QUADERNI DI METEOROLOGIA APERTA **n° 2**

REANALYSIS, OR RETROSPECTIVE ANALYSIS, IN METEOROLOGY

Massimo Crespi[1]
Gianluca Ferrari[2]

In Padova (Italy), autumn 2020

1. INTRODUCTION

1.1. Goals

Meteorology and climatology are gaining importance as quantitative variables in the design of forthcoming and future configurations of the atmosphere, and in its interactions with man, society and the planet. This process is encouraged by innovation and digital technology, which have triggered a growing demand for representative and consistent meteorological records suitable for consistent statistical processing. This need is met also by applying reanalysis or retrospective analysis techniques, which provide historical and near real-time datasets arranged in regular territorial grids of varying size, obtained from *in situ* meteorological observations.

At market level, we often come across proposals that have little to do with a methodologically correct application of retrospective analysis. In Italy, if we rule out most of the world of risk management, this is due to the inadequate specific preparation of users, caused by a lack of criticism and supervision by the academic world and the institutions. This publication is not a scientific text, but a professional guide to the methods of acquisition and use of data obtained through reanalysis, and its intention is to address the issue of reanalysis in pragmatic and operational terms.

Other types of dataset, which have less statistical consistency but are still usable in defined and circumscribed contexts, are also considered, with the aim of encouraging greater discernment by the user. These datasets are grouped under the heading of "pseudo-reanalysis" and should not be confused with solid retrospective construction, with which they compete easily with regard to price but not quality.

As the representativeness of meteorological data is always linked to the use made of said data, it should be pointed out that this document does not refer to meteorological applications that have an immediate operational purpose, the value of which is purely informative. It looks rather at the use of meteorological information as part of a more complex process capable of representing a given state of the atmosphere at a given time and therefore providing formal, and obviously conventional, content in contractual, insurance, legal and other evidence-based contexts[3].

In order to better define this framework, it should be noted that the text is based on countless in-house experiences in supplying reanalytical services to risk management companies. The part covering regulations refers to the indications of the WMO[4] and the computations used have been produced by European and international research centres.

Among the professional uses of climate reanalysis, the most prominent role is expected to be occupied, in time, by the knowledge of the parameters related to global warming. Consequently, it will be destined to a wide and transversal audience, before whom it must qualify as a credible and shared reference platform. This raises a number of fundamental considerations. One of these is the urgent need to initiate public awareness of meteorological and climatological issues, addressing all those who can or must help mitigate global warming. Efforts must be made to enrich the content and multiply the options available, contributing to broader inclusion, the sharing of choices and, consequently, increased success. These processes are not defined by law. The worlds of culture and communication play a key role in them, being called upon to broaden horizons and quickly overcome the narrow confines of meteorology, which has the technical role of monitoring and certifying, but its limited representativeness and institutional weaknesses are unable to sustain the cogency of decisions that are incisive and sometimes radical.

Reanalysis, which tells us about the past and forecasts the future, aims to fulfil this role using plain language, not as some kind of vague oracle, shrouded in a technological haze, but seeking common roots and vocabulary, to encourage a reading open to transversal paths and other sensibilities.

Once again, meteorology supplies technology, knowledge and sensitivity to offer a unified and global view of the world, in keeping with its tradition since the birth of the IMO[5] in 1879, and its subsequent co-option into the WMO[4], in 1951, as the intergovernmental technical agency of the UN[6].

1.2. Concise cultural contextualisation of the vectorial and discretized reading of the physical world in the contemporary era

Meteorological reanalysis is the tool used to quantify our planet's meteorological and climatic dynamics, both overall, as planetary-scale phenomena, and on a more specific and local level. It gives us an organic and unitary view of meteorological parameters, distributing them in a uniform grid, suitable for statistical applications and sophisticated processing. This method is based on a historicist analysis of the past, as the earth’s atmosphere, like the earth itself, evolves over relatively long periods of time, offering the attentive observer a view of its trends. Assuming a panoptic continuity between past and future, from a long-term or infinite perspective, the temporal parameter ceases to exist[7]. This paradox, which entraps even Zeno's tortoise[8], enables the development of historical series and datasets crystallised in regular grids, backed by a correct, robust statistical approach that meets our needs.

The future nests in the past and a retrotopic effort[9] is to make it easier to read, decipher and compare; in this sense, reanalysis is a shared, common, objective field of verification and comparison. At times when temperatures are particularly high, it is important to be aware that the knowledge, accessibility and transparency of meteorological data, along with its accuracy, should be among the standards for measuring the characteristics of an open society that is making slow and gradual human progress[10].

Climate is no longer a mere abiotic component of the earth's ecosystem, a factor with which humanity has shared thousands of years and to which it has adapted, learning to cope with even its strongest manifestations, albeit with varying periods of resilience. Tolerances, thresholds and return periods could be exceeded in the years to come, and territorial, environmental, economic and geopolitical balances could be shattered. So the problem is not technical; on the contrary, it must move outside the borders of digital technological hyper-realism and open up to an epistemological paradigm that guarantees understanding, sharing and awareness by individuals and society as a whole.

The product of reanalysis represents the matrix for objectively plotting and comparing all the different sensibilities and policies. After all, meteorology and climatology are not detached from the historical, cultural, social and economic context; they condition it and are, in turn, influenced by it. Growing general attention is prompting a multiple reading, which could highlight aspects, contents and perspectives that have so far escaped mere technical analysis and might have a positive and essential impact in terms of solutions and their effectiveness. Goals of this kind require a strong thematic contextualisation and a historical association that penetrates and highlights this sensitivity with other levels of knowledge, in order to allow its cultural appropriation and widespread identification. Being the fruit of its own time, there is no lack of convergence and affinity with the reanalytical approach, a small sample of which is proposed here in relation to more recent years.

A fitting definition for this search for correlation, already ripe for the ensuing digital age, was coined in the 1970s by Henry Leborit[11] who expressed the need to: "... *imagine new conceptual grids, new structures that capture the essential contribution of biological disciplines as a whole, not separately, but in an integrated form, from physics to the human species in the biosphere, in the time of evolution and that of the individual, in the gratifying space of a man and of all men, the planet*". It tells us many things: firstly, it presents a holistic vision of the grid as a concept; it corresponds to the need for a discretized reading of the world, capable of portraying large-scale syntheses and the finer details revealed through small-scale presentation; it converges in man and what we know about him.

This is a structuralist reading, used, ahead of its time, by Claude Lévi-Strauss[12] , in applying "savage thought" to the great eschatologies, and discovering a humanity or, better, a group of individuals, in constant search of their place in the universe, and hence a better definition of it. This reading is widely confirmed, discussed and developed by Zygmunt Bauman[13] in his portrayal of the individual lost in the sea of liquidity and of a "modernity" that consists in a constant chase towards an elusive "post-modernity".

If we wish to mention another convergent starting point in this innovative vision of the physical world, we should consider the figurative arts, which encompass this experience in their forms and content, offering an effective, open and universal vision, in keeping with their nature. This seed was sown at the end of the 19th century in the refined pictorial research of Georges Seurat who, during the lengthy period between Impressionism and Expressionism, used pointillism[14] to entrust the abstract power of the pixel to canvas. In its dual role as vector and follower, the pixel transcends the progressive shades of colour to lend subjectivity and authority to every point of the palette and, with Vincent Van Gogh, to every brushstroke. It is a technique which, thanks precisely to its pointillist matrix, can display different views of reality, with more or less detail, more or less depth: from the scant brushstrokes of a landscape synthesis, indicating the application of a fairly loose mesh for each individual stroke (*Figs. 1 and 2*), to an analytical breakdown of colours, placed side by side in an infinitesimal grid, in which the individual pixels almost vanish in the overall image, only to reappear upon closer analysis of the detail (*Figs. 3 and 4*).

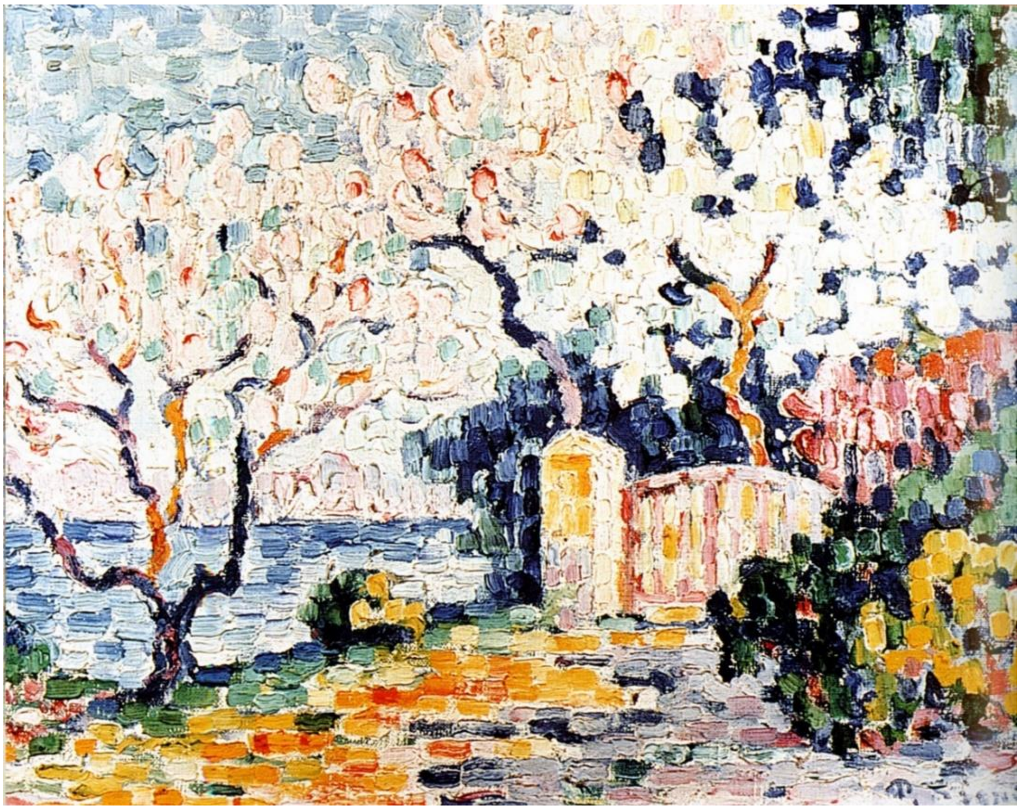

Fig. 1 Paul Signac: "Trees in bloom". Late 20th century

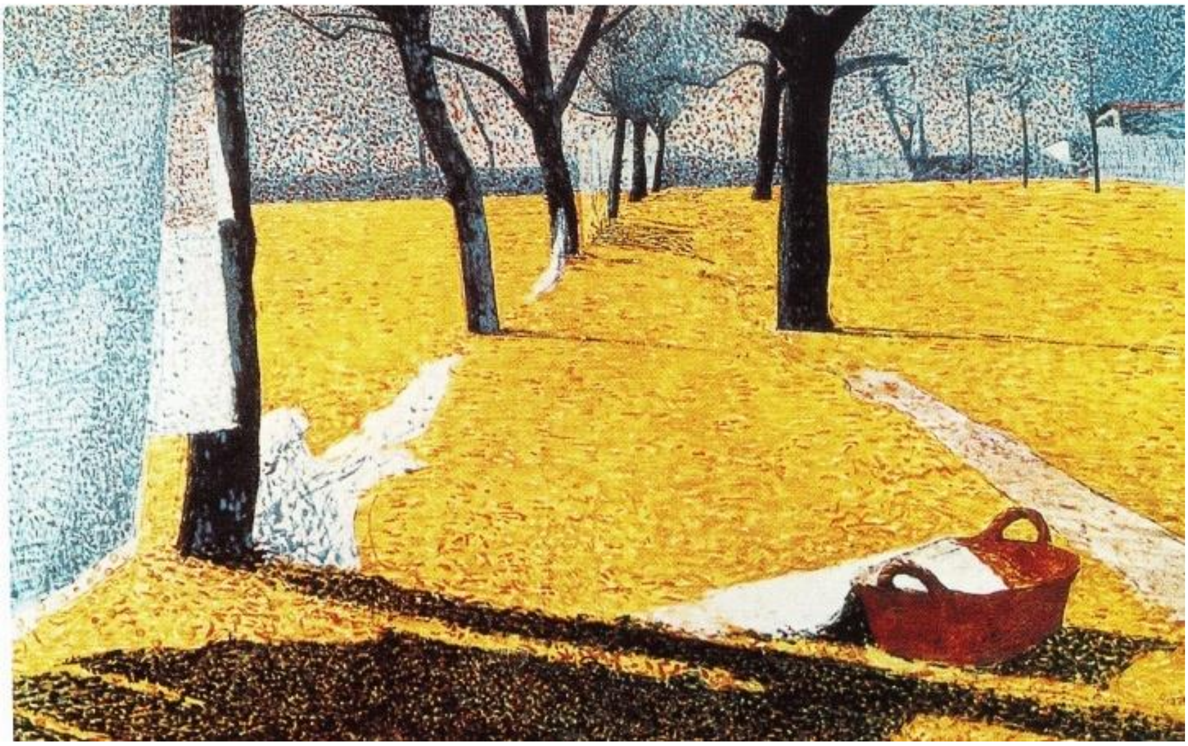

Fig. 2 Giuseppe Pellizza da Volpedo: "Washing in the sun". 1894

Fig. 3 Georges Seurat: "A Sunday Afternoon on the Island of La Grande Jatte". 1884-86

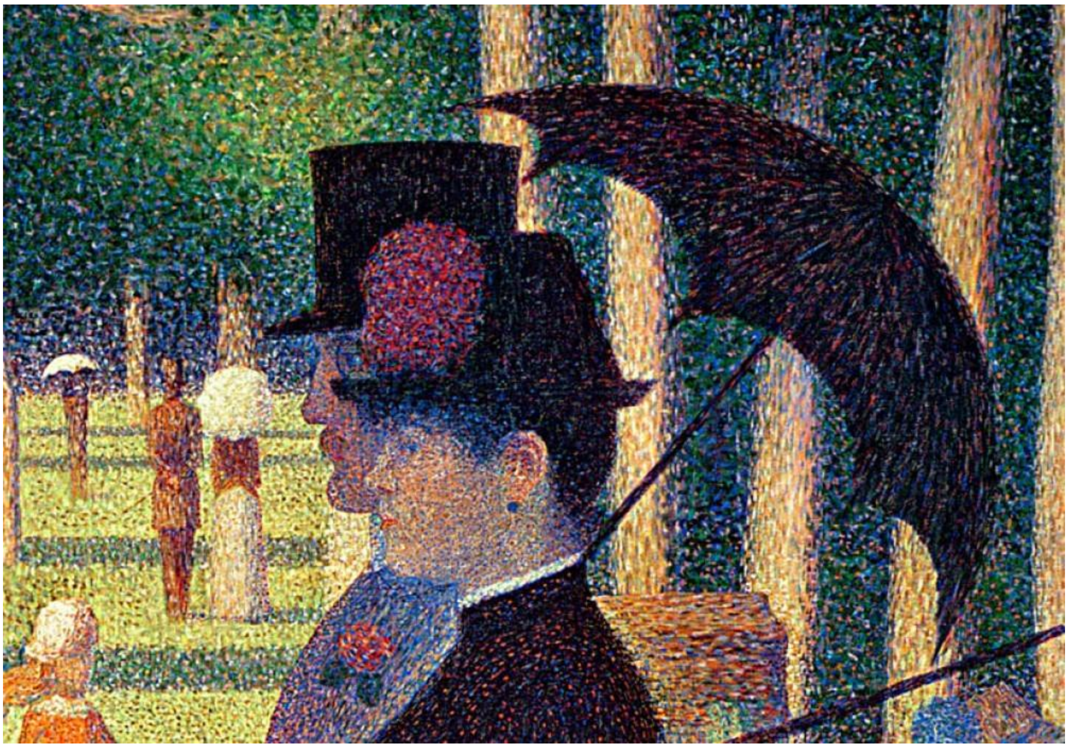

Fig. 4 Georges Seurat: "A Sunday Afternoon on the Island of La Grande Jatte". Detail. 1884-86

Reality materialises and chooses its own forms of representation, the product of a threefold contemporaneity: of the contingent expressive and cultural moment, which provides the technique and technology used, of recent and remote influences, with their nostalgia and their signs, and, lastly, of the often elitist and isolated future vision.

These are profound archetypes, proposed here in their most recent or contemporary interpretations, but which have distant and remote origins in myth, and in its gradual evolution towards transcendence. Nowadays, this arcanum is entrusted, at least in part, to digital technology, and converges in an anthropological portrayal of man in a frantic search for his place in the universe, through a process characterised not so much by an inescapable temporal constant as by a shift in spaces and contexts, the relentless liquidity of which drives us to constantly seek safer islands, defined by the rigour of metadata.

