

MODIGRAPH Y MULTIMETEO: UNA NUEVA MANERA DE HACER PREDICCIONES OPERATIVAS

José A. García-Moya

Servicio de Modelización Numérica del Tiempo. INM

RESUMEN

Los procedimientos operativos para realizar la predicción general en el INM han variado profundamente en los últimos veinte años. Sin embargo, la etapa final de la “cadena operativa de predicción” no ha variado tan radicalmente. El predictor continua redactando todos y cada uno de los productos de predicción manualmente. La necesidad de disponer de boletines de predicción general en varios idiomas llevó al INM a participar en un proyecto europeo llamado Multimeteo y cuya finalidad era precisamente la automatización de las predicciones en formato texto para su obtención en varios idiomas, español, francés, inglés, alemán y holandés. Posteriormente el INM amplió el número de idiomas para incluir el catalán, el gallego y el euskera. En esta presentación se analizarán las ventajas e inconvenientes que pueden derivarse del uso operativo de ambas herramientas y la manera en la pueden ayudar en las labores de predicción general del INM.

1. Introducción

Los procedimientos operativos para realizar la predicción general en el INM han variado profundamente en los últimos veinte años. A partir de la implantación del sistema McIDAS la cantidad de información manejada por los predictores ha aumentado notablemente. Actualmente se manejan grandes cantidades de datos en formato gráfico, tanto provenientes de los modelos numéricos de predicción como de las redes de satélites, radares y estaciones automáticas.

Sin embargo, la etapa final de la “cadena operativa de predicción” no ha variado tan radicalmente. El predictor continua redactando todos y cada uno de los productos de predicción manualmente. Como resulta que cada día aumenta el número de productos de predicción en formato texto que deben redactarse, tanto en el CNP como en los diferentes GPVs, el resultado es que el predictor invierte cada día más tiempo en los aspectos lingüísticos de la predicción y menos en los aspectos meteorológicos. Este proceso, que redundante en perjuicio del predictor y de la predicción, parece no tener fin.

Finalmente, hemos de tener en cuenta que el proceso de redacción en lenguaje claro de los diferentes productos de predicción no aporta ningún valor meteorológico añadido y, a veces, resta claridad a la exposición de la predicción meteorológica.

Todo esto y la necesidad de disponer de boletines de predicción general en varios idiomas llevaron al INM a participar en un proyecto europeo llamado Multimeteo y cuya finalidad era precisamente la automatización de las predicciones en formato texto para su obtención en varios idiomas.

La entrada para el programa Multimeteo es una matriz numérica de variables de predicción para la zona objeto que se obtienen directamente del modelo HIRLAM. Naturalmente un procedimiento operativo de predicción general no puede descansar en la automatización absoluta, de manera que se diseñó Modigraph que es una herramienta gráfica interactiva que permite al predictor modificar la matriz numérica antes de la entrada a Multimeteo.

En esta presentación se analizarán las ventajas e inconvenientes de ambas herramientas y la manera en la que su uso puede ayudar en las labores de predicción general del INM.

2. El proyecto Multimeteo y Multimeteo XXI

Desde hace unos diez años la generación de textos en lenguaje claro usando técnicas informáticas se ha desarrollado notablemente. El impulso inicial se debió al interés de las empresas de venta por correo de disponer

de herramientas automáticas para la confección de las respuestas comerciales a sus clientes. Naturalmente la condición es que el texto generado se parezca mucho al que generaría un redactor humano experto.

En el caso de la predicción meteorológica operativa, el aumento notable del número de productos a generar manualmente y la necesidad de que muchos de ellos estén disponibles en otros idiomas diferentes llevó a algunos servicios meteorológicos a plantearse en 1994 la conveniencia de un proyecto de investigación y desarrollo cuya finalidad fuera la creación de un software de generación de predicciones en lenguaje claro en diferentes idiomas. Se formó entonces un grupo constituido por los servicios meteorológicos de Francia, España, Bélgica y Austria y las empresas Lexiquet (Francia) y CL Servicios Lingüísticos (España). Se redactó un proyecto de investigación para tres años (1995-1998) que obtuvo financiación de la Unión Europea y al que se llamó Multimeteo. Posteriormente, se aprobó y financió también una ampliación del citado proyecto para el año 1999 (Multimeteo XXI).

