A Preliminary Experimentation for Meteorological Data Visual Analytics

Pisano Francesca Maria, Cembalo Assuntina, Romano Gianpaolo
Italian Aerospace Research Centre,
via Maiorise, 81043 Capua (CE)
ITALY

f.pisano@cira.it, a.cembalo@cira.it, g.romano@cira.it

ABSTRACT

Everyday sensors and detectors register huge quantities of geo-referenced meteorological data up to 40-45 GB a day, available in different formats (.csv, .nc, .grb, .xlxs, .tiff, .shp, etc.), at different frequencies; that data is to be appropriately analyzed to extract information as support in critical domains such as aviation. Meteorologists have expressed analysis requirements referring to visualization and graphical representations of data as univariate or multivariate spatial distributions of physical observations and simulated variables. Meteorologists can potentially gain advantage in adopting the innovative approach of Visual Analytics based on interactive and visual data representations to identify spatio-temporal trends and correlations. In the present work, a preliminary Visual Analytics solution, combining Tableau and in-memory database Exasol, is designed and implemented to offer worksheets e dashboards, whose visualization requires few seconds versus hours and days needed to generate traditional meteorological maps and graphs. Satisfactory experimental results are shown for COSMO (Consortium for Small-scale Modeling) data and ECMWF (European Center Medium Weather Forecast) SYNOP (Surface synoptic observations) data for a total volume of almost 1 GB and more than 21 millions of records.

1.0 INTRODUCTION

Weather information is nowadays extremely relevant for safety in a number of domains, because it can effectively help to avoid incidents and mitigate consequences related to extreme weather events. Both infrastructures and transportation systems can benefit from weather awareness in order to manage adverse weather conditions. The European Union Aviation Safety Agency has a Weather Information to Pilots project team [1] that "has reviewed the existing means to mitigate against the effects of the weather, including the use of on-board weather radar, other on-board sensors, the information provided to pilots pre-flight, and the information that is available in-flight". In the U.S.A., the Aviation Weather Center delivers consistent, timely and accurate weather information for the world airspace system [2].

Meteorology is a complex scientific area, making use of both observation and prediction data, covering large geographical areas, with updating period from 10 seconds to 1 day. The visual approach is usually adopted by meteorologists, who are used to work with graphs and maps to investigate and communicate complex phenomena in a synthetic and integrated manner [3], nevertheless they experience the limit of depicting static information.

Numerical Weather Prediction systems can take advantage from the remarkable progressive growth registered in recent years by the hardware computational power [4], that allows the storage of huge amount of data in various domains. Unfortunately, existing analytical systems can meet difficulties in managing such increasing quantity of data, so innovative solutions become necessary to adequately exploit all the available information coming from different sources, in different formats, at different frequencies: the so called *Big Data* [5]. Visual Analytics (VA) can effectively helps as it is the *Science of analytical reasoning facilitated by interactive visual interfaces* [6]. VA aims at stimulating the human cognitive abilities, in identifying patterns and detecting relationships in visualizations, in combination with the data management and analysis methods and technologies. Meteorology can just gain great benefits from the Big Data Visual Analytics [7], providing the breakthrough technical solutions to be applied in order to extract relevant information for experts and end-users from new datasets.

In the present work, an experimental approach to manage huge quantity of meteorological data is proposed using unconventional storage solution like a NewSQL database [8] and a Visual Analytics tool in order to satisfy meteorologist requirements with respect to the interactivity and the time performance of analytics.

2.0 RATIONALE

Traditional Information Visualization objectives [9] are about the effective representation of analytical activity results or the confirmation of domain expert hypothesis. Towards the end of 90s, a group of visionary researchers at the Pacific Northwest National Laboratory [10] detected the gap between the existing Information Visualization capabilities and the emergent demanding analytics requirements. After 09/11, growing national security and safety needs gave birth to projects dedicated to the elaboration of innovative visual approach to analytics worldwide ([11], [12], [13], [14]). Visual Analytics pushes forward the application of visualization techniques towards the creation of analytical models supporting information exploration and extraction by means of graphs and maps. Visual Analytics is a multidisciplinary research area ([15], [16], [17]) integrating Data Management and Analysis, Statistics, Spatio-temporal Data Processing, Visualization and Human-machine Interaction, which aims at passing over the simple representation of available data, facilitating meaningful interactions so that the end-user experiences innovative information understanding tools. In this sense, the right choice of the graphs and maps and of the interactive solutions has to be strictly related to user needs. Visual Analytics is able to effectively manage heterogeneous data formats by means of different representations; spatial and time data are especially suitable to visual depiction. VA mainly pursues the tasks of Exploration and Analysis: the former includes data selection, pattern and trend identification, outlier detection; the latter faces issues as visual support to analytical algorithms and models evaluation and the effective results representation. Therefore, VA can expand users' explorative analytical experience and possible creative processes. A particularly relevant aspect of Visual Analytics is about the communication of the obtained analytical results; methods and tools supporting the data storytelling [18] are necessary to improve domain expert abilities to disseminate the outcomes of their work.