1.3. The Italian scenario

The theme of reanalysis is touched marginally by Italian academia, with a vision limited to applications that are often detached from the economic and social context. For their part, the public meteorological institutions have insufficient knowledge or at least inadequate operational skill. This is due to the severe flaw represented by the absence of a National Meteorological Service and its breakdown into a myriad of local offices and agencies, subject to constant and, unfortunately, reiterated attempts at coordination, which are pointless when they lack the necessary constitutional seal. A lack of a national meteorological and climatological vision thwarts the need for a higher level of knowledge, limits the commitment of offices to operational activities and reduces international exchange, presence and coordination.

This structural poverty has opened the doors to commercial incursion by European and non-European companies, which present large-scale reanalysis products or products obtained through pseudo-reanalysis.
In actual fact, Italy's historical heritage of meteorological data would allow the elaboration and proposal of highly refined reanalysis products. Italy has an important meteorological monitoring system made up of thousands of *in situ* stations, both public and non-public, the national network of meteorological radars and numerous private lightning monitoring networks;

Moreover, the "Public Information Guidelines" [15] issued by the Agency for Digital Italy of the Presidency of the Council of Ministers, as well as the ISTAT "Charter of services"[16] provide a very advanced regulatory programme in terms of the dissemination of data and the involvement of society in adding value to meteorological observations, which should be considered as common public property.

1.4. The international and European scenario

The WMO[4] devotes considerable attention to this technology. Its prospects and development are of strategic importance as it is required to coordinate a set of observation systems distributed, in different densities and with varying types of instrumentation, throughout the world, and used for a variety of purposes which are not always climatological. Reanalysis is capable of systematising and adding value to this articulate geography. This is because the datasets obtained produce continuous field estimates based on several sources of observation, data for all points in space and time, and also make it possible to obtain meteorological variables that are not easily or regularly observed. Consequently, it helps increase meteorological knowledge in areas which had previously been poorly monitored or monitored in relation to secondary parameters. This said, it most certainly cannot replace a good observation system which, despite various weaknesses, is the only way to extract a good reanalytical product.

A key point in the WMO policy is represented by the indication that the datasets obtained have to be digitised to standard formats, in order to make as many observations as possible freely available and ensure that they are supplemented by an adequate framework of metadata. The WMO also presents a list of possible users:

- independent assessors,
- members of the public,
- businesses,
- society in general,
- organisations in general,
- operators in the sector,
- those with an interest in meteorology and climatology,
- those with a historical, cultural or statistical interest,
- those with an interest in global warming.

From an analytical viewpoint, weather and climate datasets obtained from observations cannot be considered as absolute values. They should be seen as points of convergence of assumptions related to the measurement of a specific atmospheric variable for a determined purpose. As such, the organisation would like to point out that there is no one set of universal analyses that can be applied in all cases. As this goal is unrealistic and inapplicable, the creation of several specific datasets for different uses, or sets of comparable data placed side by side, is recommended. This option should lead to a qualitative improvement of the datasets, as it subjects them to examination by users in the various fields of application.

At European level, this is a multifaceted issue: on one hand, the ECMWF[17], the independent intergovernmental technical body that operates effectively and to the very highest standards in the creation of global reanalysis datasets, and on the other, the Presidency of the Commission, responsible for defining communication policies. Both have worked together for several years now to develop an open and free data network in all sectors, particularly science. This commitment is summarised in the recent EU Directive 1024/2019 on "open data and the re-use of public sector information", which offers an enlightened and open vision and defines the supporting role to be undertaken by public administrations in their relations with the public, the cultural and economic sectors, and society as a whole.

Outside of Europe, and particularly in the English-speaking world, leading institutes and research centres[18] have been pursuing extensive liberalisation for years, both with regard to data obtained from observations and reanalysis datasets.

2. METEOROLOGICAL REANALYSIS (OR RETROSPECTIVE ANALYSIS)

2.1. Origin and outlook

The reanalysis technique was introduced and used to produce the first datasets in the second half of the 1990s, at a time when computer technology had developed sufficiently to support this type of particularly complex and computationally demanding modelling analysis.

The origin of this requirement lay in the scientific community's need to operate on data that is continuous and homogeneous in terms of time and space, overcoming the problem related to the high rate of discontinuity associated with traditional measures.

Subsequently, these datasets gained significant weight due to the considerable increase in demand from a broad industrial market for more definite, precise and representative meteorology. Advanced digital application sectors are now able to maximise the benefits of these products. In risk management, they play a strategic role in the assessment of new formulas or types of policy (index, parametric) and in the development of financial products like weather derivatives. In smart agriculture, they feed the DSSs[19] extensively used in agronomy and at environmental level. In the activities of Utilities and Land reclamation and irrigation authorities, and consequently in energy and the water cycle, they support the design and management of networks and resources, as well as procurement and sales strategies. This extends to all smart digital applications.

Lastly, with the insurgence of greater social sensitivity towards global warming, the cultural acquisition of climatological sensitivity is evolving fully, both among certain social elites and several advanced political entities, which have begun to accept this need, while not yet undertaking it sufficiently at practical level. The issue, which seems to have gone beyond scientific standardisation, is still the preserve of descriptive and generalist communication, which must be supplemented with a solid quantitative reference structure, to define its dimensions and to monitor any mitigating effects that might be implemented.

2.2. Definition

In meteorology, reanalysis or retrospective analysis defines the scientific method used to create a global archive of how meteorological parameters change over time. It combines simulation models with real observations to generate a synthetic assessment of the state of the atmosphere. The datasets obtained populate a regular grid system, the size and characteristics of which depend on the specific application required. In logical terms, reanalysis is able to present a more coherent past that we were aware of, and to use this knowledge to outline an equally historically reliable future.

Reanalysis is not a form of observation, but it generates data based on observations. These observations supply the essential informative content of the products and also define their qualitative level or precision. The complexities and uncertainties of the observation system, the selection of data, quality control and the correction of bias, can be crucial to the outcome of the process. In other words, reanalysis without observations is purely a modelling process and the resulting dataset is destined to be affected by all the associated weaknesses.

Reanalysis allows the development of datasets of past weather and climatic trends, both near real-time and historical, being able to reach as far as the first series of reliable instrumental observations. Consequently, it represents a crucial tool for studying climate change and understanding climate mechanisms, and, as such, can be considered one of the main developments of recent meteorology and climatology.

Fig. 5 Schematic representation of a global grid obtained using the reanalysis process

2.3. Highlights

Starting from an adequate basis of observations, reanalysis is capable of generating meteorological datasets with the following characteristics:

- complete coverage of the earth’s surface,
- extensive representation of the effective weather and climate trend across the entire geographic area covered, even in areas with no *in situ* monitoring networks,
- elimination or reduction of discontinuity and a lack of spatial-temporal homogeneity.

Another particular feature is the distribution of this data in regular grids of different size, variable on the basis of the specific application and accuracy required or possible (*Fig. 5)*.

Temporal depth also depends on the type of use, in that it has to be functional to the visibility of the weather-climate dynamics researched, which could be flattened within an excessively dilated scale.

2.4. Method

From the methodological point of view, reanalysis collects all the data measured by the various meteorological monitoring systems (*in situ* and remote sensors); these are assimilated and processed by a numerical atmospheric forecasting model, and may be post-processed using auxiliary co-variables (DEMs[20]); the output of the process comprises the distribution of the data processed in a uniform grid, of variable scale, which supplies information for each point of the domain, regardless of the presence, in that point, of a meteorological station or a sensor (*Fig. 6*).

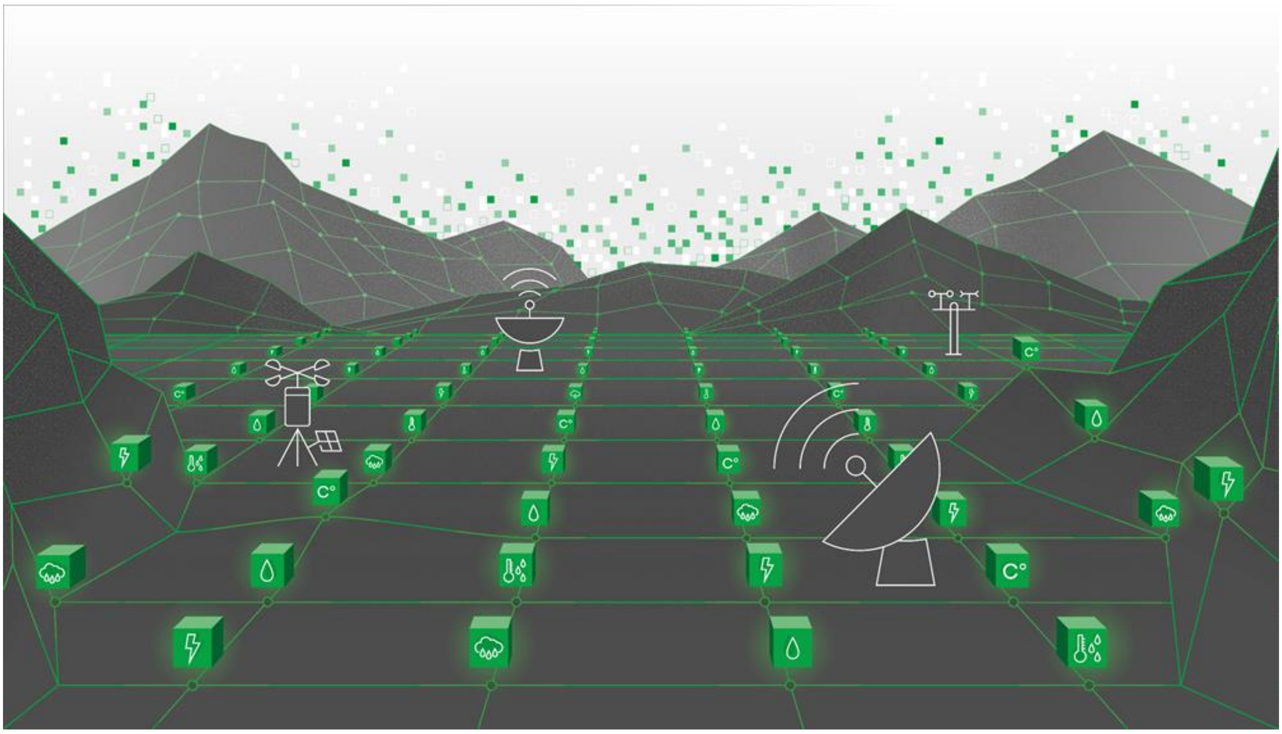

Fig. 6 The set of observations is the foundation for creating data grids using reanalysis

Two essential characterising elements are required for correct application:

- a deep and broad set of observations, which can come from different sources (*in situ* and remote),
- an atmospheric forecasting model and relative data assimilation system, the configuration of which is defined clearly and does not undergo any change during the process.

Furthermore, in specific professional uses, it is often necessary to follow the reanalysis process with post-processing operations, such as:

- downscaling using DEMs[20] with a higher spatial resolution than the original dataset,
- reduction of bias with statistical processes (MOS[21]),

- Correction of "derivative" fields, processed on the basis of the reanalysis modelling component alone, without the contribution of observations (e.g., solar radiation, precipitation, etc.).

2.4.1. Observations

As already highlighted, reanalysis is not an observation, but it generates data based on observations, which are the foundation of the process. In short, the aim of reanalysis is to consistently portray the observations available.

Being the main feature of this technique, distinguishing it from other meteorological grid production processes, observations must undergo in-depth selection. Suggesting the reading of "Characteristics and Representativeness of Precision Meteorology in Italian National Context"[3], the meteorological data used in a quality reanalysis process must come from adequate monitoring networks and use suitable and robust datasets that fall within the following categories:

Characteristics of *in situ* networks:

- **certified**: these networks are subject to formal certification procedures with regard to the type of instrumentation installed, the positioning of the survey sites, maintenance procedures and data validation,
- **WMO compliant:** these are networks belonging to associations, research institutes, NGOs, land management companies that comply with the installation, management, maintenance and validation procedures defined in the WMO Guidelines,
- **Official**: these are networks belonging to governmental bodies and organisations that are legally responsible for meteorological-environmental monitoring. The fact that they are official does not always guarantee the quality of the network and its correspondence to WMO parameters.

Characteristics of remote sensing networks, used in first assimilation and in post-processing:

- **meteorological radars,**
- **lightning detection,**
- **geostationary and polar meteorological satellites.**

Data characteristics:

- **accessibility**: standard formats,
- **continuity**: included in a consistent historical series,
- **availability**: made available according to open data criteria,
- **usability**: can be acquired promptly,
- **impartiality**: not attributable to any party,
- **transparency**: accompanied by metadata,
- **unambiguousness**: they lend themselves to a single interpretation.

Dataset characteristics:

- **continuity**: in time and space,
- **coverage**: adequate and consistent,

- **invariance**: over time of the same native dataset,
- **homogeneity**: representativeness remains constant in space and time,
- **representativeness**: defined.

Despite observing these standards, the assimilation system should also carry out a quality control on the data available and eliminate the bias from certain sets of observations (e.g.: satellite), allowing the reanalysis to obtain only those observations that are consistent with the real state of the atmosphere.

2.4.2. The atmospheric forecasting model and data assimilation system

In reanalysis, the data assimilation system is coupled with an atmospheric forecasting system; the first "forces" the system to be consistent with the observations, while the second aims to obtain physical consistency between the variables, keeping them in line with the laws of physics that govern atmospheric motion. A sufficiently realistic model is able to extract information from the parameters observed at local level and extend them to neighbouring locations, also over the course of time.

It should be noted that, in the case of precipitation, the model does not assimilate the data of the parameter but develops it as a "derivative", applying the resolution of differential equations that describe the atmospheric dynamics using other basic variables, such as temperature, humidity, wind and pressure. If assessed in a meteorological context, this method determines a significant level of uncertainty in the case of convective precipitation, becoming more contained in stratiform events.

In a climatological context, on the other hand, the offsetting of uncertainties on individual events makes this type of dataset suitable for use.

The model and basic parameters can also be used to obtain quite accurate estimates of variables that are not directly measured by the observation systems, such as solar radiation, cloud cover, ground temperature and temperatures at different altitudes.

Every reanalysis dataset uses its own specific model, characterised by the type of grid, its spatial resolution, the number of vertical levels, the height of the top level, the formulation of the physical parameters and the choice of contour conditions. Complex models usually supply higher performance results.

Fig. 7 shows an example of a data assimilation process coupled with an atmospheric forecasting model. This is a 4D-VAR system (referring to the three spatial dimensions plus time), because the assimilation of the observations extends beyond the initial instant of the forecast over a continuous window of time. The dashed red line represents the trajectory of the forecast that the model would have given without the observations. It is "corrected" during the time window of the reanalysis cycle, minimising the differences with respect to the observations. The forecast of each cycle (which can last 12 hours, for example) supplies the "first guess" of the next cycle. The process generates the meteorological reanalysis for broad time windows.

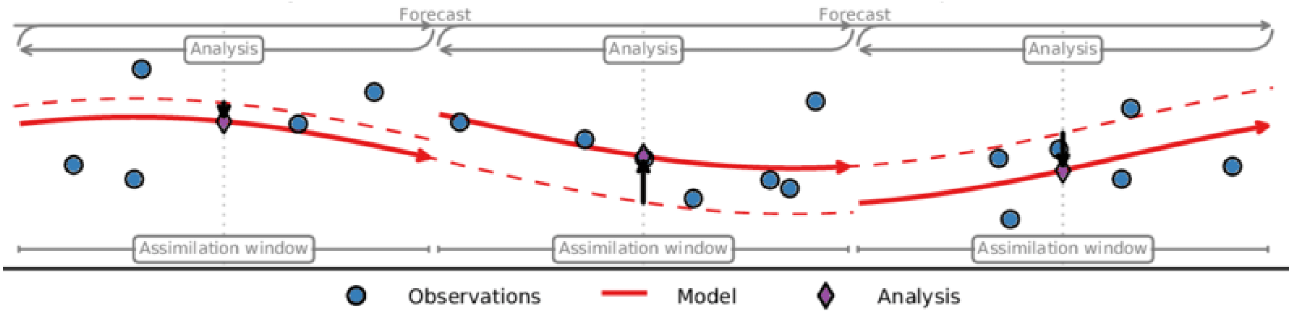

Fig. 7 Simplified schematic representation of a 4D-VAR incremental system. Source: Fujiwaraa et al. - Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems

Another essential element of reanalysis consists in the fact that the configuration of the "forecasting model-data assimilation system" binomial must be kept constant during the entire process, to ensure spatial-temporal homogeneity and prevent other spurious results. This is particularly important in the production of datasets aimed specifically at the study of climate and its variability, and at the identification of any trends linked to climate change.

Despite the application of this method, it should be noted that the portrayal of climate signals is inevitably influenced by technological updates and the reorganisation of the global observation system. This has an impact on reanalysis and on the other forms of historical analysis of observations. The global weather and climate parameters of the atmosphere, also used to estimate climate change indicators, cannot be measured directly so it is necessary to use models and statistical analyses that extract information from the measurements which, in turn, present the well-known limits in terms of representativeness and uncertainty. These issues have been subject to scientific debate, yet reanalysis is currently considered the most reliable method; the difficulty that can be encountered in assessing the uncertainty can be mitigated by a contemporary comparison with the more traditional climatic datasets obtained from observation alone.