Los idiomas elegidos para el proyecto fueron Inglés, Francés, Alemán y Español. En la tabla 1 se presenta el resumen del presupuesto de los proyectos y la parte financiada por la Unión Europea.

Proyecto	LQ	CL	MF	INM	IRM	ZAMG	Total	50% UE
Mmeteo	1 307	93	512	344	229	314	2 800	1 400
M XXI	482	68	178	160	130	130	1 148	574
Total	1 789	161	690	504	359	444	3 948	1 974
Pts Millones	297	27	114	84	60	74	655	328

Tabla 1: Presupuestos y financiación de los proyectos Multimeteo y Multimeteo XXI en k€

Como puede verse en la tabla la UE financió el 50% del presupuesto de cada participante. Después los servicios meteorológicos decidieron financiar con su parte el resto del presupuesto de las empresas, de manera que éstas quedaban financiadas al 100%.

Una vez terminados los dos proyectos el INM disponía de una herramienta informática para la generación de boletines de predicción en formato texto en inglés, francés, alemán y español. Se estudió entonces la posibilidad de ampliar el software a los idiomas cooficiales del Estado Español, es decir, catalán, gallego y euskera. El INM financió esta ampliación que se terminó en diciembre de 2000.

3. Validación de los resultados de Multimeteo

Desde el diseño mismo del proyecto Multimeteo se tuvo en cuenta la validación de los resultados obtenidos con el software como una de las piezas básicas del mismo. Puesto que la finalidad última del proyecto era la redacción de boletines de predicción en lenguaje claro tan parecidos como fuera posible a los que redactaría un predictor experimentado con los mismos datos de partida, la validación se dividió en dos partes netamente diferenciadas. La validación lingüística que consistía en la corrección gramatical y sintáctica de los boletines generados en cada uno de los idiomas, y la meteorológica que consistía en que el texto generado estuviera “meteorológicamente” de acuerdo con los datos de entrada de la aplicación.

3.1 Validación meteorológica

Cinco meteorólogos de los GPVs de Barcelona, Cantabria, Valladolid y Sevilla fueron elegidos para validar meteorológicamente los resultados de Multimeteo en dos periodos de cuatro semanas cada uno (diciembre de 1999 y marzo de 2000). Se validaron más de dos mil boletines provinciales, calificándose estos entre 1, muy malo y 5, muy bueno, siendo 3 aceptable. El promedio de validación de todos los boletines fue de **3.8**. En cuanto a las diferentes partes de las que se compone un boletín de predicción típico (nubosidad y meteoros, temperatura y viento) los resultados fueron **4.3, 4.4 y 4.3**, respectivamente. El hecho de que el promedio no coincida con la media aritmética de las partes se debe a que se calificaba cada parte en concreto y luego el boletín entero, de manera que un boletín con las tres partes calificadas como 4 se calificaba más bajo como conjunto.

De los resultados descritos anteriormente se deduce que **Multimeteo es perfectamente válido para generar los boletines provinciales de predicción del INM**. De las diferentes reuniones de validación surgieron no obstante consecuencias que es necesario resaltar. La primera es la dificultad en lograr un compromiso entre la longitud del texto de la predicción y la precisión de ésta. Otra dificultad consistió en lograr un acuerdo sobre algunas reglas de agregación (caso del uso del término “intervalos nubosos”).

Los boletines generados en los otros idiomas del proyecto fueron validados desde el punto de vista meteorológico por seis meteorólogos del INM expertos en los idiomas en cuestión. La finalidad era comprobar que había una “concordancia meteorológica” entre el boletín en español y el correspondiente a los otros idiomas. La consecuencia es que habían una concordancia muy buena entre ambos boletines.

3.2 Validación lingüística

Como ya hemos mencionado más arriba la validación literaria de los resultados de Multimeteo consistía no solo en asegurar la corrección gramatical y sintáctica de los boletines generados en cada uno de los idiomas sino que comprobar que se trataba de boletines perfectamente inteligibles por el ciudadano medio. Hay que tener en cuenta que los boletines se generan en cada uno de los idiomas a partir de un tronco de objetos lingüísticos, o sea, que no se trata de una traducción de un idioma a otro.