Visual Analytics is applicable to issues featured by huge volume of information, high variety, heterogeneous sources and large number of parameters; therefore it seems to be desirable for domains as Climatology and Meteorology, which are featured by high quantity of data, daily collected, to be compared over long time periods. In recent years, some applications of Visual Analytics to the meteorological domain have been elaborated to support situation and weather awareness. The visualization solution named "VIDa" (Visual Interactive Dashboard) is proposed in [19]; multiple views of time sequences, spatial filters and maps support end-users seeking spatio-temporal patterns for short-term forecasts. The MEVA (Multifaceted Environmental Data Visualization Application) [20] offers interactive visual interfaces for meteorological observation and simulated data to users without advanced skills in 3-D visualization systems. The Ship and Weather Information Monitoring decision support system [21] aims at specifically applying the visual approach to weather awareness for naval transportation system. Naval fleet routes are elaborated taking into account timely meteorological conditions; in order to do that, data from ECMWF and updated route information are integrated and alternative routes can be depicted by means of interactive visual interfaces.

3.0 DESCRIPTION

In the present work, an experimental approach to manage huge quantity of meteorological data is proposed using unconventional storage solutions like in-memory analytic database [8] and a Visual Analytics tool in order to satisfy meteorologist requirements with respect to the interactivity and the time performance of analytics. A brief introduction to specialist requirements is provided together with the description of the data sources selected for the experimentation. Then the solution architectural design and the adequate tools selection on the basis of expert needs are discussed.

3.1 Domain requirements

In collaboration with the Italian Aerospace Research Centre Meteorology Laboratory, expert requirements have been gathered; they are firstly about the exploration of the spatial distribution of one or more physical and numerical variables. Typically, meteorologists work with graphs and maps to depict available information in a synthetic and integrated approach; the power of visual representation is limited by the static nature of images. Indeed, specialists recursively elaborate graphs to gain exhaustive insight into data. Experts ask for dynamically and interactively selecting the geographical region and the variables values of interest, and they request to download the obtained graphs and maps as the most widespread image formats. Moreover meteorologists need plots representing observations and simulated data as scatterplots and temporal series to seek for correlations and trends; often mean values and other statistical indices of collected registrations are to be shown. Specialists also require faster analytical tools enabling the elaboration of graphs and maps dynamically, following the expert own needs during the user exploration and analysis activities. All the operations such as processing, analysis and visualization are strongly time-consuming when recurring to traditional analytics methods and technologies.

3.2 Meteorological data

The activity of requirements analysis is joined to the analysis of data sources, which have been selected by meteorologists as a representative specimen of relevant information to process. Different kinds of data have been provided, including observations and numerical datasets of different variables.

A brief description follows:

- **SYNOP data**: observed data stored in the MARS (Meteorological Archival and Retrieval System) [22] repository managed by the ECMWF (European Center Medium Weather Forecast) [23]. Registrations are collected hourly and include one alphanumeric code to respresent date, geographical area, significant phenomena occurrence, etc. and 29 fields (station coordinates, temperatures, wind speed, pressures, etc.). Records are gathered by 877 monitoring stations, positioned in Europe and Northern Africa. The dataset includes 4 months of registration in 2016, for a total amount of more than 2 millions of records and 250 MB volume.
- **COSMO data**: simulated data from the Consortium for Small-scale Modeling model [24]. Data represents temperature and rainfall registered by 51520 monitoring stations in Central Europe during the year 2016. It amounts to about 19 millions of records and 700 MB.

3.3 Analysis and design

Architectural solution has been derived from the domain expert needs and data sources analysis. Such an integrated design approach drives to a solution able to take into account expert and end-users requirements combined with the effective available information. In the following, the technical issues to be faced and the proposed solution in terms of architectural components and related tools are described.