2.4.3. Post-processing

As we are often reminded, every meteorological product should be used for and is functional to its specific application and, as such, different users (insurance, energy, hydrology, agriculture, etc.) require dedicated post-processing. Post-processing procedures make it possible to fulfil specific requirements, like the following:

- An increase in spatial (or temporal) resolution of the variables of interest compared to the original dataset. This procedure is known as downscaling and can be applied dynamically, via a high resolution modelling system that uses the data from the original reanalysis dataset, and as input or contour values, or statistically. Processing is often carried out, starting from global or continental dataset, to obtain more detailed information at national or regional level.
- Statistical correction of the output for a further reduction of the bias of "operational" variables. This correction can be applied in different ways, one of the most popular techniques being MOS[21].
- Punctual and statistical correction of the meteorological variables produced by "derivative" reanalysis, i.e., the modelling component alone, in that the observations are not acquired from the

assimilation system. A significant example is precipitation which, in the reanalysis datasets, is generated dynamically by the modelling component on the basis of basic variables such as temperature, humidity and wind. It is, therefore, necessary to apply this type of correction to precipitation in the post-processing phase, using the observations of rain gauges and meteorological radars, in order to eliminate the "spurious" drifts of the model.

Sometimes these procedures result in a situation in which the variables are no longer physically consistent, as they act selectively on target parameters. The effect achieved leads to an improvement in the result of professional and operational uses, supplying data that is more in line with reality. This may not be the case in scientific research activities.

In order to portray a specific case, we are going to use the example of the post-processing technique used by Radarmeteo[22] in the processing of the precipitation field. The RainGis® algorithm was developed to merge the data between rain gauge networks and radar data. This algorithm indicates the precipitation accumulated with a spatial resolution of 1 km; the field obtained in this way replaces that processed using the reanalysis modelling component alone.

The process envisages the following sequence:

- acquisition of radar-meteorological data,
- processing of the precipitation field accumulated from radar data alone (conversion using the *Marshall-Palmer* relationship[23]),
- acquisition of data from rain gauges,
- correction of the radar precipitation field with data from rain gauges (*merging*).

In this specific case, post-processed merging data performs much better in the portrayal of the real levels of precipitation observed. The correction of the radar precipitation field with data from rain gauges allows the consistent elimination of drifts and uncertainties derived from radar measuring while maintain information relating to the spatial variability of the precipitation. Consequently, we are able to reconstruct a precipitation field that adequately portrays localised precipitation both in terms of quality and quantity, an operation which is particularly difficult for all other processing techniques (models, satellites, interpolation of stations, etc.).

2.5. The benefits

The application of reanalysis has brought numerous benefits to weather forecasting, for which data represents the end of the process, and to countless other applications with meteorological and other matrices, for which statistically and digitally structured meteorological information is necessary. Here are the main benefits:

- it uses all the observation sources available, increasing their economic value and expanding the vocation for multiple uses,
- the output presents a high level of consistency with historical measurements,

- it carries out a cross validation between various types of sensor, which can also be independent of one another,
- the data is returned in complete, homogeneous, regular grids; for each point throughout the territory it is possible to obtain meteorological data that is representative of that area, irrespective of the presence of *in situ* weather station or other sensors,
- data spatialisation is carried out in compliance with the physical laws that regulate atmospheric motion and not using simple isotropic interpolation methods,
- the resulting datasets are homogeneous in time, being based on a constant modelling setup for the entire analysis,
- in addition to the variables measured, it is also possible to obtain those processed using the modelling component only (direct solar radiation, evapotranspiration, leaf wetness, etc.),
- it supplies a good basic meteorological knowledge of areas with scarce or inadequate network coverage,
- the datasets are immediately usable by operators, analysts and statisticians, even without a consistent background in physics, meteorology or climatology, the elements of which are dealt with beforehand.

2.6. The main reanalysis datasets

From the point of view of the geographic domain, reanalysis datasets can usually be grouped into two types:

- Global, when the computational grid or domain covers the earth's entire surface,
- Regional, when the computational grid or domain covers limited portions of the globe.

2.6.1. Global reanalysis

Global reanalysis datasets are usually processed by the world's main weather centres, the only structures which have adequate economic, scientific and computational resources to sustain such complex modelling processes. Here we are going to consider the two most important:

- ECMWF[17] - European Union
- NCEP[18] - USA

The processes are distinguished by a grid with a very broad spatial resolution of over 30 km. They perform two main functions:

- supporting the meteorological and climatological analyses of synoptic scale phenomena or phenomena which have an impact on large areas of the globe,
- supplying contour data and initialisation data for regional reanalysis.

Consequently, they can be used to study global warming and synoptic phenomena (hurricanes, non-tropical cyclones) in the scientific field and other spheres, such as risk assessment in the insurance sector.

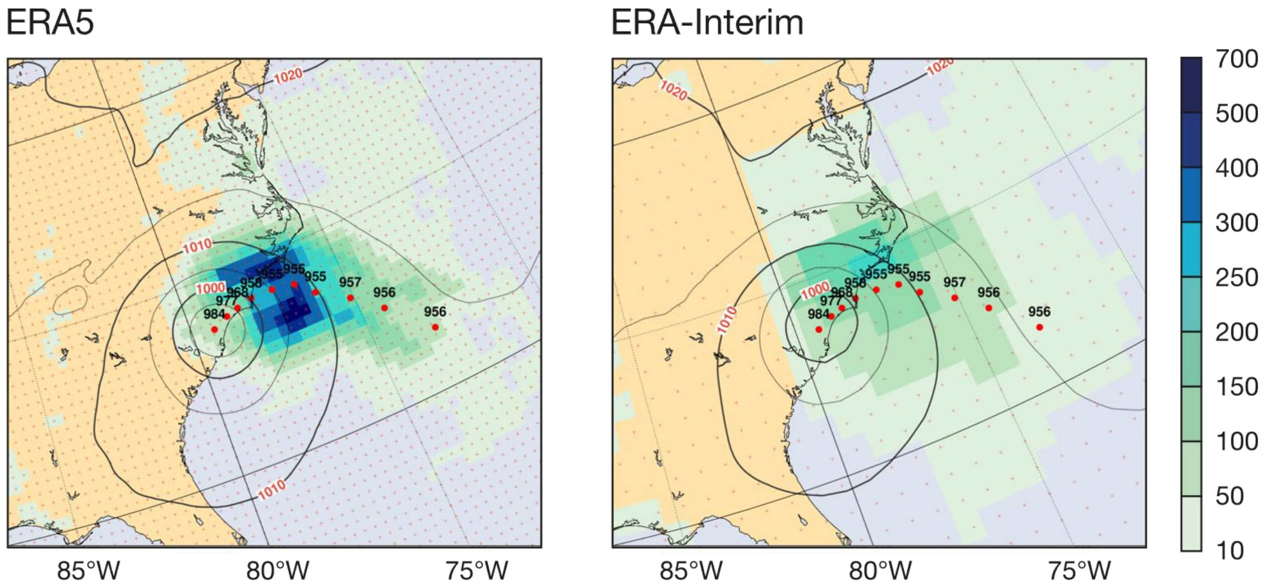

Fig. 8 Comparison between reconstructions of hurricane Florence at 09 UTC on 15 September 2018 between ERA5 (left) and ERA-Interim (right). Thanks to the higher resolution and developments of the model's physics, the ERA5 dataset supplies a more detailed reconstruction of the cyclone's intensity (lower pressure values, contours in hPa) and of the precipitation field (higher accumulations in mm, represented by coloured pixels in shades of blue). Source: ECMWF

Since the 1990s, both centres have developed global reanalysis datasets, which have gradually been upgraded, becoming more advanced; particular reference should be made to the following:

- NCEP[18] / NCAR[18] *Reanalysis 1 (Kalnay et al., The* NCEP/NCAR *40-year reanalysis project, Bull. Amer. Meteor. Soc., 77, 437-470, 1996*). Developed in partnership between NCEP and NCAR, it uses the *pro tempore* configuration of the analysis/forecasting system to assimilate the data collected from 1948 to the present day; it presents a spatial resolution of 2.5° (c.a. 250 km) and a time step of six hours.
- NCEP[18] / CFSR[24] (*Saha, S., et al. The* NCEP *climate forecast system reanalysis, Bull. Amer. Meteor. Soc., 91, 1015-1057, 2010*). The dataset was created using a high-resolution global model which simulates the coupled dynamic of the atmosphere, ocean, land surface, sea and cryosphere; it covers the period of time from 1979 to the present day, with a spatial resolution of c.a. 38 km and an hourly time step.
- ERA[25]-*Interim (Berrisford et al., 2011*). This is a global atmospheric reanalysis based on the 2006 version of the IFS (*Integrated Forecasting System* of ECMWF[17]); the system includes a four-dimensional variational analysis (4D-Var) with an analysis window of 12 hours; the spatial resolution of the dataset is 80 km and covers the time window from 1979 to 2019 with a step of three hours.
- ERA[25]*5 (Copernicus Climate Change Service, 2017*). The new dataset developed by ECMWF[17] to replace the ERA-*Interim* reanalysis; it is based on a 4D-Var data assimilation system and uses the 41r2 cycle of the IFS[26] in operation at ECMWF in 2016; ERA*5* benefits from a decade of development in model physics, core dynamics and data assimilation relative to ERA-*Interim*; it also has a significantly enhanced horizontal resolution (31 km compared to the 80 km of ERA-*Interim*) and an hourly time step; the dataset covers the period of time from 1979 to the present day (*Fig. 8*).

2.6.2. Regional reanalysis

During the last few decades, the development of global reanalysis datasets has been accompanied by the development of reanalysis projects on a regional scale. These respond to the growing need to obtain weather and climate datasets for specific geographic areas and represent the applicative tool that has allowed the use of reanalysis in various operational spheres, such as insurance, energy, water management and numerous smart applications.

The main specific characteristics of these products are that:

- they have higher spatial-temporal resolutions than global products,
- they usually require fewer grid points, consequently reducing the computational resources needed for processing, which can also be carried out by local authorities, companies and weather centres,
- they encourage more widespread use of meteorological data in many spheres,
- compared to global reanalysis, they allow the formulation of a historical background also for all those phenomena with a limited spatial-temporal scale (e.g., storms) that global models simulate in a parametrised manner or are unable to simulate at all.

In order to process a regional reanalysis, it is necessary to enter the contour conditions of the geographic domain considered into the model. The equations that guide the simulation can only be solved if they have access to the atmospheric conditions of the neighbouring grid cells; this requires the use, along the boundaries of the domain, of the data contained in a global dataset or a regional one which includes the area subject to processing within its territory. Obviously, this problem does not occur in global reanalyses as their domain is represented by the entire globe.

Different regional reanalysis datasets have been developed at European level, including:

- UERRA[27] *Regional Reanalysis for Europe*: this dataset was developed using the UERRA-HARMONIE modelling system of the ECMWF[17]. The observations are assimilated using a 3D-VAR system, while the contour data is supplied by the ERA[24]40/ERA-Interim global reanalysis dataset (*Fig. 9*). The spatial resolution of the dataset is 11 km and the period covered runs from 1961 to 2019, with an hourly time step. The geographic domain covers the whole European continent, part of North Africa and Greenland.
- COSMO REA[28]6: high-resolution reanalysis developed by the German weather service (DWD[29]) based on the COSMO model of which it also exploits the system used to assimilate observations, with additional models for the analysis of snow coverage, sea surface temperatures and soil moisture. The contour data is supplied by the ERA[25]-Interim global reanalysis dataset; it covers the period of time from 1995 to 2019, with an hourly time step. The dataset was developed with the geographic domain centred on the European continent and it has a spatial resolution of 6 km.

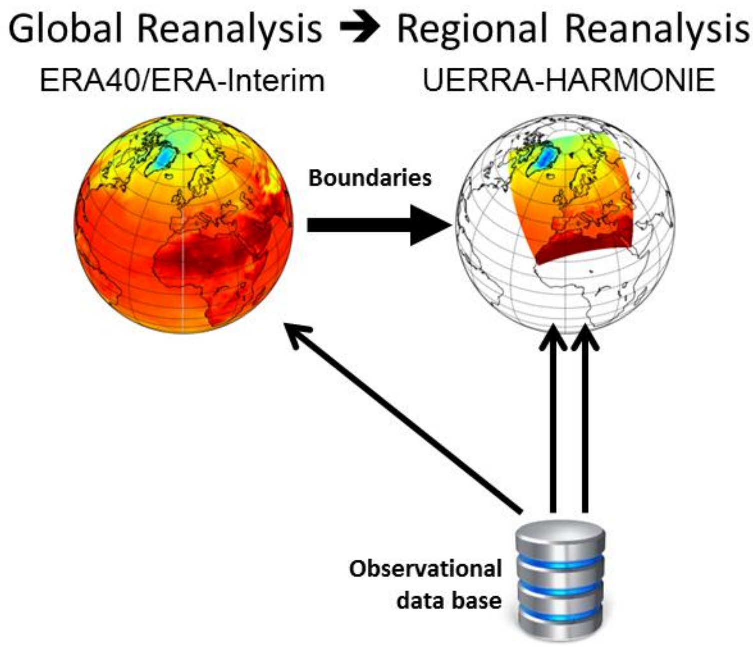

Fig. 9 Synthetic diagram of data used for the assimilation of the UERRA-HARMONY *system. Source: Copernicus Climate Change Programme: User Learning Service Content*

3. PSEUDO-REANALYSIS

The demand for historical weather and climate data is increasing significantly. This is a variable which, thanks partly to digital technology, has acquired strategic importance in industrial policies and in the development of new business models. As revealed in the previous chapters, the production of these datasets requires the application of a precise scientific methodology known as reanalysis or regression analysis. It is definitely possible to build historical and near real-time datasets also using different methods and procedures but these products offer lower and sometimes poor quality.

Nevertheless, datasets produced in this way, referred to here as "pseudo-reanalysis", are sometimes used in more general applications. A brief description is given to provide a knowledge of their characteristics and allow the assessment of possible fields of use, along with their limits, in that:

- they can be proposed as reanalysis datasets,
- the buyer may not be able to tell the difference,
- very low prices might lead people to prefer them.
-

Systematics of pseudo-reanalysis datasets

- ***In situ* observation datasets**
  - *In situ* meteorological sensor data, as measured,
  - Grids of interpolated *in situ* meteorological sensor data, without the modelling component.
- **Datasets detached from *in situ* observation**
  - Data from Hindcast or Historical re-forecast,
  - Data from the forecasting model output archive.
- **Indirect and/or remote observation datasets**
  - Satellite data,
  - Meteorological radar data.

3.1. *In situ* observation datasets

3.1.1. In situ **meteorological sensor data, as measured,**

In situ meteorological sensor data, more commonly known as weather station data, is definitely one of the best-known types of data and, therefore, the most sought after. It consists of the value of a given parameter, measured precisely at different time intervals (daily, hourly, etc.), and the resulting historical series. Due to application in reanalysis as the only form of input, and therefore in the formation of robust datasets, data from ground stations presents a number of weaknesses.

Applicability

- Historical series obtained from station data alone can only be applied in a type of analysis that is limited to the area around the station itself, rapidly losing consistency as the area of interest expands.

Limitations

- Intrinsic shortcomings:
  - limited, deficient, interrupted time series that cannot be reconstructed in any other way,
  - malfunctions and measuring errors that cannot be checked with data from other sources,
  - availability of few variables, usually temperature and precipitation,
  - replacement and modification of sensors over time.
- Shortcomings in terms of representativeness:
  - being point data, it is often inadequate to represent a meteorological phenomenon as a whole,
  - representativeness quickly ceases to be precise the further away from the station we move and the greater the orographic complexity, and with small scale (temporal) phenomena,
  - discontinuous territorial distribution.
- Environmental shortcomings:
  - alteration in the surrounding environment (constructions, agriculture, trees, etc.),
  - site relocation,
  - availability on land only.

3.1.2. Grids **of interpolated *in situ* meteorological sensor data, without the modelling component**

The formation of grids with the mere interpolation of data gathered by *in situ* sensors can be used to overcome some of the fragilities that characterise the use of data alone.

Applicability

- Production of regular data grids, with which it is possible:
  - to obtain complete coverage of the territory,
  - to mitigate the lack of homogeneity in the distribution of measurement stations,
  - to make up for the absence of data in certain areas,
- Production of continuous fields in time, with good historical depth,
- Low computational costs,
- Short processing times.

Limitations

- Availability of datasets limited to standard variables (temperature and precipitation) only; the dataset often fails to include other parameters and rarely even includes the variables derived,
- Isotropic distribution of data; while using auxiliary co-variables to "guide" the interpolation of data (e.g., DEM[20] for temperature), spatial distribution takes place evenly in various directions, unlike the physical processes that guide atmospheric motion,
- Need for a high density of stations, particularly for precipitation,
- Reduced representativeness in areas covered by few sensors,
- Coverage of land only,
- Time scale often limited to the day, in that the historical data available is often grouped together on a daily basis,

- Production of grids with a spatial resolution that is not particularly high (10-25 km),
- Geographic domain limited to national or even regional level, due to the restrictions or differences imposed by neighbouring countries and, within Italy, by the different bodies that supply data at local level.