La validación literaria en español consistió en reunir a seis personas ajenas completamente al mundo de la meteorología para que leyeran y calificaran 100 boletines (50 generados por Multimeteo y los 50 generados operativamente por el INM en los mismos días y las mismas provincias). Se trataba de calificar su comprensibilidad desde el punto de vista del mensaje que transmitían (la predicción del tiempo). Otra vez se calificaban entre 1, muy malo y 5, muy bueno, siendo 3 aceptable. Naturalmente, los evaluadores no sabían de qué manera se habían generado los boletines. El promedio de los boletines generados por Multimeteo fue de **4.0** y el de los operativos de **3.6**. El comentario más común fue que los boletines generados por Multimeteo eran concretos y suficientemente claros.

La validación literaria de los boletines en lenguas extranjeras fue realizada por colegas de los servicios meteorológicos del proyecto (francés por MeteoFrance y alemán por ZAMG); en el caso del inglés se contrató a tres profesores nativos de inglés que efectuaron los trabajos. **La consecuencia general fue otra vez que Multimeteo generaba boletines perfectamente inteligibles por nativos del idioma en cuestión.**

3.3 Los idiomas cooficiales de España

Durante el año 2000 la empresa LexiQuest, por encargo del INM, acometió la ampliación de Multimeteo al catalán, gallego y euskera. La validación de estos nuevos idiomas se produjo durante el último trimestre de 2000 y el primero de 2001. En el caso del gallego y el catalán la validación se llevó a cabo por personal del INM de los CMTs de Galicia, Cataluña y Baleares. Para el caso del euskera, ante la inexistencia de personal del INM con un conocimiento suficiente del idioma, se recurrió a personal externo pero con conocimientos de meteorología. El resultado de la validación de todos los idiomas fue igualmente satisfactorio.

4. La interfaz gráfica Modigraph

Los datos de entrada que alimentan la aplicación Multimeteo se presentan en forma de una tabla de parámetros meteorológicos previstos cada 3 horas para la zona en la que se va a aplicar Multimeteo, por ejemplo una provincia de España. La matriz se extrae directamente de la salida de un modelo numérico (en el caso del INM Hirlam) y se pone a disposición del predictor para su modificación. Para que es proceso sea lo suficientemente ágil es necesaria una herramienta gráfica mediante la cual pueda procederse a la manipulación de la matriz resultante del Hirlam. Este es un punto clave para el uso operativo de Multimeteo ya que en un entorno operativo no puede perderse tiempo con manipulaciones que no sean de tipo gráfico.