3.3.1 Architectural design

Analytics requirements from meteorologists are focused on demanding data visualization and representation needs with respect to spatial and statistical distributions of one or more variables in order to seek temporal trends and possible correlations. As mentioned before, the possibility of storing huge quantity of data, both for physical observations and numerical forecasts, covering long time periods, represents an outstanding opportunity to elaborate more accurate forecasting models for meteorologists. The complexity of analytics activities is often faced using graphs and maps to catch synthetic and integrated view of data, eventually for subsequent more detailed analysis.

In this context, Visual Analytics offers unheard-of interactive visualization solutions, enabling a rich and customizable experience of data exploration, that can effectively empower expert activities. Indeed, *Visual exploration* can support the information interpretation process in a user-friendly manner, using geographical maps, which easily represent multivariate problems in two and three-dimensions. Points stand for single observations corresponding to positions in a specified reference system; rectangles represent observation groupings whose extension is related to the number of included points; lines link observations and polygons can aggregate rectangles. All those geometric elements can be combined to create graphs, that can in turn form analytical interactive dashboards. The selection of specific elements and graphs has to be driven by the expert needs. Geographical maps represent positions of interest, related to monitoring stations or sensor coordinates, and use shapes, dimensions and colors to depict one or more variables values. In addition, operators as sum, average, maximum, minimum can be displayed and more textual information is provided passing the cursor over the map. Filters, highlights and dynamic animations facilitate the data exploration experience. Map portion selection is performable to focus on a reduced geographical region of interest, synchronizing statistical indices and graphs.

The temporal trend of a variable can be represented by means of linear graphs that can be aggregating with respect to the time dimension (year, month, day, hour, etc.) updating the aggregation operators (sum, average, etc.). More variables are added using side-by-side graphs, and/or using colors and shapes. Time is depicted using discrete or continuous approach. Box-plots to compare variable distributions for different time spans and scatterplots compare two variables for correlation identification; they can be enriched with information using colors and shapes again. The user takes advantage of the graph synchronization, so that selections are applied to all available visual representations on the screen.

On the other hand, the time performance challenging requirement fulfillment can be sought applying innovative solutions for the management of huge volume of data. Actually, COSMO simulated data amounts to about 700MB, not adequately managed by Visual Analytics solutions. It results in exceeding processing time that dramatically affects the user interactivity. In order to overcome such a drawback, *in-memory database* technology [8] is investigated to achieve faster data management, exploiting the computer central memory. In-memory databases belong to NewSQL database category [25], which assure as high performances as the NoSQL databases, keeping the well-known relational approach and SQL query language. Such databases effectively support the Volume, Velocity, and real-time Big Data analytical needs [26].

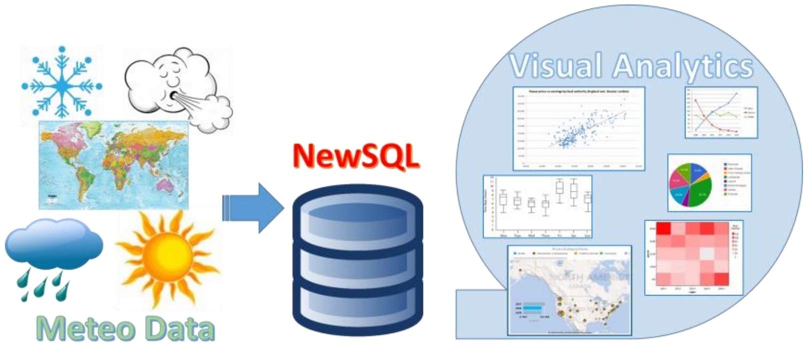


Figure 3-1: Architectural solution

One last issue is worthy to be faced: the solution release. As a matter of fact, it can be helpful for the end-users to have easily shareable analytical results at disposal, but this could be still an issue when high volumes of data are to be shared. A shareable solution including analytical results and data of interest should be investigated.

3.3.2 Tools

For the experimentation of the above described architectural solution (Figure 3-1), a technological benchmarking activity has been carried out. It has resulted into the selection of the free in-memory database Exasol [27] and the commercial Visual Analytics tool Tableau [28], which has native connection to Exasol database [29].

Exasol is a high-performance in-memory database, able to manage data much faster than traditional relational databases making use of mass memory. Exasol requires the installation of a Java Virtual Machine: Oracle Virtual Box has been exploited. Exasol can be used both from command line or by means of the SQL editor console *ExaPlus*.