3.2. Datasets detached from *in situ* observation

3.2.1. Hindcast or Historical re-forecast

In oceanography and meteorology, the term Hindcast or Historical re-forecast usually refers to the product of a numerical atmospheric simulation model covering a particular historical period during which no observation was assimilated. In other words, this is the re-performance of forecast model runs over a particular period of the past. The element that makes this product different from reanalysis is the limited nature of the observations, which are used only in the model initialisation phase; consequently, their guiding role, i.e., their correction and guidance of the process, constantly aligning it to a t_o that corresponds to the station measured, which is as real as possible.

Applicability

- Better knowledge of the model's climatology, of its behaviour, to provide a more accurate interpretation of the forecasts developed in the operational phase,
- Availability of output data in regular grids that are both spatially and temporally complete,
- Distribution of information in the grid according to the physical processes that regulate atmospheric motion, unlike datasets derived from the mere interpolation of station data.

Limitations

- Model drifts, meaning the limited adhesion to that actually observed. In the case of phenomena characterised by small spatial-temporal scales (e.g. convective precipitation), the model can simulate events that may be very different from those that actually occurred. Consistency with reality is poor both for daily and long-term data, as no statistical corrections are made, using MOS[21] techniques for example,
- Significant computational resources are required for analysis over broad time windows and for high spatial resolutions.

3.2.2. Data from the forecasting model output archive

The archive of operational forecasting model output is a similar product to the Hindcast, simply archiving forecast output products during their period of operation.

During that time window, however long it may be, the model setup is subject to changes; these are mainly software upgrades to improve the forecasting performance. This is the difference between this product, which has undergone changes that may improve the output of the model, and the Hindcast product, in which the model is re-run in a predetermined configuration.

Applicability

- The generation of the dataset does not require any computational resources in addition to those used for the operational forecasting model; in short, the dataset completes itself, albeit over a period of years.

Limitations

- Model drifts: as with Hindcast, the dataset presents a limited consistency with that actually observed,
- Lack of homogeneity: being the simple archiving of the forecasting output processed during the operation of the model, its new setups, such as the spatial resolution of the computational grid, will cause a lack of spatial and temporal homogeneity, which might be significant.

3.3. Indirect and/or remote observation datasets

3.3.1. Satellite data

Although satellites have been used for several decades, it is only recently that a considerable increase in the number of sensors and orbits has increased their usability. The satellites used in meteorology fall into two categories: polar and geostationary. Polar satellites orbit close to the poles and intersect the equator almost perpendicularly; their revolution time is about 100 minutes. Geostationary satellites, on the other hand, follow a circular orbit over the equator, with a revolution time of 24 hours, so they always “capture” the same portion of the globe, moving together with it. Satellite observations use the different channels of the electromagnetic spectrum, particularly the infra-red and visible channels, to detect certain meteorological variables through indirect measurements.

Applicability

- Complete coverage, even independent of other weather data sources; this benefit is emphasised in parts of the globe where monitoring is limited, such as poorer countries, remote areas, deserts and oceans,
- Acquisition of meteorological variables such as cloud cover, temperature, precipitation and wind speed,
- Acquisition of specific variables which would otherwise be hard to measure, such as cloud top temperature, significant information in monitoring severe thunderstorms, or water vapour content,
- Immediate acquisition of large-scale data.

Limitations

- Temporal continuity of datasets linked to the frequency of passage of polar satellites; the problem is solved to some extent by the increase in their number. Some applications, however, are limited by it,
- Significantly uncertain measurements, due to the fact that they are indirectly estimated (e.g., light scattering for measuring wind speed); for some datasets, such as precipitation, the uncertainty is

partially mitigated by correction/calibration by other sensors, such as weather radars or rain gauges; this is only possible in areas where these sensors are present, and is also usually achieved with a limited number of ground sensors,

- Partial geographical availability of certain measurements, such as that of wind, which is limited to the oceans,
- Interference from cloud cover, precipitation, particular concentrations of water vapour; these are factors which have a negative impact both on measurement uncertainty and the continuity of the data; in extreme cases, these conditions can lead to loss of measurement,
- Medium-low spatial resolution, 10-25 km at most,
- Limited historical depth: the datasets compiled using the latest technologies go back a maximum of five years. Those with a greater historical depth frequently present a lack of temporal homogeneity. This is caused by the upgrading of instrumentation over the years and, consequently, the initial use of obsolete technologies,
- Experimental nature; much of the technology used is still being tested and researched. This means that there are operational limitations, or limited possibilities for supplying continuous and efficient services.

3.3.2. Meteorological radar data

Meteorological radar is used for the real-time monitoring of precipitation systems within its range, which for C-band radars is usually around 100-200 km. The operating principle is based on the interaction between the electromagnetic radiation emitted by the instrument, characterised by a wavelength compatible with the typical size of hydrometeors, and the precipitation systems in the atmosphere (rain, hail, snow). Based on the intensity of the signal returned to the aerial and the time between the emission of radiation and its reception, the instrument estimates the intensity and geographical position of precipitation phenomena with a good degree of accuracy. The WMO[4] considers it to be an unconventional instrument, as it is not used in all countries. This does not, however, invalidate its usefulness, as it is fundamental in operational applications, both to obtain an extremely clear and synthetic real-time picture of storms and to track their imminent evolution (now casting). Radar makes a significant contribution also in terms of reanalysis of precipitation fields.

Applicability

- The high spatial and temporal resolution that characterises radar data allows the adequate detection of very localised and intense precipitation phenomena that often escape conventional rain gauge networks. It is therefore capable of providing a fairly precise picture of the spatial variability of the precipitation field which, in the case of convective phenomena, can be very strong and not easily portrayed with other measurement systems,
- It is an important precipitation measurement source in areas without rain gauges.

Limitations

- Precipitation is estimated on the basis of an indirect measurement (scattering of the radiation emitted by the aerial), so there can be a significant amount of uncertainty,

- Certain factors, such as the presence of mountains/hills, signal disturbance generated by other telecommunications systems and even the intensity of the phenomena themselves can negatively influence the uncertainty, the spatial homogeneity of the measurement and the area covered. While there are techniques that can mitigate some of these effects, in some cases they can significantly impair the quality of the measurement, leading to large underestimates or overestimates. If the precipitation systems are particularly intense, they can reduce the radar signal and cause an underestimation of phenomena which are “in the shadow” of the sensor's field of view. Hail can lead to an overestimate of the measurement of precipitation as it returns a more intense signal to the radar aerial than other hydrometeors with the same water content; the exact opposite happens with snow,
- Radar provides the measurement of the precipitation taken at a certain altitude, which may be quantitatively different from the precipitation that falls on the ground,
- Radar networks are usually installed for the purposes of real-time monitoring and civil protection. Consequently, the use of radar data for generating precipitation datasets has only been applied relatively recently and has limited historical depth. Datasets with greater historical depths often suffer from a lack of temporal homogeneity due to the technological upgrade of the sensors,
- Radar measurements are only available on land and over coastal sea areas,
- The high cost of purchasing and maintaining radar networks can hinder the spread of this technology, which is still somewhat limited to more developed countries,
- The lack of homogeneity of the instrumentation used from one country to another, and sometimes even within the same country, can make it difficult to produce homogeneous datasets.

4. CASE STUDY

4.1. Assessment of the performance of certain types of datasets

4.1.1. Precipitation

Among the most important meteorological variables, precipitation definitely occupies a prominent position. It should be noted that it is as important as it is difficult to monitor, forecast and reconstruct historically. This is due to the marked lack of spatial and temporal homogeneity that often characterises it, and is linked directly to the considerable complexity of the physical and microphysical processes associated with its generation.
In an extremely simplified but rather widespread scheme, precipitation can be divided into two large groups:

- stratiform,
- convective.

As always, in meteorology there are no such things as closed compartments and there may well be convective structures within stratiform precipitation. Convective precipitation is characterised by a large area and relative spatial homogeneity, making ex-post analysis fairly easy. Convective precipitation, which is sometimes characterised by extreme localisation, is completely different. In particular conditions, the precipitation values that can affect a restricted area can be very high, while in adjoining areas there may be little or no rainfall at all.

The methodology
To assess the performance of the various processing methods, the resulting datasets were compared in relation to a single case study. The case-study in question was a precipitation event spread over a large part of the country, with a prevalence of connective phenomena. The total precipitation (24 hours) on 21 September 2020 was considered. A total of 64 reference stations[30] (*Fig. 9*) fitted with rain gauges were identified and the following data for each one was compared:

- the data measured against the data extracted from the grid point of the dataset at the control station, for datasets consisting of grids,
- the data measured against the data taken from the nearest *in situ* station, for the dataset created using this method, and therefore not structured in a grid (so only the first of the list of datasets).

The data from the stations was not used in the creation of the datasets, so it was used to form the independent control sample. The quantitative comparison was carried out using some standard metrics described later on.

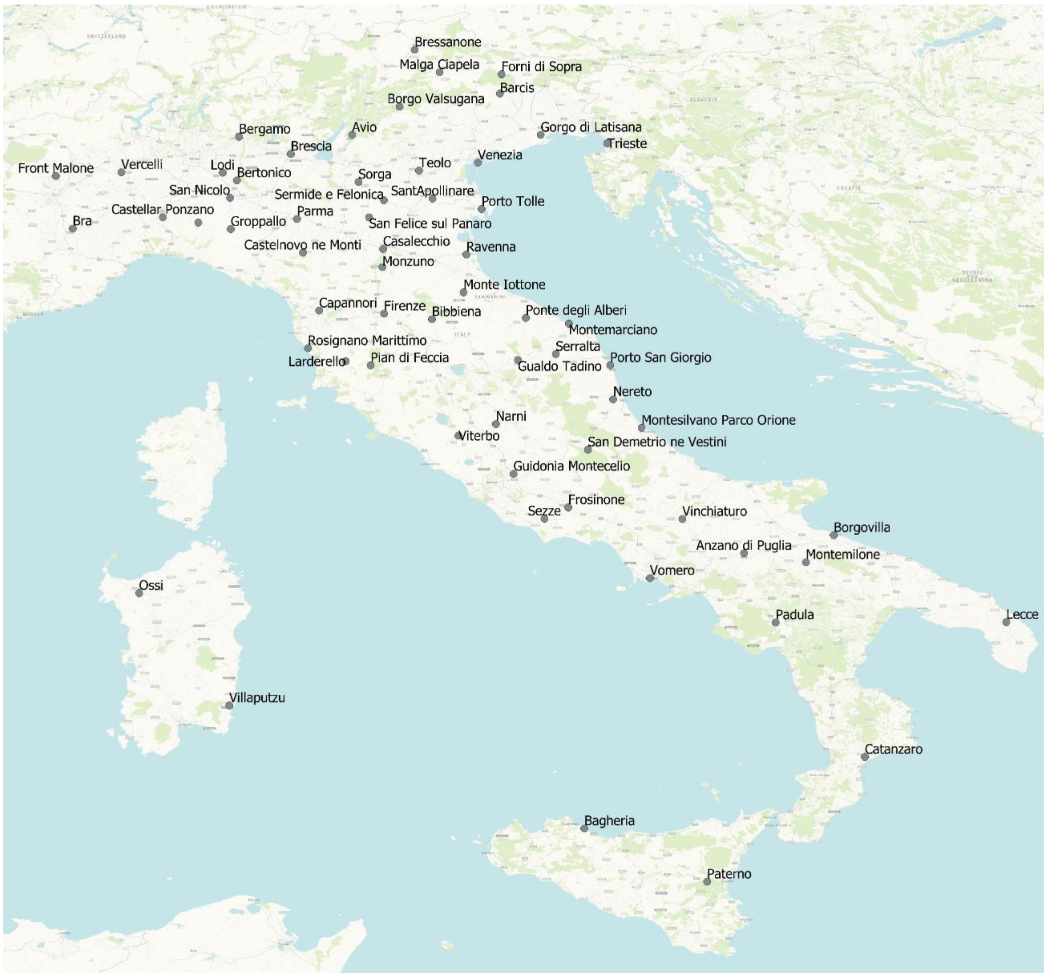

Fig. 9 The 64 reference stations used to compare the performance of the various datasets for the "precipitation" parameter

Datasets

The specific datasets used for the case study were the following:

- Daily rain gauge data from *in situ* stations, as measured: dataset consisting of the precipitation data measured by the stations closest to the 64 reference stations[30]. In other words, the nearest rain gauge station was identified for each of those stations, and the data measured by it was used to create the dataset. This dataset "simulates" the method of using the data from the nearest station to reconstruct a meteorological event in a particular location.
- Grid of interpolated rain gauge data from *in situ* stations: dataset consisting of the interpolation of rain gauge data from about 4000 *in situ* stations[30] falling within at least one of the following categories: certified, WMO[4] compliant and official. These stations do not include the 64 reference stations. The data is presented in a regular grid with a resolution of 1 km.
- Grid of daily rain gauge data from the forecast model output archive: dataset consisting of a regular grid with a resolution of 1 km showing the simulated rain gauge values:
  - from the forecasting run of the WRF model[31] (initialised with ICON data[32] and implemented by Radarmeteo[22]) carried out at 00z hours on 21 September 2020 for the first 12 hours;
  - from the forecasting run of the same model carried out at 12z hours on 21 September 2020 for the second 12 hours.

- Grid of daily rain gauge data reconstructed using satellite measurements: dataset consisting of daily rainfall supplied by the IMERG dataset[33] presented in a regular grid with a resolution of 1 km (downscaling from the native resolution of 0.1°).
- Grid of daily rain gauge data reconstructed using radar measurements: dataset consisting of daily rainfall obtained from the measurements of the national radar system alone and presented in a grid with a spatial resolution of 1 km.
- Grid of rain gauge data reconstructed by integrating measurements from radar and rain gauge (RainGis®): dataset consisting of daily rainfall obtained by integrating station data and radar data, created using the RainGis® algorithm and presented in a regular grid with a spatial resolution of 1 km. This is the method developed and used in operations by Radarmeteo[22] for the post-processing of the precipitation field within the reanalysis dataset.

Fig. 10 shows the maps reconstructed using grid-based datasets (all of them apart from the first, which is based on data from the *in situ* stations), with the aim of highlighting the fact that the same precipitation event is presented in different ways by different methods.

The checks

In order to quantitatively check the performance of the datasets, the precipitation values measured by the 64 test set stations[30] and the values returned by the different datasets at the grid point closest to the control station were compared.

The analysis was based on the following metrics:

- Coefficient of determination (R^2): this provides an estimate of the dataset's ability to provide values close to those measured. The value of R^2 varies from 0 to 1: values close to 1 indicate that the dataset closely approximates the values measured by the 64 control stations,
- MAE – Mean Absolute Error. The lower the index value, the more accurately the dataset is able to estimate the precipitation that has fallen at test set points, R
- MSE – Root Mean Square Error. Like the MAE, it provides an estimate of the mean error, emphasising the biggest errors thanks to the quadratic term. Low index values indicate good performance by the dataset.

A contingency table for four reference precipitation thresholds (1 mm, 5 mm, 10 mm, 20 mm) was created to obtain further analysis metrics. The contingency table makes it possible to obtain four combinations between the data observed and the data provided by the dataset analysed:

- Hit: the dataset and the measuring station both show a value equal to or above the reference threshold,
- Correct Negative: the dataset and the measuring station both show a value below the reference threshold,
- Miss: the dataset shows a value below the reference threshold, while the data measured is equal to or above that threshold (meaning that the dataset has underestimated the event),
- False alarm: the dataset shows a value equal to or above the reference threshold, while the data measured is below that threshold (meaning that the dataset has overestimated the event).