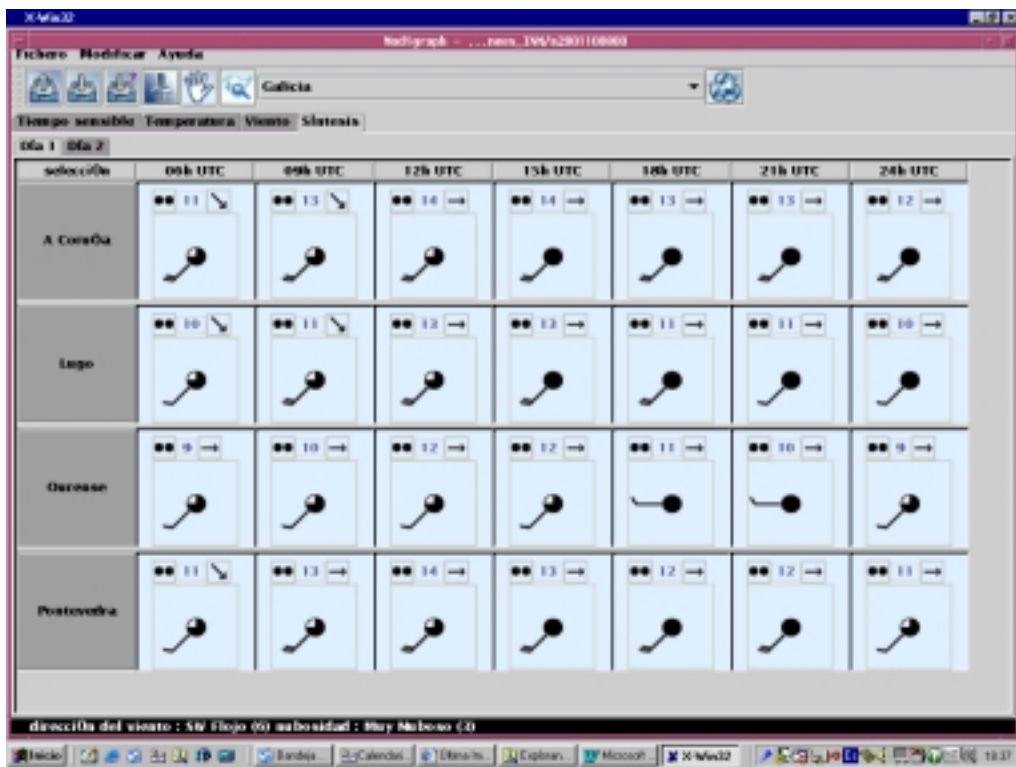


Figura 1: Pantalla de presentación de la aplicación Modigraph.

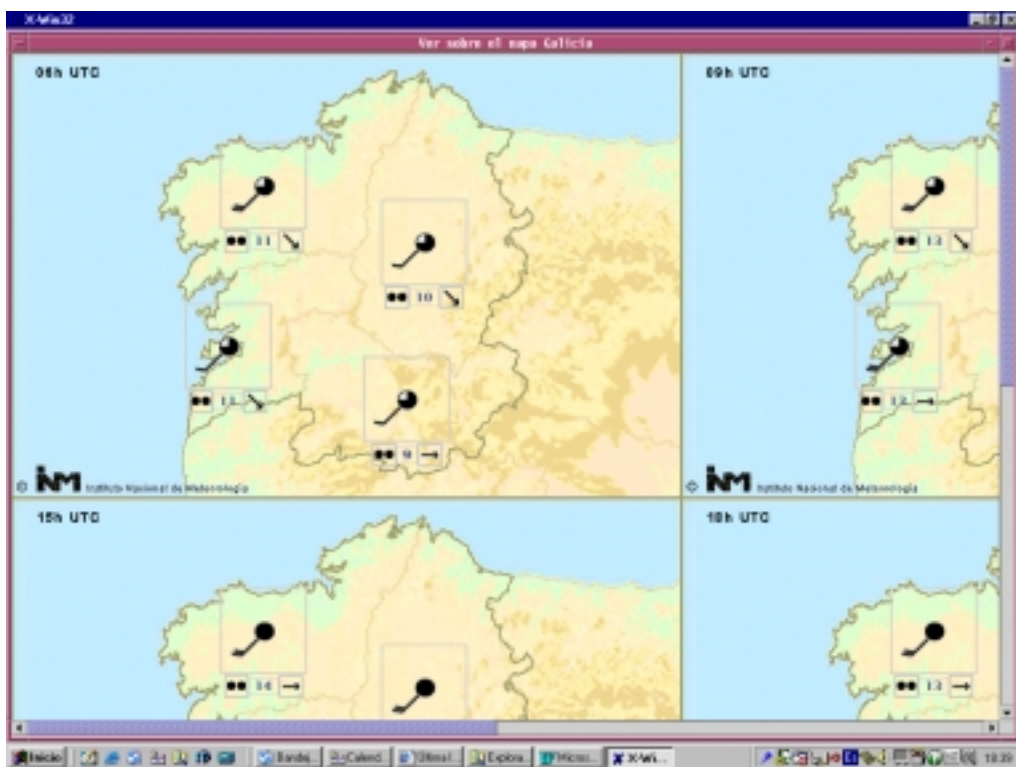


Figura 2: Presentación de la matriz en forma de mapa, para cada uno de los alcances.

Con este objetivo se diseñó y programó el software Modigraph. Se trata de una interfaz gráfica programada en java y c++ (puede ejecutarse en cualquier tipo de plataforma) que permite de manera simple la manipulación de la matriz de entrada de Multimeteo. En la Figura 1 se presenta la pantalla inicial de la interfaz.

En la Figura 1 puede verse el aspecto de la pantalla inicial de Modigraph una vez elegido el fichero de datos y la Comunidad Autónoma sobre la que se va a trabajar. Esta pantalla presenta un resumen de la predicción para cada provincia y para cada alcance en forma de tabla. La aplicación también presenta la predicción en forma de mapa de la Comunidad Autónoma tal como se ve en la Figura 2.