Tableau Desktop is one of the most widespread Visual Analitycs solutions worldwide. It is ranked as the second “Analytics and Business Intelligence Platform” for completeness of vision and ability to execute in the corresponding Gartner Magic Quadrant [30]. Tableau offers many graphs to be combined in effective worksheets including filters, highlights, legends and parameters and advanced features to create dashboards from them. Tableau lets save Visual Analytics applications as collection of worksheets and dashboards, optionally joining the associated data for release to users who do not have access to source data. Moreover, Tableau capabilities include *Self-service Data Preparation*, so user can create derived fields, dimensions and measures by themselves.

In the end, *Tableau Reader* [31] is a free software tool allowing the access and utilization of the visual analytical applications produced using Tableau Desktop also to users who do not own Tableau license.

4.0 RESULTS

COSMO data has been cleaned eliminating special characters and about 19 million of rainfall records have been stored in the Exasol database, using the ExaPlus editor, taking less than 40 seconds. The Tableau connection to Exasol allows to accelerate data visualization from minutes to few seconds. The most relevant advantage is about the interactive visual approach to data exploration and analysis; end-users can navigate worksheets and dashboards investigating temporal trends and comparing variables over geographical maps, dynamically selecting statistical indices and filtering values. In this way, experts can exploit the interpretation power of the visual approach and save hours of waiting for traditional graphs and maps elaboration and finally speed up the analysis work. Both for the COSMO and SYNOP data, interactive geographical maps have been created to represent the spatial distribution of meteorological variables statistical indices, aggregated on the selected time interval, using shapes like circles and squares to depict the value: bigger is the circle, higher is the index value. Filters are applicable to select a specific week or month of observation.

In Figure 4-1, the worksheet related to the animation of the COSMO monthly-based average rainfalls is represented. The legend on the upper right corner lets user select a specific month or to play the visual animation of the monthly average rainfalls sequence. The color legend indicates the depicted average values. The map appears as a solid shape because of the dense grid of monitoring stations; zooming in the background image, single points corresponding to monitoring station become visible.

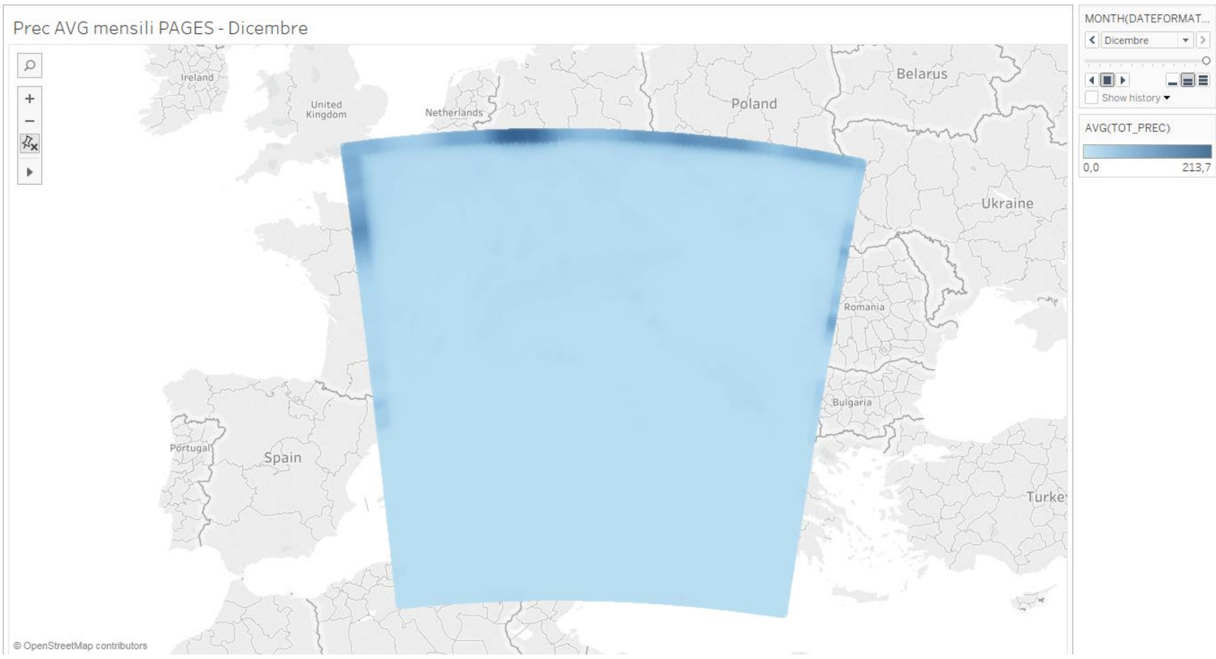


Figure 4-1: Animated sequence of simulated rainfall average value spatial distribution

In Figure 4-2 a SYNOP variable spatial distribution is shown; the circle area is proportional to the minimum value in a specific week/month of observation that user can select using the filter on the right.