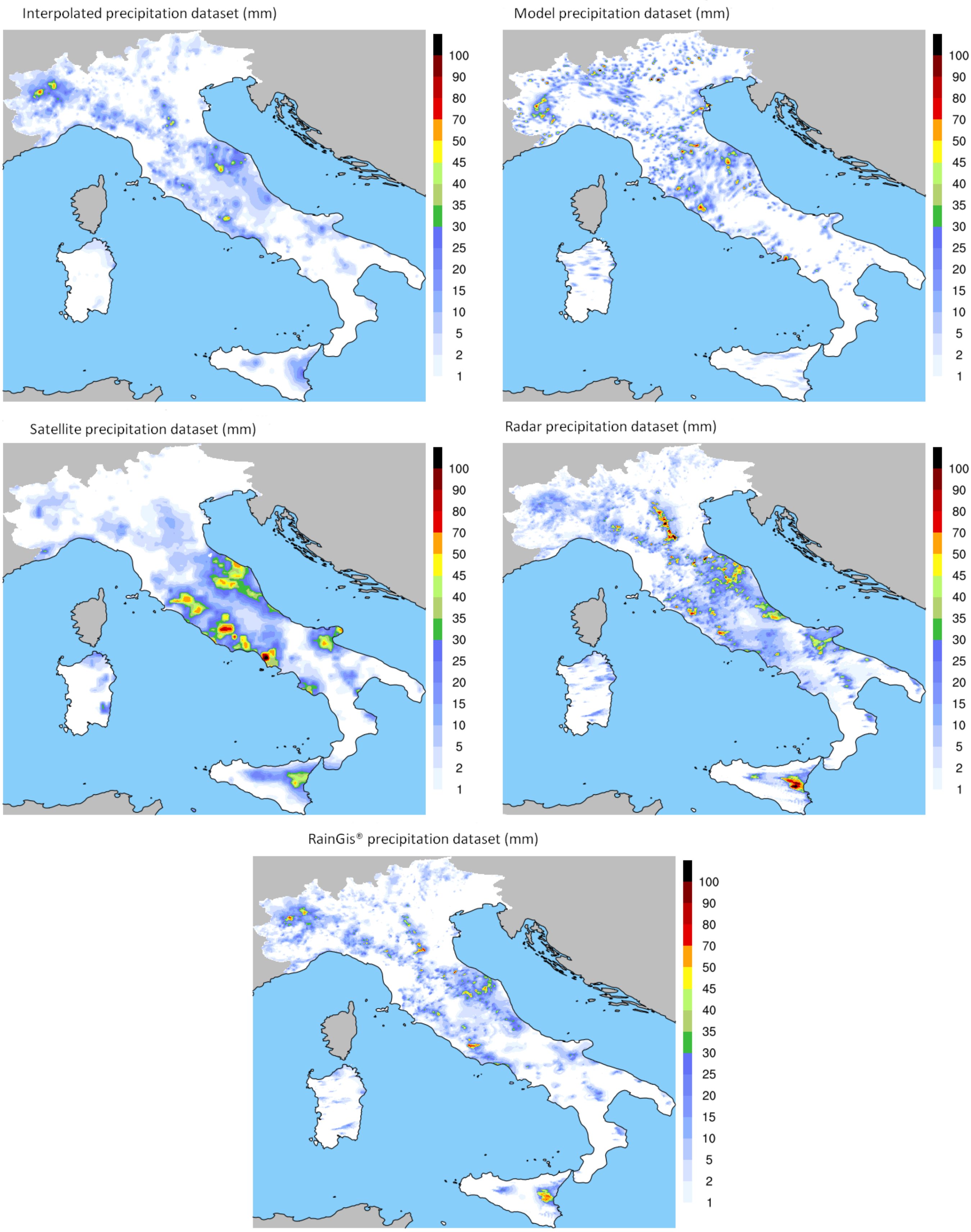

Fig. 10 Daily precipitation on 21/09/2020 reconstructed in the various datasets

The statistical metrics obtained from the results of the contingency table, which allowed further analysis of the performance of the datasets, are:

- ETS – *Equitable threat score* (*Gilbert skill score*): provides a measurement of the events correctly portrayed by the dataset,
- FAR – *False Alarm Rate*: provides a measurement of the extent to which the dataset returns precipitation estimates above the reference threshold compared to the number of correct estimates,
- FBIAS – *Frequency BIAS*: returns an estimate of the dataset's tendency to overestimate (values >1) or underestimate events (values between 0 and 1).

In theory, a dataset without error would be characterised by:

ETS=1

FAR=0

FBIAS=1.

The results obtained are shown in *Fig. 11*.

Final considerations

The dataset obtained from the RainGis® algorithm, developed by Radarmeteo[22], which integrates rain gauge observations with radar observations, shows the best performance in all the metrics used. Part of this performance can be explained by the good correlation between the dataset and the precipitation data obtained from radar measurements only: this dataset is part of the input data of the algorithm and, although its error is not negligible (RMSE >10 mm), it displays a good coefficient of determination (R^2 = 0.63) and good results in the contingency table metrics, usually second only to the database obtained from the RainGis® algorithm. This is presumably due to the radar's excellent ability to “intercept” the strong spatial discontinuity of convective precipitation.

The datasets consisting of data from the closest stations and interpolated data show similar results, tending to favour the interpolated data, and are in the middle of the performance ranking of the various datasets.

The dataset consisting of satellite measurements shows significant errors and quite poor performance in the metrics obtained from the contingency table.

The dataset consisting of archive modelling data is the dataset with the poorest performance among those analysed: the coefficient of determination and the ETS index, which are very close to zero, show that the data correlate very poorly with the control measurements. This occurs particularly with convective phenomena, which are adequately simulated at macro-scale level but not always accurately identified at local level.

The precipitation field displayed by the different datasets in the province of Turin for the event of 21 September 2020 is shown by way of example. It is indicated using a chromatic scale (from yellow to purple) superimposed with the total daily precipitation values recorded by rain gauges in the area (numeric point values). The aim is to compare the different degrees of consistency between the data obtained from the different datasets and that actually observed in the event.

Fig. 11 Diagrams of the verification metrics of the different datasets. The representative value of the best performance is highlighted for each metric.

Key:

a. PREC_NEAR: rain gauge data from in situ station, as measured;

b. PREC_INT: grid of rain gauge data from in situ station, interpolated;

c. PREC_MOD: grid of rain gauge data from the forecasting model output archive;

d. PREC_SAT: grid of rain gauge data reconstructed on the basis of satellite measurements;

e. PREC_RAD: grid of rain gauge data reconstructed on the basis of radar measurements;

f. PREC_REA: grid of rain gauge data reconstructed by integrating measurements from radar and rain gauges (RainGis®).

A comparison of the precipitation fields portrayed by the different datasets in the Turin area reveals the following:

- The dataset obtained from interpolated data (*Fig. 12*) follows the values measured by the rain gauges exactly, because it is based on their interpolation. However, the method used is unable to detect events or characteristics of the precipitation field that are not detected by the rain gauges: in other words, the interpolation method "fills" the space between one rain gauge and another isotropically and gradually, starting from the known point values only. This means, for example, that the field cannot highlight peaks caused by storm phenomena that are not intercepted by the rain gauges: this is a particularly negative point in a case of localised convective precipitation like that analysed.
- The dataset obtained from modelling data (*Fig. 13*) shows a very low correlation with the data measured on the ground: precipitation peaks are positioned further south than they actually were, leading to a general underestimation of precipitation in areas where it rained and an overestimation in areas where there was little or no precipitation.
- The dataset obtained from satellite measurement data (*Fig. 14*) shows a general and evident underestimation of precipitation values, although the spatial correlation is slightly better than the model dataset.
- The dataset obtained from radar data alone (*Fig. 15*) shows an excellent spatial relationship with rain gauges despite the fact that, at least in this case, it presents a general underestimation of the values. In contrast to the dataset obtained by interpolating the rain gauge data, it is able to identify localised events and local characteristics of the precipitation field.
- The dataset obtained from the RainGis® algorithm (*Fig. 16*), which supplements the radar dataset with rain gauge data, provides a precipitation field capable of highlighting localised events, thanks to the radar measurement, and is quantitatively more accurate, thanks to the calibration carried out by the rain gauges. Consequently, it presents an excellent degree of spatial correlation and fewer discrepancies in the estimates compared to the data from the *in situ* stations.

Moving on from ex-post meteorological analysis to climatological analysis, especially over several decades, it is logical to expect results that are generally less diversified, although the marked lack of spatial homogeneity caused by the convective component could continue to strongly influence the degree of uncertainty, especially when using purely modelling datasets such as hindcast or historical re-forecast, which offer poorer performance.

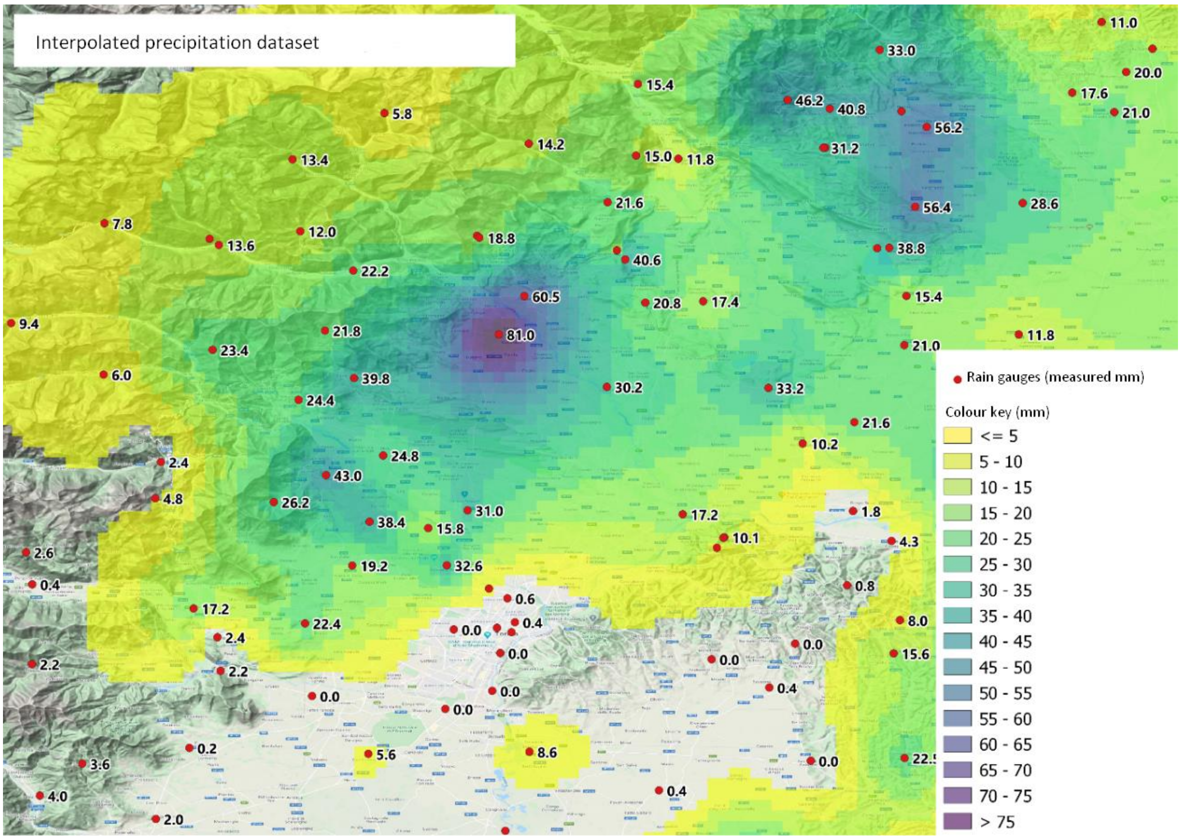

Fig. 12 Map of the precipitation field obtained from the dataset of interpolated data from in situ stations superimposed with rain gauge point data.

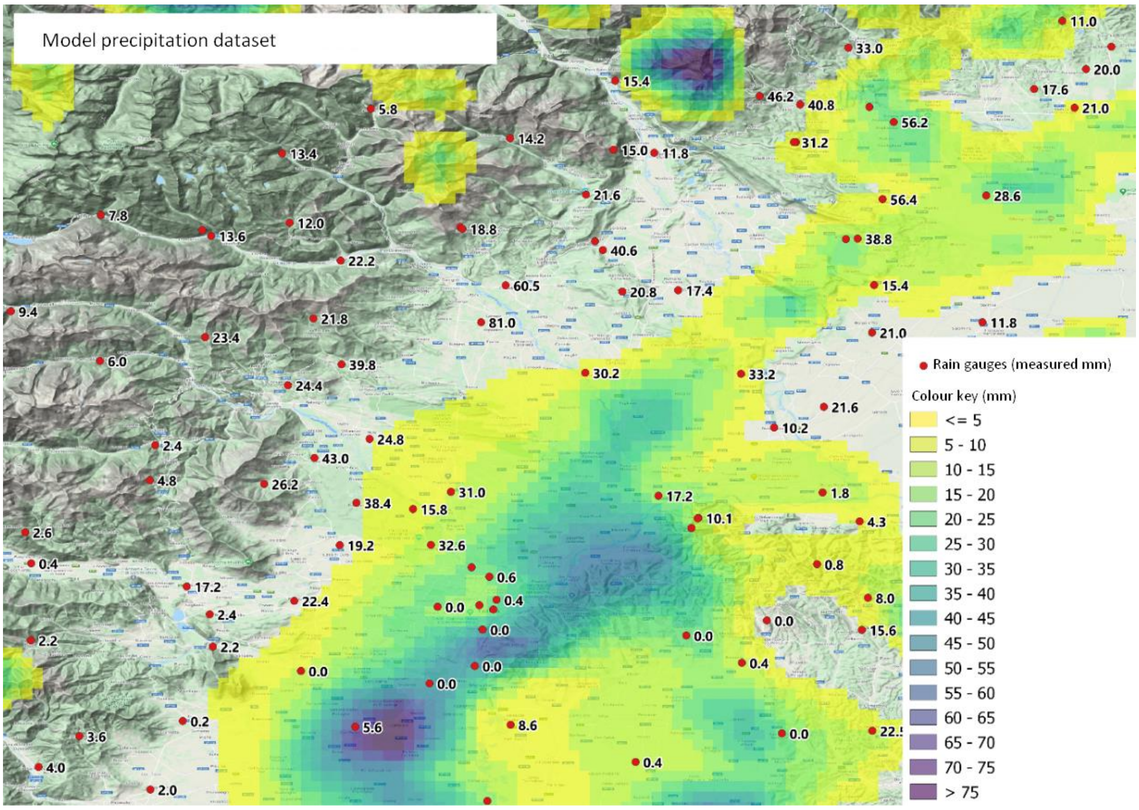

Fig. 13 Map of the precipitation field obtained from the dataset of the output archive of forecasting models superimposed with rain gauge point data

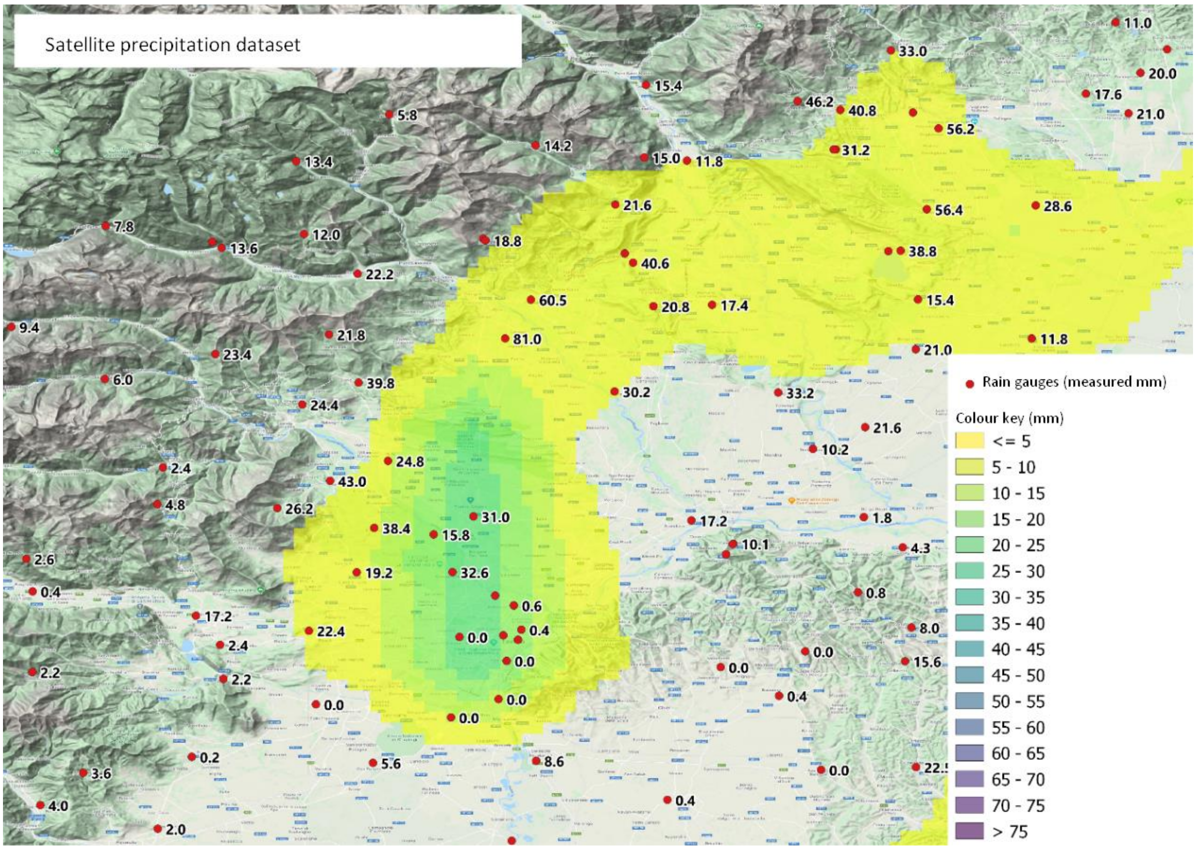

Fig. 14 Map of the precipitation field obtained from the dataset of satellite data superimposed with rain gauge point data

