temperaturas globales y precipitación utilizando dieciséis modelos diferentes, fundidos utilizando una función de desviación standard definida por el usuario.	10
2.2. El framework Ultrascale Visualization Climate Data Analysis Tools (UVCDAT). Uniendo múltiples protocolos y APIs, el framework permite a los científicos visualizar e interpretar datos climatológicos y observaciones en múltiples modelos.	11
2.3. Visualización de un conjunto de datos oceánicos ICON utilizando ParaView. En el panel de la izquierda se muestran los datos, las selección realizada y la topografía de la Tierra. Los tres viewports son utilizados para especificar la selección, basada en una matriz de scatterplot y coordenadas paralelas.	12
2.4. Visit implementa una arquitectura cliente servidor, permitiendo así trabajar con grandes conjuntos de datos, y cortos tiempos de respuesta.	13
2.5. Visit permite al usuario trabajar con un amplio rango de casos de uso.	14
2.6. Visualización interactiva en COVISE de la interacción entre la superficie terrestre y las precipitaciones en el Oeste de Africa.	16
2.7. Ensemble-Vis provee una plataforma de visualización y análisis de datos a través de la combinación de múltiples técnicas de visualización estadística y un alto nivel de interacción con el usuario.	17
2.8. ArcGIS mostrando la precipitación anual observada en 2014, para la región de la República Argentina.	18
2.9. GRADS. Pronóstico diario de precipitación para la región de Sudamérica.	19
2.10. IDV. Mapas de precipitación simulada cada 3 horas, sobre el distrito de Mae Sot, Tailandia, durante 24/09/2014:19.00 a 25/09/2014:10.00.	20

3.1.	Ilustración de la técnica RAR para un pronóstico de 24 hs. El pronóstico numérico se muestra en la primera fila, definido por los 16 puntos negros. El área central, delimitada por líneas punteadas, define la región para la cual se desea obtener el pronóstico probabilístico. Los cuatro pronósticos análogos mas similares se muestran en la segunda fila. Las precipitación acumulada observada para los análogos elegidos se presentan en la tercer fila. Finalmente, tres pronósticos probabilísticos, para umbrales de 15 mm, 25 mm y 35 mm se muestran en la última fila.	24
3.2.	El área es dividida en subregiones de $1^\circ \times 1^\circ$, para los cuales se calcularán pronósticos probabilísticos locales. Posteriormente, los pronósticos serán unidos para conseguir un pronóstico global.	25
3.3.	Estrategia de armado de pronóstico probabilístico en función de la validez de un evento dado en los distintos puntos de la grilla. En el algoritmo básico de RAR solo se considera la región central de $1^\circ \times 1^\circ$	25
3.4.	Estrategia de armado de pronóstico probabilístico en función de la validez de un evento dado en los distintos puntos de la grilla, utilizada para el método de suavizado, en la cual se calculan probabilidades para toda el área de $3^\circ \times 3^\circ$	26
3.5.	Ilustración del algoritmo de smoothing. En este caso se buscan las probabilidades en los puntos naranjas. Los nueve paneles son las nueve regiones para las cuales análogos han sido calculados, que se sobreponen con la grilla de puntos naranjas. Los puntos restantes se denotan con color azul. Los análogos son calculados por similaridad en los puntos negros de mayor tamaño; el centro de cada región análoga se denota con una cruz roja. Para el punto denotado por el rectángulo verde, la probabilidad final es calculada como la suma ponderada de los nueve estimados.	27
4.1.	Visualización de pronósticos de precipitación probabilísticos del Servicio Meteorológico Nacional. Creados usando GrADS y Octave [?].	29
4.2.	El diagrama muestra cómo Alberó se integra en el nuevo workflow interactivo. Distintas vistas vinculados son coordinados en un fluido diálogo entre el usuario y la aplicación, para ayudar al usuario en el proceso de toma de decisiones y la optimización de la técnica RAR	30
4.3.	Dominio para el cual se obtuvo el set de datos de pronósticos retrospectivos en el período 2002-2016, definido por 76°W - 52°W y 56°S - 20°S	31
4.4.	Übermap. La visualización global muestra los pronósticos probabilísticos para los diversos umbrales de precipitación y rangos de acumulación. También permite visualizar el pronóstico determinístico, las observaciones y el error cuadrático medio (MSE), utilizando los botones correspondientes. . .	32
4.5.	Interactive Forecast Layer, mostrando un primer plano de: (a) pronóstico probabilístico mostrando una grilla interactivo dividida en sub-regiones utilizadas por el RAR, (b) distribución espacial del MSE (c) pronóstico numérico, (d) observaciones. Haciendo clic en una de las sub-regiones, el usuario tiene acceso a información detallada del ensamble de análogos y observaciones utilizadas en el cálculo del pronóstico para la sub-región. . . .	34