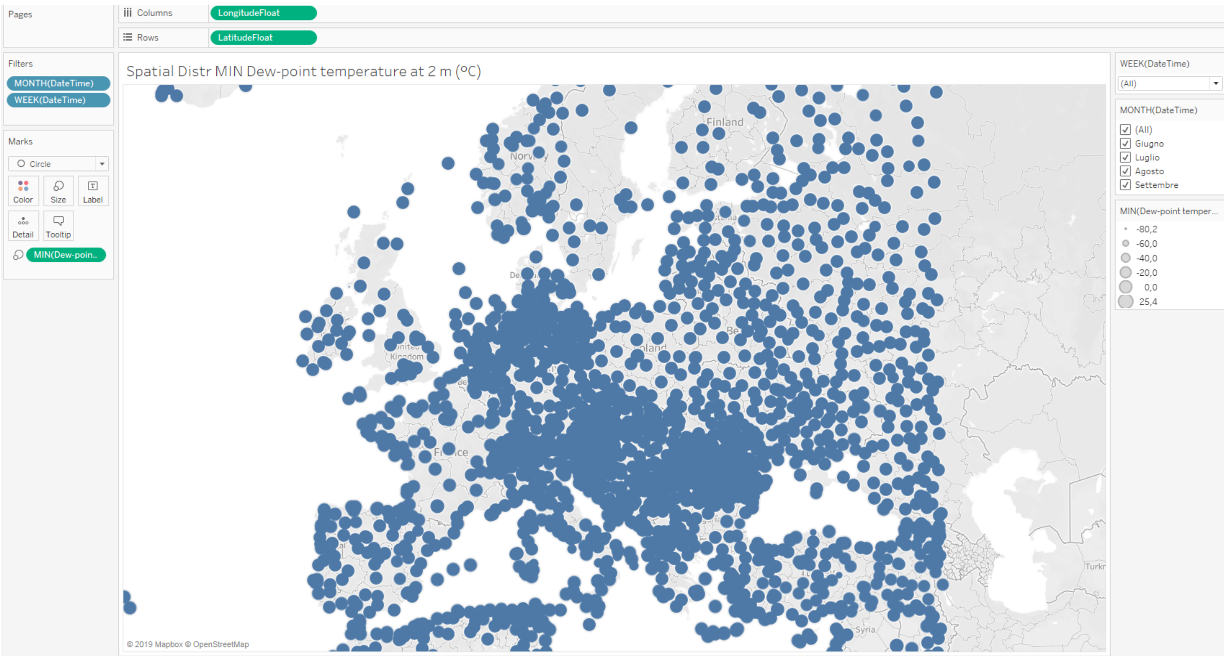


Figure 4-2: One variable spatial distribution with time filter

On the map in Figure 4-3, a second physical variable has been added to let user catch information about two observed dimensions at one glance. The first variable values are represented by the circle areas, while the second variable values are depicted by colors as indicated by the palette color legend. Interactivity is empowered by filters over the latitude and longitude values, which allow the user to select geographical area of interest, and by animated sequence of monthly average value spatial distributions of the two variables.