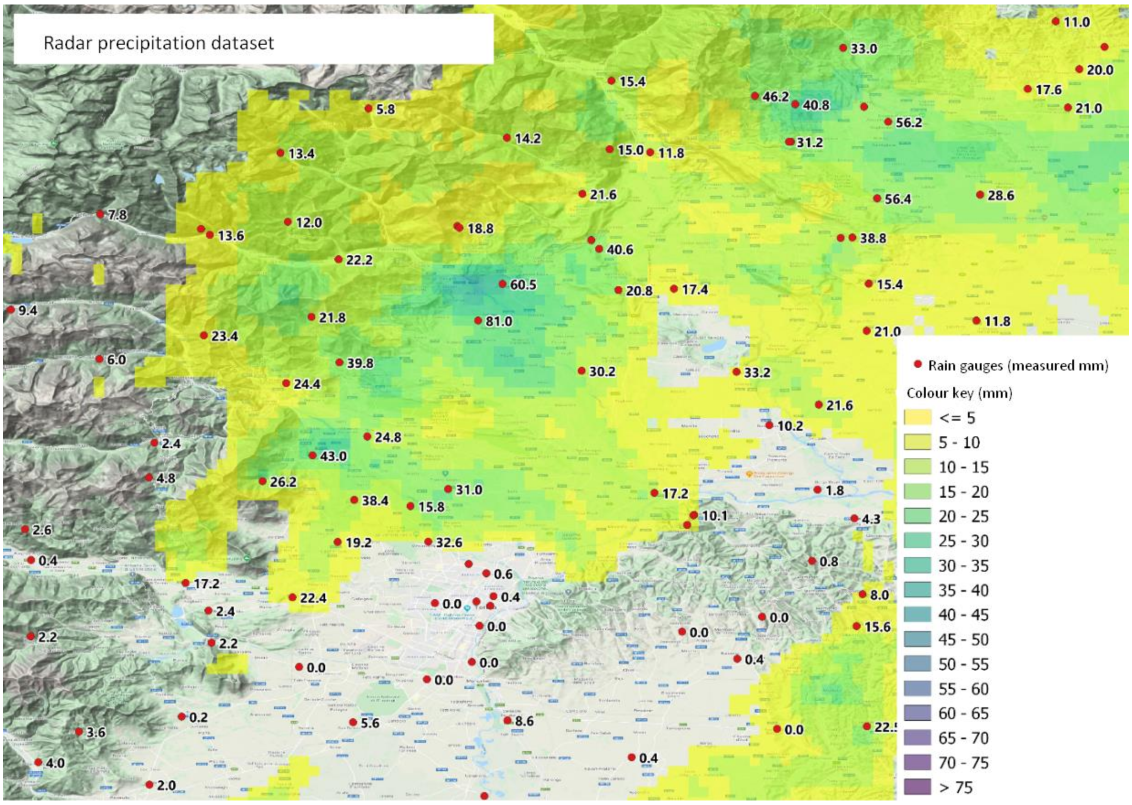

Fig. 15 Map of the precipitation field obtained from the dataset of radar data superimposed with rain gauge point data

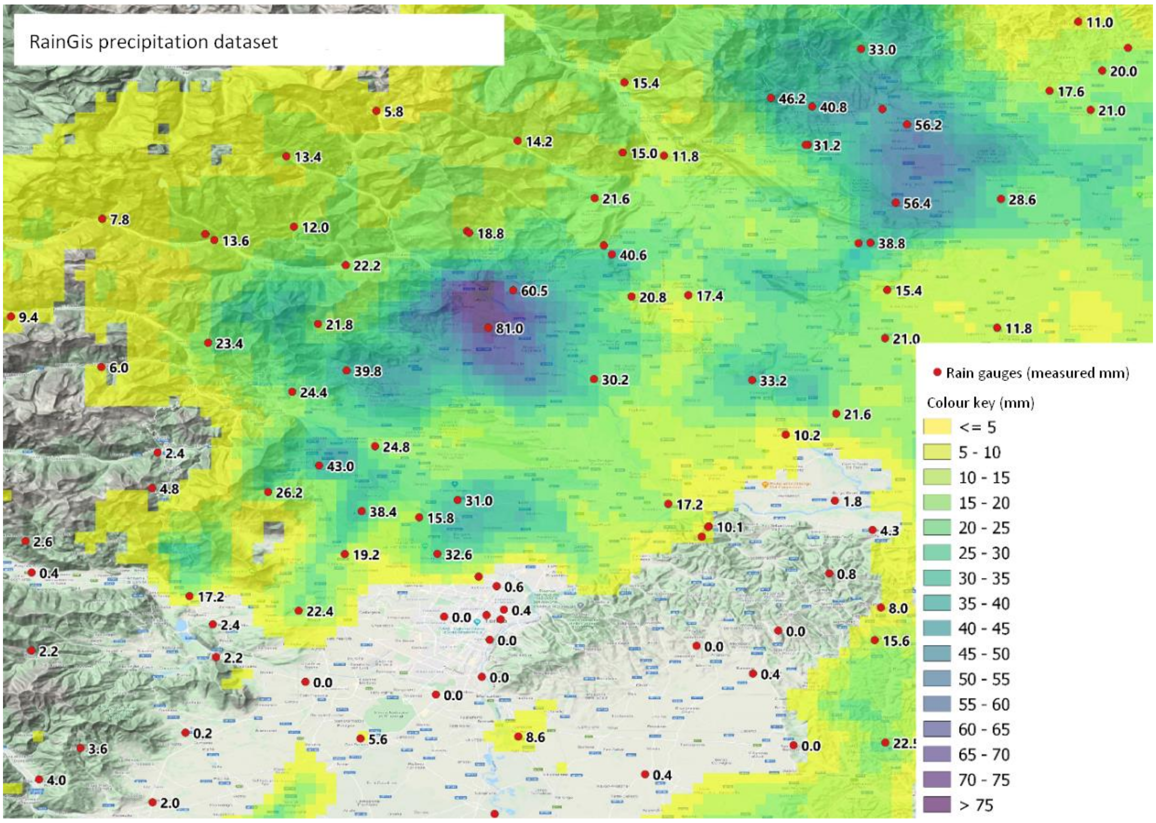

Fig. 16 Map of the precipitation field obtained from the dataset of the integration of radar and rain gauges (RainGis®) superimposed with rain gauge point data

4.1.2. Temperature

The spatial variability of temperature, especially in areas where the orography is not particularly complex, is definitely less significant than that of precipitation; it must also be said that there are no remote sensing systems for this parameter capable of supplying information with as high a spatial resolution as that which can be deduced by the meteorological radar for precipitation.

The methodology

Once again, the performance of a series of different datasets was compared over the 24 hours of 21 September 2020. A total of 93 reference stations[30] (*Fig. 17*) fitted with a temperature sensor positioned 2 metres above ground level were identified, and the following data for each one was compared:

- the data measured against the data extracted from the grid point of the dataset at the control station, for datasets consisting of grids,
- the data measured against the data taken from the nearest *in situ* station, for the dataset created using this method, and therefore not structured in a grid (so only the first of the subsequent list of datasets).

The data from these stations was not used in the creation of the datasets, so it was used to form the independent control sample. The quantitative comparison was carried out using some standard metrics described later on.

The meteorological parameter of reference is the hourly temperature observed during the 24 hours of 21 September 2020. The case study day was chosen on the basis of the precipitation; characterised by unstable and stormy weather conditions, the hourly temperature trends deviate from the curve observed on stable and sunny days, and the spatial variance is higher than usual. This situation offers a good test ground for the various datasets.

Datasets

The specific datasets used for the case study were the following:

- Hourly temperatures at a height of 2 m from the *in situ* station, as measured: dataset consisting of the hourly temperatures measured by the stations closest to the 93 reference stations[30]. In other words, the nearest thermometer station was identified for each of those stations, and the data measured by it was used to create the dataset. This dataset simulates the method of using the data from the nearest station to reconstruct a meteorological event in a particular location.
- Grid of hourly temperatures at a height of 2 m, interpolated by *in situ* stations, without the use of regression techniques that consider the auxiliary variable of the altimetry of the territory: dataset consisting of the interpolation of temperatures from about 4000 *in situ* stations[30] forming the national unified DB, from which the 93 reference stations are excluded. The data is presented in a regular grid with a resolution of 1 km.

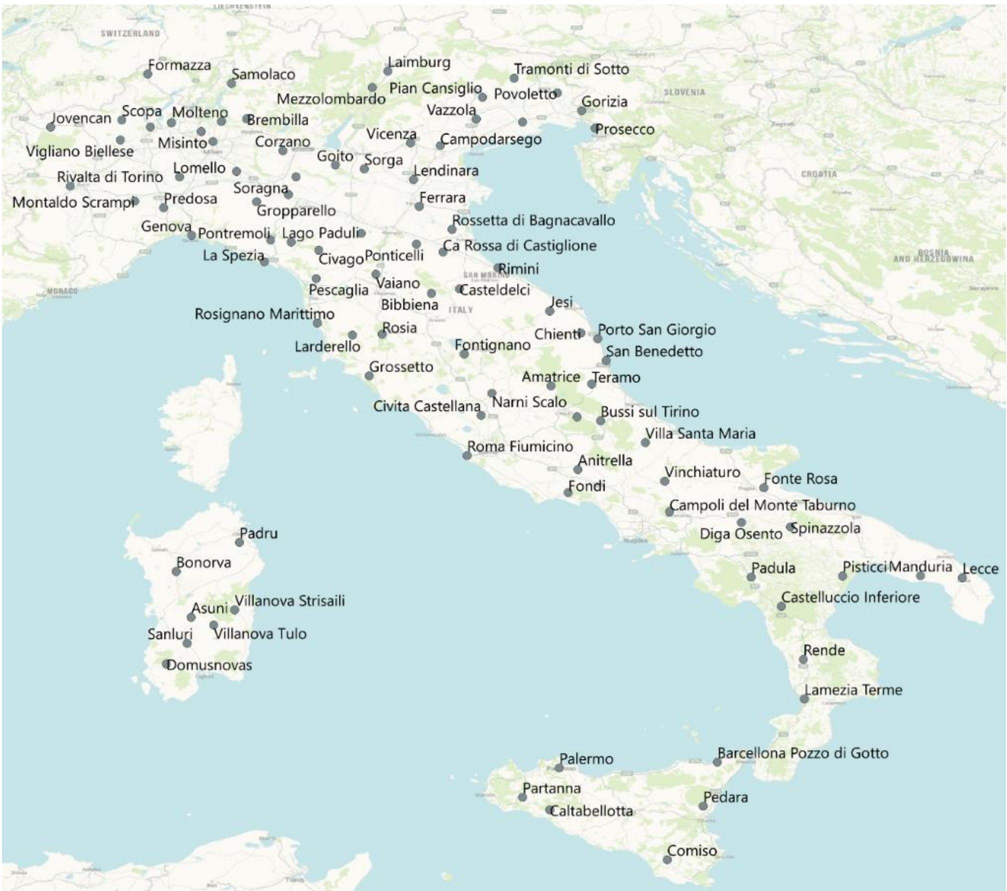

Fig. 17 The 93 reference stations used to compare the performance of the various datasets for the "temperature" parameter

- Grid of hourly temperature data at a height of 2 metres interpolated from *in situ* stations, with the application of regression due to the auxiliary variable of altitude: as above. This dataset considers the vertical trend of the temperature for the reconstruction of the field through DEM[20] with a spatial resolution of 1 km.
- Grid of hourly temperature data from the forecast model output archive: dataset consisting of a regular grid with a resolution of 1 km showing the simulated hourly temperature values:
    - from the forecasting run of the WRF model[31] (initialised with ICON data[32] and implemented by Radarmeteo[22]) carried out at 00z hours on 21 September 2020 for the first 12 hours;
    - from the forecasting run of the WRF model[31] (initialised with ICON data[32] and implemented by Radarmeteo[22]) carried out at 12z hours on 21 September 2020 for the second 12 hours.
- Grid of hourly temperature data at a height of 2 metres reconstructed through reanalysis: dataset created by reanalysing the hourly temperature data of about 4000 *in situ* stations[30] falling within at least one of the following categories: certified, WMO[4] compliant and official. These stations do not include the 93 reference stations. The data is presented in a regular grid with a resolution of 1 km

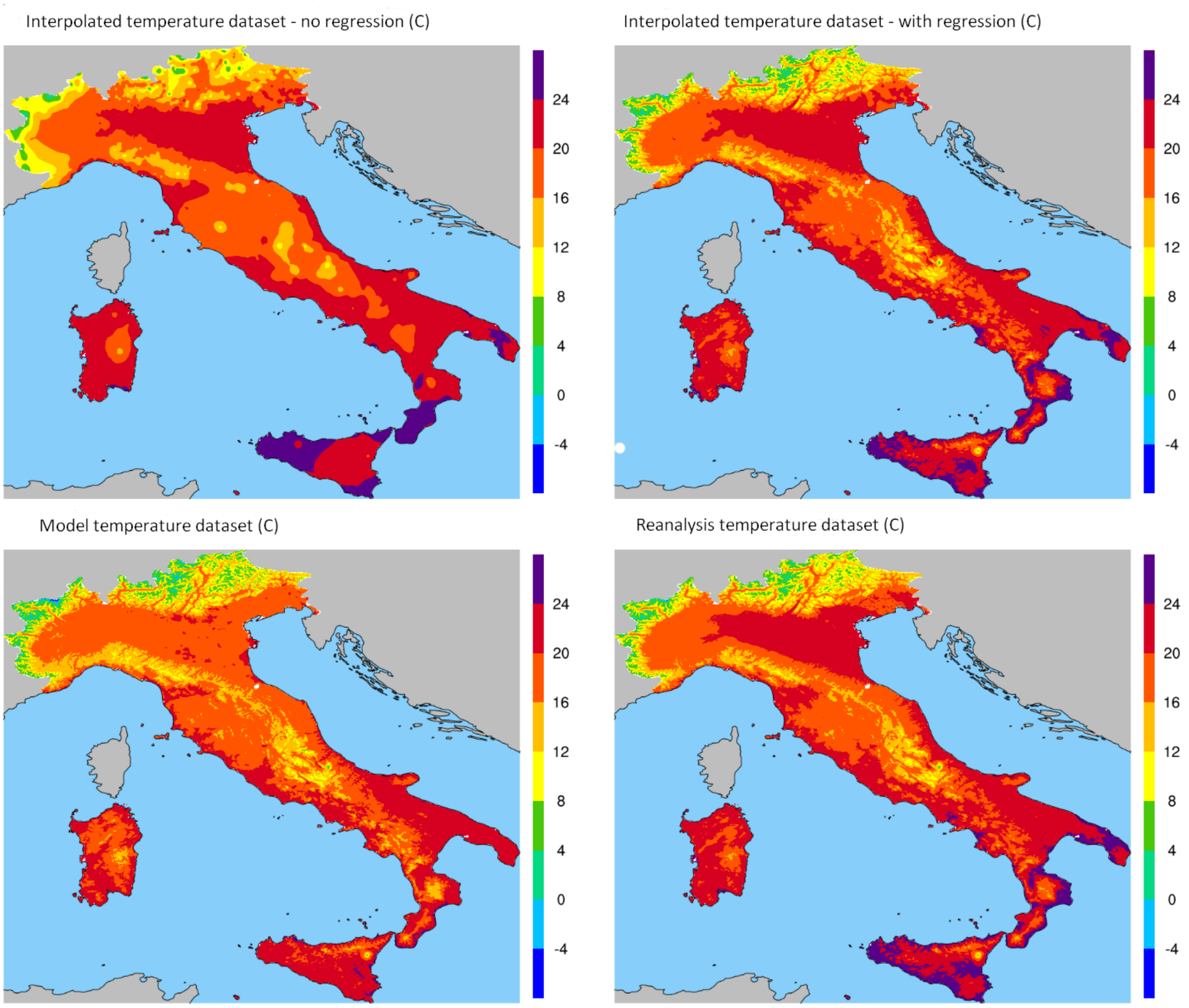

Fig. 18 Average daily temperatures on 21/09/2020 reconstructed in the various datasets.

Fig. 18 shows the maps of the average daily temperatures obtained from the different grid-based datasets (all of them apart from the first, which is based on data from the *in situ* stations). Different methods produce different fields, but the differences are less tangible than those that emerged for precipitation; it is evident that the interpolation dataset that does not use regression with the auxiliary variable of altitude provides a more approximate portrayal of the influence of orography on this field. *Fig. 19* shows the regression plots that highlight the level of agreement between the datasets and observations. In theory, a dataset with no deviations from observations would be associated with a diagram made up of points aligned perfectly with the line with equation y=x+0 and R^2 = 1. Greater dispersion of the points in the diagram and a lower R^2 value indicate a poorer performance by the dataset.