4.6.	Analog Viewer para la región seleccionada, en el segundo rango de acumulación [24 hs a 48 hs). Muestra la información utilizada internamente en el método RAR: la lista de análogos elegidos y sus observaciones, los valores medios y el sesgo, vinculada al Interactive Forecast Layer.	35
4.7.	Diferentes mapas de colores evaluados por nuestros expertos del dominio durante el proceso iterativo de refinamiento de nuestras opciones de diseño. Para representar la probabilidad de precipitación acumulada, prefirieron el mapa de colores HCL de dos matices (C) , que va desde el transparente a verde a violeta.	37
4.8.	Una captura de Alberio mostrando algunos resultados de la técnica RAR, para un conjunto dado de parámetros de configuración.	38
4.9.	Diagrama de arquitectura de VIDa (Visual Interactive Dashboard)	39
4.10.	Abstracción de las partes involucradas en la atención de pedidos. La estructura permite incrementos de funcionalidad modulares.	40
4.11.	Para obtener los análogos de una región de $1^\circ \times 1^\circ$, la región de $3^\circ \times 3^\circ$ que la contiene es comparada con la misma región en cada uno de los días en la ventana temporal de los años disponibles.	45
4.12.	División en 5×5 puntos (puntos naranjas) en el primer paso del algoritmo. El punto negro se asocia a los análogos de la región de $3^\circ \times 3^\circ$ compuesta por los puntos naranjas. El punto azul denota el centro de dicha región.	47
4.13.	Dimensión para cada uno de los puntos rojos.	48
4.14.	VIDa: Atención de un pedido de generación de tile.	50
4.15.	Ejemplo de integración con TileMill. En la izquierda, Alberio utilizando el mapa generado por TileMill, configurado para mostrar únicamente fronteras internacionales y provincias. En la derecha, interfaz web de TileMill y el CartoCSS utilizado.	51
5.1.	Una captura de Alberio mostrando algunos resultados de la técnica RAR, para un conjunto dado de parámetros de configuración.	54
5.2.	Interactive Layer Forecast mostrando el pronóstico determinístico para el intervalo de [24-48 hs), válido para el mes de Abril, 1st, 2013. El usuario encuentra que el pronóstico numérica subestima la cantidad precipitación acumulada dado que los valores de precipitación más altos son inferiores a 45,1 mm.	55
5.3.	Pantalla de Alberio mostrando las precipitaciones extremas ocurridas durante el primer y tercer día del mes de Abril de 2013. (a) El usuario identifica un pronóstico probabilístico con una gran área de precipitación acumulada por encima de 60 mm. (b) El usuario hace zoom-in y selecciona un área de interés. (c) el usuario considera que el pronóstico numérico sistemáticamente subestima la cantidad de precipitación acumulada, dado que la observación media es superior a la media de los análogos.	55

Bibliografía

- [1] Thomas, J., Cook, K.: *Illuminating the Path: Research and Development Agenda for Visual Analytics*. IEEE-Press (2005)
- [2] Keim D.A., Mansmann F., Schneidewind J., Thomas J., Ziegler H. (2008) *Visual Analytics: Scope and Challenges*. In: Simoff S.J., Böhlen M.H., Mazeika A. (eds) Visual Data Mining. Lecture Notes in Computer Science, vol 4404. Springer, Berlin, Heidelberg
- [3] Sacha D., Stoffel A., Stoffel F., Kwon B. C., Ellis G., Keim D. A. *Knowledge Generation Model for Visual Analytics*. IEEE Transactions on Visualization and Computer Graphics, 20(12):1604-1613, 2014.
- [4] Magnani L. *Abduction, Reason and Science: Processes of Discovery and Explanation*. Springer, 2001.
- [5] Peircem C. S. *Collected papers: Elements of Logic*. Volume 2, pages 56–59. Belknap Press, 1965.
- [6] T. Gneiting and A. E. Raftery. *Weather forecasting with ensemble methods*. Science, 310(5746):248–249, 2005.
- [7] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.
- [8] T. Palmer and R. Hagedorn. *Predictability of Weather and Climate*. Cambridge University Press, 2006.
- [9] T. M. Hamill and J. S. Whitaker. *Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application*. Monthly Weather Review, 134(11):3209–3229, 2006.
- [10] T. Hamill, G. Bates, J. Whitaker, D. Murray, M. Fiorino, and T. Galarneau. *A Description of the 2nd-Generation NOAA Global Ensemble Reforecast Data Set*.
- [11] R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie. *CMORPH 0.25 degree 3-hourly precipitation estimates*. NOAA CPC, 2002.
- [12] L. Aldeco, J. Ruiz, and C. Saulo. *Probabilistic forecasts during the monsoon season: the analogs technique as a tool for precipitation prediction over southeastern-south america*. In X International Conference on Southern Hemisphere Meteorology and Oceanography. American Meteorological Society, 2012.
- [13] F. J. Anscombe. *Graphs in statistical analysis*. The American Statistician, 27(1):17–21, 1973.
- [14] D. Murray, J. McWhirter, S. Wier, and S. Emmerson. *The integrated data viewer: a web-enabled application for scientific analysis and visualization*. In 19th Conference on International Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology. American Meteorological Society, 2003.