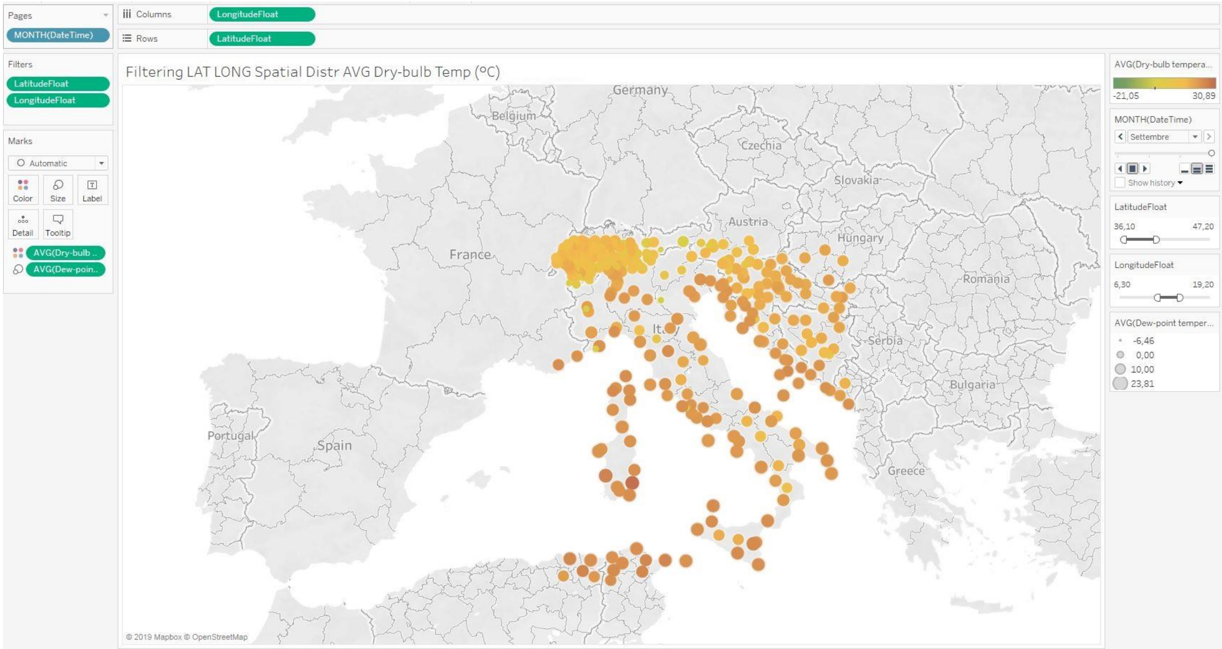


Figure 4-3: Animated sequence of two variable spatial distributions with spatial and time filters

Analytical dashboard can be built by suitably combining worksheets. Starting from a map (Figure 4-4), user can select a region of interest using the cursor and such a selection works as a filter to a second dashboard representing analytical graphs (Figure 4-5). In this way, user is supported exploring an increasing level of detail, because a visual selection is much simpler and more intuitive than numeric filters application.