The checks

In order to quantitatively check the performance of the datasets, the hourly temperature values measured by the 93 test set stations[30] and the values returned by the different datasets at the grid point closest to the control station were compared. The analysis was based on the following metrics (the results obtained are shown in *Fig. 20):*

- Coefficient of determination (R^2): this provides an estimate of the dataset's ability to provide values close to those measured. The value of R^2 varies from 0 to 1: values close to 1 indicate that the dataset closely approximates the values measured by the 93 control stations[30],
- MAE – Mean Absolute Error. The lower the index value, the more accurately the dataset is able to estimate the hourly temperature measured at test set points.
- RMSE – Root Mean Square Error. Like the MAE, it provides an estimate of the mean error, emphasising the biggest errors thanks to the quadratic term. Low index values indicate good performance by the dataset.
- BIAS: returns an estimate of the statistical distortion of the dataset. Values above 1 indicate that the methodology used has an average tendency to overestimate, values below 1 indicate a tendency to underestimate.

Final considerations

The dataset obtained through reanalysis shows the best performance in the R^2 (0.962), MAE (0.8°C) and RMSE (1.0°C) metrics. The reanalysis shows a slightly positive bias, indicating a tendency to overestimate the data.
The datasets consisting of data observed at the stations (T_NEAR, T_INT_NOREG, T_IN_REG) present similar results, but with performances that tend to improve as the complexity of the algorithm increases.

The simplest analytical method, using data from the stations which are the closest to the control stations only, presents the highest uncertainties (MAE 1.4°C).

The interpolation dataset presents an intermediate uncertainty value (MAE 1.3°C).

The interpolation dataset which applies regression based on altitude presents the slightest error among the three (MAE 1.1°C).

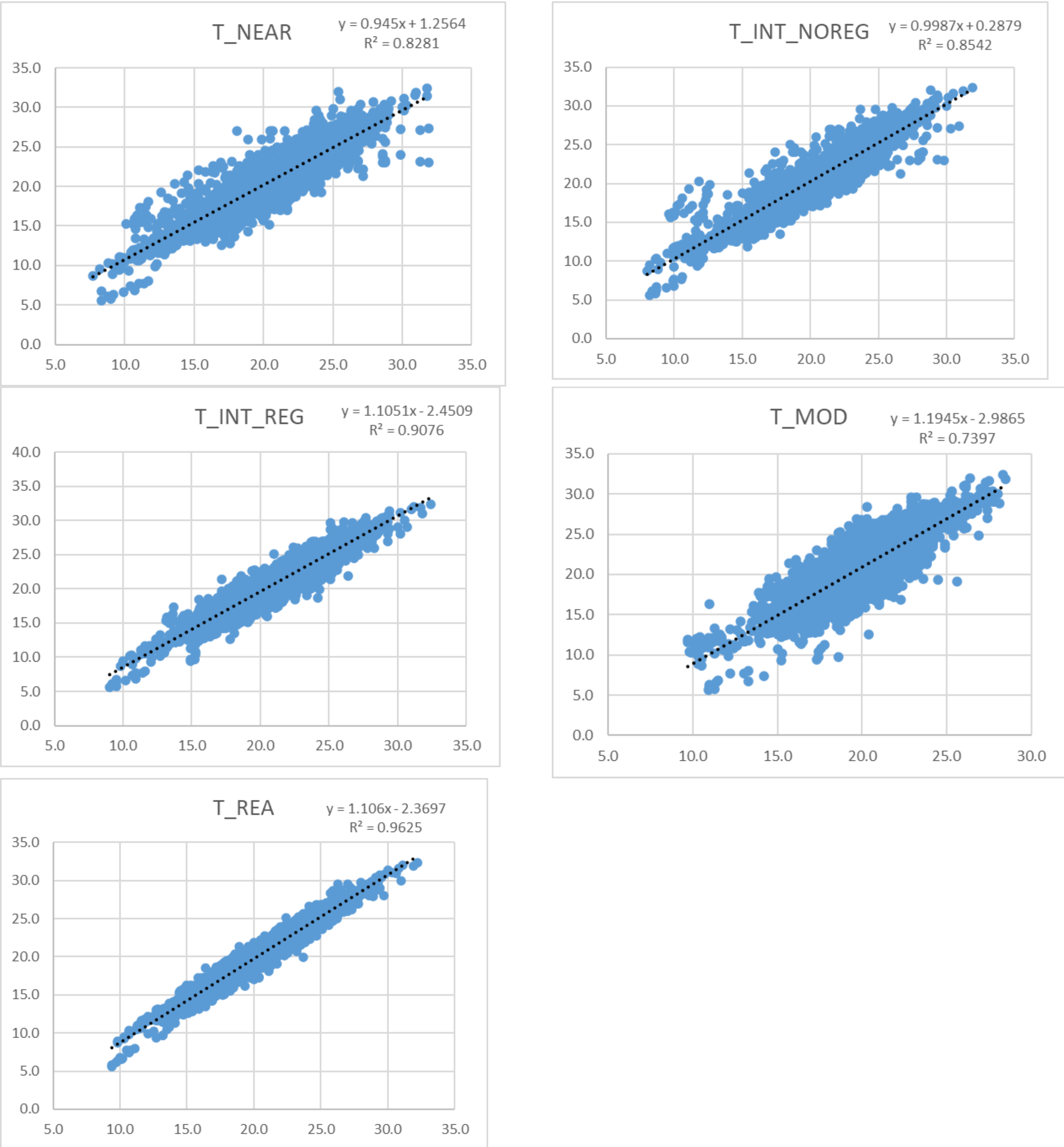

Fig. 19 Observation/dataset regression plots

Key:

a. *T_NEAR: hourly temperature data at 2 m, from station;*
b. *T_INT_NOREG: grid of hourly temperature data at 2 m, from station interpolated without regression with the auxiliary variable of altitude;*
c. *T_INT_REG: grid of hourly temperature data at 2 m, from station interpolated with regression with the auxiliary variable of altitude;*
d. *T_MOD: grid of hourly temperature data at 2 m, from forecasting model archive;*
e. *T_REA: grid of hourly temperature data at 2m, from reanalysis.*

The comparison of the BIAS is of interest: ~~;~~ it shows an underestimate in the first two datasets, and an overestimate in the dataset where regression based on altitude was used, showing that this technique improves the overall performance of the dataset (smaller errors) but tends to invert the sign of uncertainty (from underestimate to overestimate).

The dataset consisting of archive modelling data offers the poorest performance among the datasets analysed; the error is significantly higher than in the other datasets (MAE 2.0°C, RMSE 2.5°C) and the BIAS also shows an evident tendency to underestimate the values.

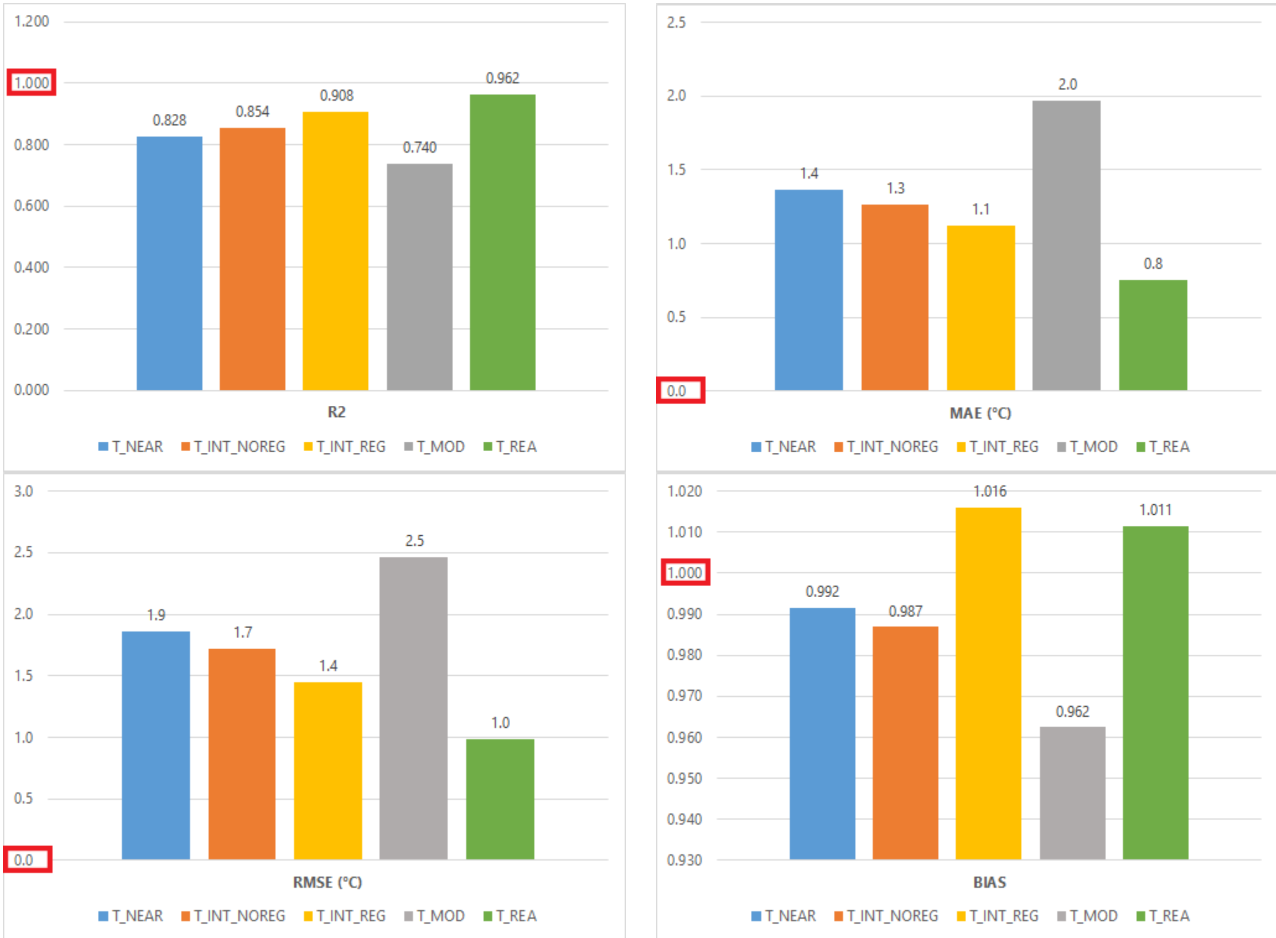

Fig. 20 Diagrams of the verification metrics of the different datasets. The representative value of the best performance is highlighted for each metric. Key: see previous figure

4.2. Parametric agricultural insurance policies based on high-resolution reanalysis datasets

Italy is undoubtedly one of the first countries in the world to have included risk management among the founding elements of its agricultural policy. Since the 1970s, Italian farms have been able to insure their crop production with multi-peril policies, protecting themselves against the qualitative and quantitative loss of product due to so-called "frequent" adverse weather events, such as hail, strong winds, flash flooding and sudden changes in temperature, and to so-called "catastrophic" adverse events, which occur more rarely but take on a systemic character, such as frost, drought and floods. The development and expansion of these insurance policies for the protection of income have been facilitated by a series of agricultural support measures, with the assignment of public subsidies to partially cover policy costs. The most recent legislation on the CAP has made specific provision for the recognition of Community aid, not only for traditional insurance policies, but also for index-based policies, which cover the loss of production insured exclusively on the basis of the trend of biological or weather-climate indices.

So, in addition to traditional policies covering damage caused by specific meteorological events, it is also possible to use policies that indemnify the negative consequences of an adverse climatic trend. An adverse climate trend is identified on the basis of the alteration of meteorological parameters, such as rainfall and/or temperature in the growing season, to the extent that they deviate significantly from the optimal curve for a given crop at a given phenological stage, negatively impacting production, which can also be measured using biological indices if necessary.

A very structured approach to these issues is naturally linked to risk management in general, as it requires long-term assessments and evaluations, even in a historical and climatic context of extreme events and consequent fluidity of territorial, environmental and economic balances. This need for in-depth analysis initiated extensive debate and the establishment of dedicated think tanks involving the scientific community, weather providers, analysts, statisticians and agronomists in the analysis of new crop models, along with the simulation and estimation of production yields in relation to meteorological parameters and associated variables. The aim of these activities is to modernise the range of insurance products available in the light of new scientific evidence.

The work teams identified the datasets best suited to achieving these aims, both in operational and commercial terms, as follows:

- historical meteorological data (10 or more years), to give the crop model the necessary input to simulate past crop performance and, so, identify the indices most closely related to the development of the damages and obtain an adequate calibration of the policy mechanisms,
- near real-time meteorological data for the constant and daily monitoring of the indices and parameters, in order to supply the policy holder with regular reports on the performance of their policy.

Another significant choice was to use high-resolution reanalysis datasets, to meet two specific needs:

- the processing of indices that are as accurate as possible, offering a good representation of an agricultural sector made up of a large number of small farms, extensive spatial variety of crops and considerable micro-climatic variability, like that which exists in Italy,
- the accurate estimation of reference weather parameters (temperature, rainfall, etc.), in order to achieve a crop damage simulation which comes as close as possible to the reality found in the field. This can only be achieved using datasets based on a consistent use of observations, as is the case of reanalysis.

The use of low-resolution databases would lead to the identification of general indices which would be unable to appreciate the extreme variability of the matrix and end up becoming detached from the punctual crop performance observed in the field. A similar approach would present an average large-scale trend, providing a damage estimate that could actually be an overestimate for certain crops and an underestimate for others.

High-resolution meteorological data grids on the other hand make it possible to associate a "virtual weather station" with each individual crop area. The area to be insured is georeferenced when taking out the policy. Then the system assigns the nearest node of the grid and the relative flow of historical and near real-time meteorological data. The high resolution of the dataset, which can reach 1 km^2, allows the assignment to the insured areas of points which are a maximum of 500 metres away. This ensures provision of adequately representative meteorological data, accepted as conventional.

The expected result, on which the very concept of an *index-based* policy is founded, is that even crop damage simulated by the crop model on the basis of the input supplied by the "virtual weather station" presents a picture in line with what actually takes place in the field. This results in completely automated operation of the policy, which can be summarised as follows:

- assimilation of the daily meteorological data and update of indices,
- verification that the thresholds of the indices are not exceeded,
- if a threshold is exceeded, the model calculates a percentage of damage to the area insured, based on the intensity of the deviation of the index from the normal situation,
- the accumulated damage is updated on a daily basis,
- when the crop is harvested, the accumulated damage is converted into economic damage and the insurance company indemnifies the policy holder automatically.

This is why index-based policies are of such considerable interest. On one hand, the completely automated process guarantees a reduction of insurance costs and, on the other, it allows the constant and transparent update of the policy, guaranteeing immediate payment, with no intermediaries, as soon as the trigger parameters are exceeded.
There is absolutely no doubt that index-based policies are going to take on an increasingly important role in risk management in the years to come, now that technology has reached levels capable of adequately supporting their development. The first experiments in the agriculture sector are showing encouraging results in this direction (*Fig. 21*).

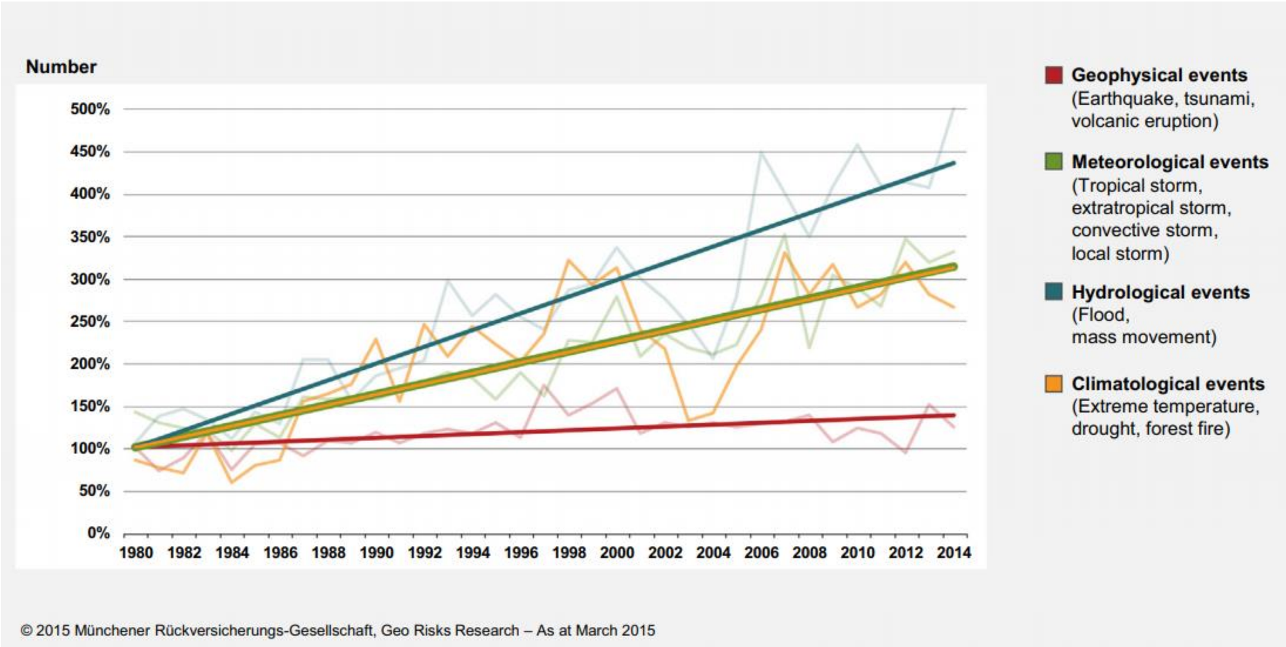

Fig. 21 Relative trend in economic losses due to natural events for the four groups of events (reference year 1980) at global level. The lines represent the linear regression of the individual curves. The trends are statistically significant for each group of events, but it is clear that weather and climate events (Meteorological events, Hydrological events and Climatological events) increase faster than geophysical events (earthquakes, volcanic eruptions, etc.). Source: Munich Re NatCatSERVICE.