-
- [15] J. Clyne, P. Mininni, A. Norton, and M. Rast. *Interactive desktop analysis of high resolution simulations: application to turbulent plume dynamics and current sheet formation*. New Journal of Physics, 9(8):301, 2007.
- [16] P. Tsai and B. Doty. *A prototype java interface for the Grid Analysis and Display System (GrADS)*. In Proceedings of 14th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, pp. 11–16, 1998.
- [17] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., 2013.
- [18] D. Williams, C. Doutriaux, J. Patchett, S. Williams, G. Shipman, R. Miller, C. Steed, H. Krishnan, C. Silva, A. Chaudhary, P.-T. Bremer, D. Pugmire, W. Bethel, H. Childs, M. Prabhat, B. Geveci, A. Bauer, A. Pletzer, J. Poco, T. Ellqvist, E. Santos, G. Potter, B. Smith, T. Maxwell, D. Kindig, and D. Koop. *The Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT): Data analysis and visualization for geoscience data*. IEEE Computer, 99(PrePrints):1, 2013. LBNL-6278E. doi: 10.1109/MC.2013.119
- [19] Y. Song, J. Ye, N. Svakhine, S. Lasher-Trapp, M. Baldwin, and D. Ebert. *An atmospheric visual analysis and exploration system*. IEEE Transactions on Visualization and Computer Graphics, 12(5):1157–1164, Sep 2006. doi: 10.1109/TVCG.2006.117
- [20] Microsoft Research: *Microsoft FetchClimate*. <https://www.microsoft.com/en-us/research/project/fetchclimate/>. Accessed in March 2017.
- [21] *Weather Spark*. <https://es.weatherspark.com/>. Accessed in March 2017.
- [22] *The Weather Channel*. <https://weather.com/>. Accessed in March 2017.
- [23] *Wetter.de*. <https://www.wetter.de/>. Accessed in March 2017.
- [24] *Weather Underground*. <https://www.wunderground.com/>. Accessed in March 2016.
- [25] *Terra3D*. <http://www.terra3d.de/>. Accessed in March 2016.
- [26] K. Potter, P. Rosen, and C. R. Johnson. *From quantification to visualization: A taxonomy of uncertainty visualization approaches*. In Uncertainty Quantification in Scientific Computing, pp. 226–249. Springer, 2012.
- [27] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. *A typology for visualizing uncertainty*. In Electronic Imaging 2005, pp. 146–157. International Society for Optics and Photonics, 2005.
- [28] S. Deitrick and R. Edsall. *The influence of uncertainty visualization on decision making: An empirical evaluation*. Springer, 2006.
- [29] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson. *Ensemble-vis: A framework for the statistical visualization of ensemble data*. In IEEE Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes., pp. 233–240, Oct 2009.

- [30] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. *Noodles: A tool for visualization of numerical weather model ensemble uncertainty*. Visualization and Computer Graphics, IEEE Transactions on, 16(6):1421–1430, November 2010. doi: 10.1109/TVCG.2010.181
- [31] P. Salio, M. P. Hobouchian, Y. G. Skabar, and D. A. Vila. *Evaluation of high-resolution satellite precipitation estimates over southern south America using a dense rain gauge network*. Atmospheric Research, 163(S1):146–161, September 2015. doi: 10.1016/j.atmosres.2014.11.017
- [32] R. Stauffer, G. J. Mayr, M. Dabernig, and A. Zeileis. *Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations*. Bulletin of the American Meteorological Society, 37 2013.
- [33] *CartoCss* <https://carto.com/docs/carto-engine/cartocss/>. Accessed in March 2017.
- [34] *TileMill* <https://tilemill-project.github.io/tilemill/>. Accessed in March 2017.