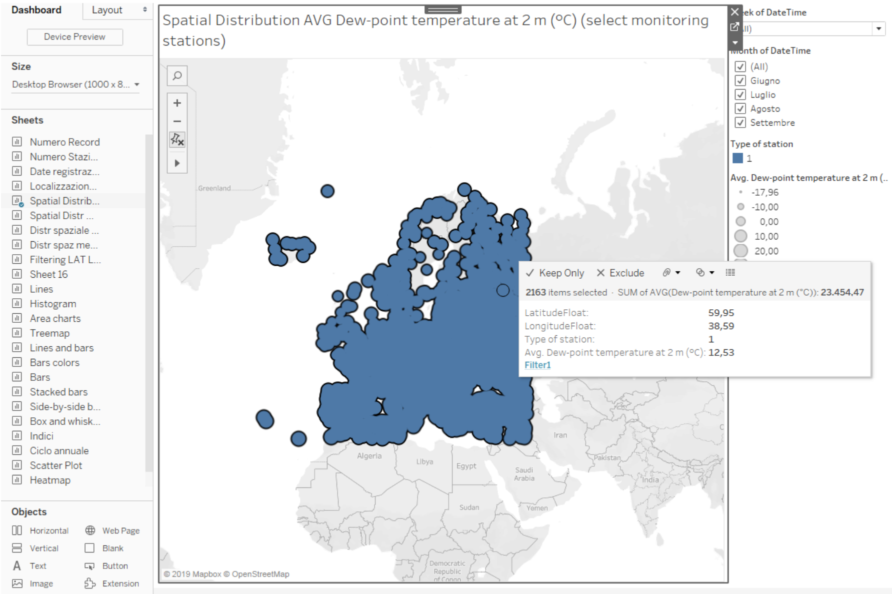


Figure 4-4: - Spatial filtering

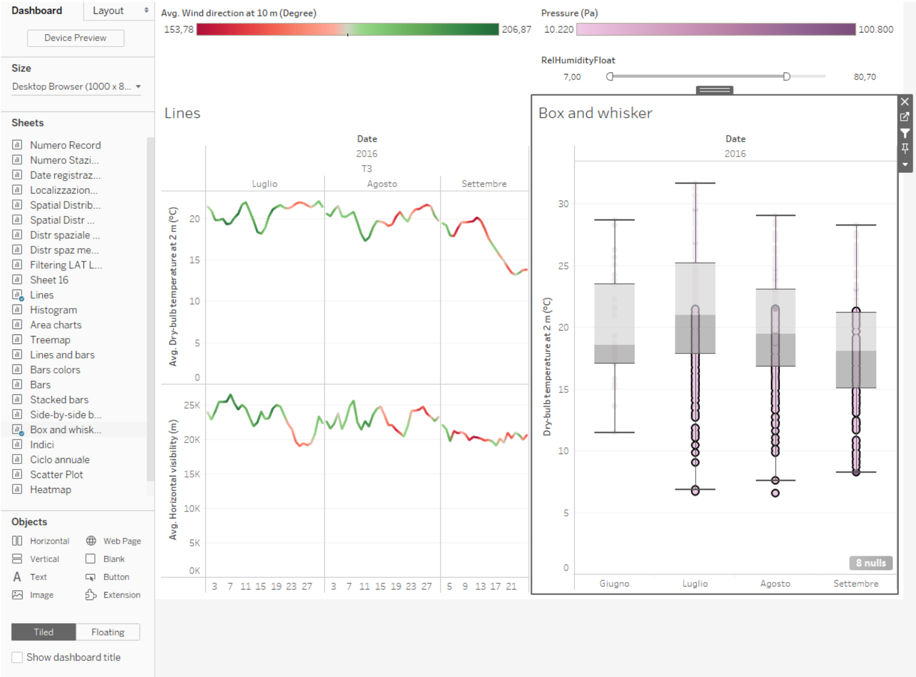
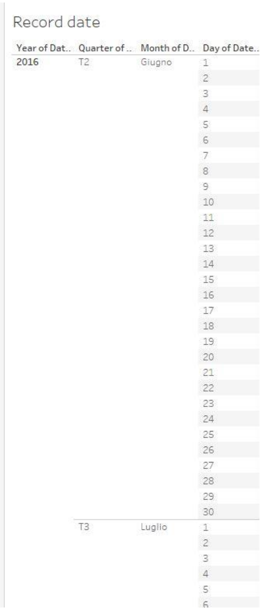


Figure 4-5 – Filtered linear and box-plot graphs

Besides some easy worksheets are provided to support the initial exploration of data; they let user navigate the registration time spans and the spatial distribution of the monitoring stations.



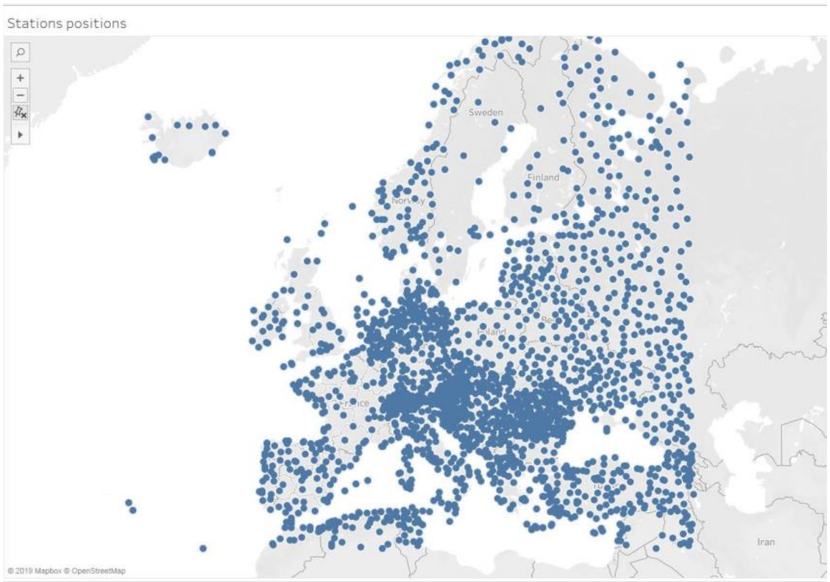


Figure 4-6: Records time span

Figure 4-7: Spatial distribution of monitoring stations

5.0 CONCLUSIONS

In the present work, elements for a Visual Analytics solution design have been described aiming at assuring ease of use for the end-user so that the information understanding and perception process from huge amount of heterogeneous data can be significantly empowered by interactive analytical interfaces. Meteorologists requirements about interactivity and time performance have been met using graphs and maps and in turn combining them into worksheets and dashboards providing effective representations of available information and facilities for eventually subsequent more detailed analysis. Using technologies based on the data storage in the central memory instead of traditional mass memory, faster elaboration and visualization of high amount of data are achieved. Hopefully, future work will be focused on the integration of advanced analytical models in the Visual Analytics application, moving forward towards the empowerment of analytics.