5. NOTES

[1] Massimo Crespi, former Inspector of the State Forestry Corps, Director of the Experimental Centre for Avalanches and Hydrogeological Defence of Arabba (Veneto Region), Director of the Meteorological Centre of Teolo (ARPA Veneto), Director of Research and Communication of ARPA Veneto, Director General of Planning and Programming of the Veneto Region, National Delegate at the UN WMO (World Meteorological Organization), Director of the Environmental and Hydrological Monitoring Centre of the European Union in Asunción (Paraguay); currently CEO of Radarmeteo.

[2] Gianluca Ferrari: Expert in meteorology applied to risk assessment, Manager of the insurance and agricultural sector of Radarmeteo.

[3] Massimo Crespi: "Characteristics and Representativeness of Precision Meteorology in the Italian National Context" - "Quaderni di meteorologia aperta" n. 1-2020 www.radarmeteo.com .

[4] WMO: *World Meteorological Organization,* UN Technical Agency tasked with global coordination of meteorology, climatology and operational hydrology.

[5] IMO (*International Meteorological Organization)*.

[6] UN *(United Nations)*.

[7] Olga Tokarczuk "Flights" 2007.

[8] Zeno of Elea "The paradox of Achilles and the tortoise" 5th century b.c..

[9] Zygmunt Bauman "Retrotopia" 2017.

[10] Karl Popper "The Open Society and its Enemies" 1945.

[11] Henry Leborit "La nouvelle grille" 1974.

[12] Claude Lévi-Strauss "The Savage Mind" 1962.

[13] Zygmunt Bauman "Liquid Modernity" 2000.

[14] Félix Fénéon – 1870.

[15] Presidency of the Council of Ministers - Agency for Digital Italy "Public data - Public Information Guidelines" 2018.

[16] ISTAT (Italian National Institute of Statistics) "Charter of Services" 2013.

[17] ECMWF (*European Center for Medium-range Weather Forecast*): Independent intergovernmental organisation.

[18] NCEP (*National Centers for Environmental Prediction*) of *NWS* (*National Weather Service*) USA; NCAR (*National Center for Atmospheric Research)* USA; JMA (*Japan Meteorological Agency*) JAP; BOM (*Bureau of Meteorology – Australian Government*) AUS.

[19] DSS *(Decision Support System)*.

[20] DEM (*Digital Elevation Model*).

[21] MOS (*Model Output Statistics*).

[22] Radarmeteo Srl (www.radarmeteo.com).

[23] *Marshall-Palmer* 1948: this is the relationship used to convert reflectivity into precipitation intensity.

[24] CFSR (C*limate Forecast System Reanalysis).*

[25] ERA *(*ECMWF *Re-Analysis).*

[26] IFS *(Integrated Forecasting System):* operational forecasting model used by ECMWF[17].

[27] UERRA (*Uncertainties Ensembles Regional Re-Analysis*).

[28] COSMO REA (COnsortium for Small scale Modelling RE-Analysis).

[29]DWD (*Deutscher WetterDienst*) the German Meteorological Service.

[30]The weather stations used in the case studies are operated by the following bodies: the Italian Air Force, the regional environmental agencies of Veneto, Piedmont, Lombardy, Emilia-Romagna and Calabria, the Regional Civil Protection Departments of Friuli Venezia Giulia, Marche and Apulia, Meteotrentino, the Autonomous Province of Bolzano, the regional agricultural-meteorological network of Apulia, the regional hydrological service of Tuscany, the Meteonetwork Association, CETEMPS.

[31] WRF(*Weather Research and Forecasting model*) developed by US organisations ad bodies.

[32] ICON (*ICOsahedral Nonhydrostatic model*) sviluppato dal DWD[28].

[33] IMERG (*Integrated Multi-satellitE Retrievals for GPM*) sviluppato dalla NASA – USA.

[34]PAC (Community Agriculture Policy)

6. BIBLIOGRAPHY

- WMO - Kennedy J.; Dee D.; et al. (2011). On the Reprocessing and Reanalysis of Observations for Climate. WCRP OSC Climate Research in Service to Society (Denver, USA).
- Austin, P. M. (1987). Relation between measured radar reflectivity and surface rainfall. Mon. Wea. Rev., 115, 1053–1070.
- Bengtsson, L, Hagemann, S., and Hodges, K. I. (2004). Can climate trends be calculated from reanalysis data? J. Geophys Res., 109.
- Bengtsson, L., and Shukla, J. (1988). Integration of space and in situ observations to study global climate change. Bull. Amer. Meteor. Soc., 69, 1130–1143.
- Berg, E., & Schmitz, B. (2008). Weather-based instruments in the context of whole-farm risk management. Agricultural Finance Review, 68(1), 119–133.
- Betts, A. K., M.Zhao, P. A.Dirmeyer, and A. C. M.Beljaars, 2006: Comparison of ERA-40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets.J. Geophys. Res., 111.D22S04.
- Black, E., E. Tarnavsky, H. Greatrex, R. Maidment, A. Mookerjee, T. Quaife, and J. Price, (2015). Exploiting satellite-based rainfall for weather index insurance: The challenges of spatial and temporal aggregation. Proc. First Int. Electronic Conf. on Remote Sensing, 1–19.
- Bollmeyer, C., and Coauthors, 2015: Towards a high-resolution regional reanalysis for the European CORDEX domain. Quart. J. Roy. Meteor. Soc., 141, 1–15.
- Bosilovich, M. G., Kennedy, J., Dee, D., Allan, R., .and O'Neill, A. (2011). On the reprocessing and reanalysis of observations for climate. Climate Science for Serving Society: Research, Modeling and Prediction Priorities, G. R. Asrar and J. W. Hurrell, Eds., Springer, 51–71.
- Bosilovich, M.G., J. Chen, F.R. Robertson, and R.F. Adler (2008). Evaluation of Global Precipitation in Reanalyses (J. Appl. Meteor. Climatol., 47, 2279–2299).
- Caesar, J., Alexander, L., Vose, R. (2006). Large-scale changes in observed daily maximum and minimum temperatures: creation and analysis of a new gridded dataset. J Geophys Res 111.
- Compo, G. P., and Coauthors. (2011). The Twentieth Century Reanalysis Project. Quart. J. Roy. Meteor. Soc., 137A, 1–28.
- Conradt, S., R. Finger, and M. Spörri, (2015). Flexible weather index-based insurance design. Climate Risk Manage., 10, 106–117.
- De Rooy, W.C., Kok, K. (2004). A combined physical-statistical approach for the downscaling of model wind speed. Weather Forecast 19:485–495.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kaallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, T.J.N., Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system (Q. J. R. Meteorol. Soc. 137: 553–597).
- Delle Monache, L., T.Eckel, D.Rife, and B.Nagarajan, (2013) Probabilistic weather prediction with an analog ensemble. Mon. Wea. Rev., 141, 3498–3516.
- Derin, Y., and Coauthors, (2016). Multiregional satellite precipitation products evaluation over complex terrain. J. Hydrometeor., 17, 1817–1836.

- Durán, L., and Rodríguez-Muñoz, I. (2016). Automatic monitoring of weather and climate in mountain areas. The case of Peñalara Meteorological Network (RMPNP). Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2015-248.
- European Centre for Medium-Range Weather Forecasts, (2017). ERA5 reanalysis. National Center for Atmospheric Research, Computational and Information Systems Laboratory, accessed 7 September 2019.
- European Commission, (2017). Uncertainties in Ensembles of Regional Reanalyses. European FP7 Project, European Commission, http://www.uerra.eu.
- Frei, C. (2014). Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. Int J Climatol 34: 1585–1605.
- Frick, C., Steiner, H., Mazurkiewicz, A., Riediger, U., Rauthe, M., Reich, T., Gratzky, A. (2014). Central European high-resolution gridded daily data sets (HYRAS): mean temperature and relative humidity. Met Z 23(1):15–32.
- Greatrex, H., J. Hansen, S. Garvin, R. Diro, S. Blakeley, M. Le Guen, K. Rao, and D. Osgood (2015). Scaling up index insurance for smallholder farmers: Recent evidence and insights. Climate Change, Agriculture and Food Security Rep. 14, 32 pp.
- Haylock, M., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008). A European daily high-resolution gridded dataset of surface temperature, precipitation and sea-level pressure, J. Geophys. Res., 113, D20119.
- Henn, B., A. J.Newman, B.Livneh, C.Daly, and J. D.Lundquist, (2018). An assessment of differences in gridded precipitation datasets in complex terrain. J. Hydrol., 556, 1205–1219.
- Herrera, S., Gutierrez, J.M., Ancell, R., Pons, M.R., Frias, M.D., Fernandez, J. (2012). Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). Int J Climatol 32:74–85.
- Hoeppe, P. (2016). Trends in weather related isasters – Consequences for insurers and society (Weather and Climate Extremes 11, 70–79).
- Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei (2008). The comparison of six methods for the interpolation of daily, European climate data, J. Geophys. Res., 113.
- Hofstra, N., M.Haylock, M.New, and P. D.Jones, (2009). Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. J. Geophys. Res., 114, D21101.
- Hou, A. Y., and Coauthors, (2014). The Global Precipitation Measurement mission. Bull. Amer. Meteor. Soc., 95, 701–722.
- Huffman, G. J., D. T.Bolvin, D.Braithwaite, K.Hsu, R.Joyce, and P.Xie, (2014). NASA Global Precipitation Measurement Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.4, 30 pp.
- Isotta, F. A., and Coauthors, (2014). The climate of daily precipitation in the Alps: Development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. Int. J. Climatol., 34, 1657–1675.
- J. Bruinsma (2012). World agriculture towards 2030/2050. FAO
- Kalnay, E., and Coauthors. (1996). The NCEP/NCAR 40-Year Reanalysis Project. Bull. Amer. Meteor. Soc., 77, 437–471.
- Keller, J. D., L. Delle Monache, and S. Alessandrini, (2017). Statistical Downscaling of a High-Resolution Precipitation Reanalysis Using the Analog Ensemble Method. J. Appl. Meteor. Climatol., 56, 2081–2095.
- Klein Tank, A., and Coauthors, (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. Climatol., 22, 1441–1453.

- Klok, L., and A. M. G. Klein Tank (2008). Updated and extended European dataset of daily climate observations, Int. J. Climatol., 29, 1182–1191.
- Krähenmann, S., Walter, A., Brienen, S., Imbery, F., and Matzarakis, A. (2017). High-resolution grids of hourly meteorological variables for Germany. Theor. Appl. Climatol., 131, 899–926.
- Lazo, J. K., Lawson, M., Larsen, P. H., & Waldman, D. M. (2011). US economic sensitivity to weather variability. Bulletin of the American Meteorological Society, 92(6), 709–720.
- Masatomo Fujiwara, Jonathon S. Wright, Gloria L. Manney, Lesley J. Gray, James Anstey, Thomas Birner, Sean Davis, Edwin P. Gerber, V. Lynn Harvey, Michaela I. Hegglin, Cameron R. Homeyer, John A. Knox, Kirstin Krüger, Alyn Lambert, Craig S. Long, Patrick Martineau, Andrea Molod, Beatriz M. Monge-Sanz, Michelle L. Santee, Susann Tegtmeier, Simon Chabrillat, David G. H. Tan, David R. Jackson, Saroja Polavarapu, Gilbert P. Compo, Rossana Dragani, Wesley Ebisuzaki, Yayoi Harada, Chiaki Kobayashi, Will McCarty, Kazutoshi Onogi, Steven Pawson. (2017). Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems (Atmos. Chem. Phys., 17, 1417–1452).
- Mesinger, F., and Coauthors, (2006). North American Regional Reanalysis. Bull. Amer. Meteor. Soc., 87, 343–360.
- Müller, M.D. (2011). Effects of Model Resolution and Statistical Postprocessing on Shelter Temperature. Journal of applied meteorology and climatology, Vol. 50, Nr. 8. pp. 1627-1636.
- Mußhoff, O., Odening, M., & Xu, W. (2011). Management of climate risks in agriculture—will weather derivatives permeate? Applied Economics, 43(9), 1067–1077.
- Newman, A. J., M. P. Clark, R. J. Longman, and T. W. Giambelluca, (2019). Methodological Intercomparisons of Station-Based Gridded Meteorological Products: Utility, Limitations, and Paths Forward. J. Hydrometeor., 20, 531–547.
- Nurmi P. (2003). Recommendations on the verification of local weather forecasts. ECMWF Technical Memorand.
- Odening, M., Muhoff, O., & Xu, W. (2007). Analysis of rainfall derivatives using daily precipitation models: Opportunities and pitfalls. Agricultural Finance Review, 67(1), 135–156.
- Parker, W.S. (2016). Reanalyses and Observations: What's the Difference? Bulletin of the American Meteorological Society 97 (9): 1565-1572.
- Pierce, D. W., D. R.Cayan, and B. L.Thrasher, (2014). Statistical downscaling using localized constructed analogs (LOCA). J. Hydrometeor., 15, 2558–2585.
- Prein, A., and A.Gobiet, (2017). Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. Int. J. Climatol., 37, 305–327.
- Q.J.R. Meteorol Soc 137:553–597., D. P. (2005). Bias and data assimilation. Quart. J. Roy. Meteor. Soc., 131, 3323–3343.
- Ritter, M., O. Musshoff, and M. Odening, (2014). Minimizing geographical basis risk of weather derivatives using a multi-site rainfall model. Comput. Econ., 44, 67–86.
- Saha, S., and Coauthors. (2010). The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., 91.
- Schiemann, R. et al. (2010). Geostatistical radar-raingauge combination with nonparametric correlograms methodological considerations and application in Switzerland, Hydrology and Earth System Sciences.
- Sinclair, S., Pegram, G. (2005). Combining radar and rain gauge rainfall estimates using conditional *merging*., Atmospheric Science Letters, 6, 19-22.

- Stahl, K., Moore, R.D., Floyer, J.A., Asplin, M.G., McKendry, I.G. (2006). Comparison of approaches for spatial interpolation of daily temperature in a large region with complex topography and highly variable station density. Agr For Meteorol 139:224–236.
- Thorne, P., and Vose, R. S. (2010). Reanalyses suitable for characterizing long-term trends. Bull. Amer. Meteor. Soc., 91, 353–361.
- Vedenov, D. V., and B. J. Barnett, (2004). Efficiency of weather derivatives as primary crop insurance instruments. J. Agric. Resour. Econ., 29, 387–403.
- Velasco-Forero, C. A., et al. (2008). A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data, Elsevier Ltd.
- Walton, D., and A.Hall, (2018). An assessment of high-resolution gridded temperature datasets over California. J. Climate, 31, 3789–3810.
- Wilks, D. S. (2006). Statistical methods in the atmospheric sciences. International Geophysics Series, Volume 91, Department of Earth and Atmospheric Sciences, Cornell University, Elsevier.
- Williams, T. M., and W. R. Travis, (2019). Evaluating Alternative Drought Indicators in a Weather Index Insurance Instrument. Wea. Climate Soc., 11, 629–649.
- Wilson, L. J., and M.Vallée, (2002). The Canadian Updateable Model Output Statistics (UMOS) system: Design and development test.Wea. Forecasting, 17, 206–222.
- Woodard, J. D., and P. Garcia, (2008). Basis risk and weather hedging effectiveness. Agric. Finance Rev., 68, 99–117.
- Xu, W., Filler, G., Odening, M., & Okhrin, O. (2010). On the systemic nature of weather risk. Agricultural Finance Review, 70(2), 267–284.
- Yan H., Gallus W. A. (2016). An Evaluation of QPF from the WRF, NAM, and GFS Models Using Multiple Verification Methods over a Small Domain (Wea. Forecasting, 31, 1363–1379).
- Yuval, and W. W. Hsieh, (2003). An Adaptive Nonlinear MOS Scheme for Precipitation Forecasts Using Neural Networks. Wea. Forecasting, 18, 303–310.
- Zhang, J., Qi, Y., Langston, C., Kaney, B., and Howard, K. (2014). A real-time algorithm for *merging* radar QPEs with rain gauge observations and orographic precipitation climatology. J. Hydrometeor, 15, 1794–1809.

Index of Contents