Se dispone además de tres pantallas, una para cada una de las partes del boletín de predicción (nubosidad y meteoros, temperatura y viento). Se eligen con las solapas superiores y en ellas se procede a la modificación de los parámetros que se desee. Esta modificación es siempre gráfica de manera que el predictor no necesita tener conocimientos especiales de informática para realizarla.

En todo momento se presenta en la pantalla de la aplicación toda la información necesaria para que el predictor no necesita consultar ningún tipo de documentación mientras trabaja con Modigraph.

5. Conclusiones

El INM dispone ya de una aplicación capaz de generar, con la calidad suficiente, boletines de predicción en lenguaje claro y en ocho idiomas diferentes. El objetivo más importante del proyecto Multimeteo, que era la generación de los citados boletines con una calidad similar a la de los generados por predictores humanos, se ha cumplido suficientemente. Con esta herramienta se podrán generar todos los productos diferentes que se desee sin por ello perturbar el tiempo que el predictor puede dedicar a labores “meteorológicas”.

El problema derivado de la manipulación de la matriz de entrada de datos, de manera que dicha matriz represente convenientemente “la predicción deseada por el predictor” se ha resuelto con una aplicación gráfica (Modigraph) que permite dicha manipulación a base de usar el ratón del PC o de la estación de trabajo.

Puesto que ambas aplicaciones se han programado en Java y c++ son independientes de la plataforma sobre la que se integren (tanto estaciones de trabajo UNIX como PC con Windows).

La elasticidad de las aplicaciones y la bondad de los resultados hacen de estas aplicaciones buenos candidatos a resolver los problemas de tiempo en las oficinas de predicción operativa del INM.

Finalmente, hay que tener en cuenta que estas herramientas buscan un aumento del tiempo disponible para labores meteorológicas en las oficinas de predicción, de manera que cada predictor solamente tenga que realizar una única predicción del tiempo a partir de la cual puedan extraerse automáticamente todos los productos que los usuarios demanden del INM.

Agradecimientos

Es evidente que un proyecto tan complejo como éste nunca hubiera podido abordarse sin la colaboración de muchas personas dentro y fuera del INM. El grupo de trabajo formado en el INM ha contado con la participación más o menos directa de más de treinta personas. Como sería muy largo de enumerar a todas, quede aquí mi agradecimiento, como director del proyecto, a todas ellas. También hay que agradecer la participación desinteresada de algunas personas de fuera del INM, en particular Onitze Salazar de TV5 y Vicente Iribar Sorazu de la Universidad del País Vasco.

Referencias

EC (1995): “Contract LE1-1612. Project Multimeteo”. European Commission, DG XIII. Ten-Telecom Program. 150 pp.

EC (1997): "Adendum of the Contract LE1-1612. Project Multimeteo". European Commission, DG XIII. Ten-Telecom Program. 90 pp.

EC (1998): "Project Multimeteo XXI". European Commission, DG XIII. Ten-Telecom Program. 70 pp.

Coch, J., (1998): "Interactive generation and knowledge administration in MultiMeteo" in Proceedings of the Ninth International Workshop on Natural Language Generation. Niagara-on-the-lake, Ontario, Canada, 1998.

Coch J., De Dycker E., Garcia-Moya J.-A., Gmoser H., Stranart J.-F., Tardieu J., (1999): "MultiMeteo: adaptable software for interactive production of multilingual weather forecasts" in Proceedings of the 4th European Conference on Applications of Meteorology (ECAM 99) , Norrköping, Suède 1999.

Chevreau, K., Coch, J., (1999): "Génération multilingue de bulletins météorologiques : le logiciel MultiMeteo" in Proceedings of the GAT workshop, Grenoble, France 1999.

Chevreau, K., Coch, J., García-Moya, J.A., Alonso, M., (1999): "Generación multilingüe de boletines meteorológicos" in Proceedings of the SEPLN, Lleida, Espagne 1999.

Coch, J., Chevreau, K., (2001): "Interactive Multilingual Generation" in Alexander Gelbukh (Ed.): Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science Vol. 2004. Springer Verlag, Berlin, Allemagne 2001, pp.239-250.
<<http://link.springer.de/link/service/series/0558/papers/2004/20040239.pdf>>