REFERENCES

[1] https://www.easa.europa.eu/easa-and-you/air-operations/weather-information-pilots#group-easa-related-content

[2] https://www.aviationweather.gov/

[3] Manzi P., Zollo A., Mercogliano P., Galdi C., "Aviation weather awareness: Development of algorithms for the detection of weather hazard through the use of EUMETSAT products", IEEE Metrology for Aerospace (MetroAeroSpace) Conference, 4-5 June 2015, Benevento, IT (2005)

[4] Mylne K., Big Data challenges in NWP and its Application, EMS Annual Meeting Abstracts Vol. 12, EMS2015-172 (2015)

[5] Big Data - A New World of Opportunities, NESSI (The Networked European Software and Services Initiative) White Paper, (2012)

[6] Thomas, J., Cook, K., *Illuminating the Path: Research and Development Agenda for Visual Analytics*, IEEE-Press (2005)

[7] Gaztelumendi S et al., "Big data in the context of Euskalmet activities", EMS Annual Meeting Abstracts Vol. 12, EMS2015-245, (2015)

[8] Stimmel C.L., *Big Data Analytics Strategies for the Smart Grid*, CRC Press, (2016)

[9] Spence R., *Information Visualization: an introduction*, 3rd Ed., Springer, (2014)

[10] https://www.pnnl.gov/computing/visual_analytics/index.stm

[11] http://www.vismaster.eu/

[12] Keim D., Kohlhammer J., Ellis G. and Mansmann F. (Editors), *Mastering the Information Age Solving Problems with Visual Analytics*, VisMaster - European Coordination Action Project, Eurographics Association (2010)

[13] Dill J., Earnshaw R., Kasik D., Vince J., Wong P.C., *Expanding the frontiers of Visual Analytics and Visualization*, Springer, (2012)

[14] http://www.va-sa.net

[15] Wong P. C., Thomas J., "Visual Analytics, Computer Graphics and Applications", IEEE (Volume:24 , Issue: 5), (2004)

[16] Keim D., Mansmann F., Schneidewind, Thomas J., Ziegler H., "Visual Analytics: Scope and Challenges", Lecture Notes in Computer Science, No.4404, pp. 76-90 (2008)

[17] Keim D., Mansmann F., Stoffel A., Ziegler H., "Visual Analytics", Encyclopedia of Database Systems, Springer, (2009)

[18] Segel E., Heer J., "Narrative Visualization: Telling Stories with Data", IEEE Transactions on Visualization and Computer Graphics, Vol. 16 , Issue: 6 (2010)

[19] Diehl A., Pelorosso L., Delrieux C., Saulo C., Ruizk J., Gröller L., Bruckner S., "Visual Analysis of Spatio-Temporal Data: Applications in Weather Forecasting", In Eurographics Conference on Visualization (EuroVis) – Carr H. , Ma K.L., Santucci G. editors, Volume 34, Number 3, (2015)

[20] Helbig C., Bilke L., Bauer H.S., Böttinger M., Kolditz O., "MEVA - An Interactive Visualization Application for Validation of Multifaceted Meteorological Data with Multiple 3D Devices", http://dx.doi.org/10.1371/journal.pone.0123811, (2016)

[21] Lundblad P.; Eurenius O.; Heldring T., "Interactive Visualization of Weather and Ship Data", 13th International Conference Information Visualisation, 379 – 386, (2009)

[22] https://www.ecmwf.int/en/forecasts/datasets/archive-datasets

[23] https://www.ecmwf.int/

[24] http://www.cosmo-model.org/

[25] Moniruzzaman A. B. M., Hossain S. A., "NoSQL Database: New Era of Databases for Big data Analytics – Classification, Characteristics and Comparison", International Journal of Database Theory and Application, Vol.6, No.4, (2013)

[26] Dillon S., Brodeur J., "Supporting the Smart Grid 2.0 using In-Memory Database System", (2015) https://www.slideshare.net/StephenDillon/supporting-the-smart-grid-with-imdbs-48353189

[27] https://www.exasol.com/en/

[28] https://www.tableau.com/it-it

[29] https://help.tableau.com/current/pro/desktop/en-us/examples_exasolution.htm

[30] https://www.gartner.com/en/documents/3900992/magic-quadrant-for-analytics-and-business-intelligence-p

[31] https://www.tableau.com/it-it/products/reader