Quaderno di METEOROLOGIA APERTA N°5

ON THE USE OF WEATHER-CLIMATE DATA IN INDEX-BASED APPLICATION CONTEXTS

Experiences, conventional and communicative aspects

Massimo Crespi
Gianluca Ferrari
Paolo Spagna

In Padova (Italy),
Winter 2022/23

CONTENTS

1. CONTEXT AND SCOPE

Over the past five years, Radarmeteo and Hypermeteo — the latter a recent spin off of the former and the current owner of the entire data asset and weather datasets — have had direct experience in the creation and supply of datasets linked to climate-related risks in agriculture and other exposed sectors throughout Italy. This was not an experimental or pilot activity, but rather a service all aspects of which were structured on the basis of specific contracts with almost all insurance companies and farmer consortia and with the relevant national ministerial control bodies. In sum, the service covered more than 90% of the capital insured in agricultural production and supported more than one million audits per year in the various phases of the insurance process, i.e. from risk assessment to policy underwriting and finally to claim settlement.

This was a true and proper stress test that, thanks to such widespread deployment, was able to highlight the fact that the system rested on a technologically sound basis, but which could still be subject to improvement as certain aspects may have been missed during initial implementation.

Another relevant aspect of this experience should also be noted, namely the so-called "oracle"; that is, the source of the weather data was the same for both the insurance company and the insured. At first glance this fact may seem improper, but instead, it has represented one of the reasons for the success of the service. It has practically eliminated all data-related sources of contention by applying reanalysis based on the single national data base implemented by Radarmeteo with third-party and independent data, i.e. drawn only from the official weather and weather radar networks, certified or in accordance with WMO (*World Meteorological Organization*) standards.

And yet, at contractual level, this is not enough. Credibility must be built on other aspects - transparency, consistency and uniformity of the response and communication contents - aspects that are not only technical-scientific but also truly informative. If this is not the case, the entire system would prove inadequate, ineffective and result in a loss of reputation. Critical issues, updates and new solutions imposed by technical-managerial complexity and extensive user base should also be handled in the same manner.

Moreover, technological, economic and digital transitions significantly affect these processes, as do the scenarios outlined by the climate emergency, changing their structural aspects as well. In fact, in this area, there is a growing push toward parameterisation — at least partial — of insurance logic, which assigns the weather trigger a decisive, evidential and sometimes definitive value.

All this has made it necessary to "revise" certain aspects through a more precise definition of the conventions, considering and because of the fact that this is an operational matter where, rather than doubts, solutions and clarity are required. To this end, the processes on which action has been taken are redefined, justified specifying the aspects that have undergone mitigation, revision or in-depth analysis.

2. DEFINITION OF STANDARDS FOR OBSERVATION SOURCES, DATA AND *DATASET* PROCESSING METHODS

2.1. Observation sources and data

The national observation set is made of different types of networks: on-site (ground-based) weather networks, weather radar networks, satellite networks, etc. The unified national database uses all data gathered from networks that fall into the following qualitative categories:

- **certified**: subject to formal certification procedures with regard to the type of instrumentation installed, the location of the survey sites, maintenance procedures and data validation;
- **WMO-compliant**: belonging to associations, research institutes, non-profit organisations, land management companies that comply with the installation, management, maintenance and validation procedures defined by the WMO Guidelines;
- **official**: belonging to the governmental bodies and organisations legally responsible for weather-environmental monitoring.

For further details, see Quaderno di meteorologia aperta No. 1 Characteristics and representativeness of precision meteorology in the Italian national context" published by Radarmeto in 2020.

2.2. *Dataset* processing methods: reanalysis and *post-processing* techniques

The characteristics and formats of the data collected from the various observation sources are anything but homogeneous; this, in turn, makes them difficult to use in their original form. In particular, since the goal is to obtain as complete and coherent a representation of the atmospheric situation as possible, processing techniques must be adopted so as to assimilate the data, integrating them and making them available in datasets that are both "ordered" and complete.
In meteorology, the term reanalysis, or retrospective analysis, defines the scientific method used to accomplish this processing; 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.
Starting from an adequate basis of observations, reanalysis is thus able to generate meteorological datasets with the following characteristics:

- complete coverage of the earth's surface;

- extensive representation of the actual weather and climate trend across the entire geographic area covered, even in areas with no on-site 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, which varies based on the specific application and the required or possible accuracy.
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 for example be flattened within an excessively dilated scale.
In order to make them more application-specific, reanalysis datasets are subject to post-processing systems. These procedures — which adopt innovative artificial intelligence techniques such as neural networks — significantly increase the representativeness of the datasets, particularly for those variables, such as rainfall, that are simulated by the modelling component of the reanalysis system alone, the drifts and uncertainties of which would undermine their direct and fruitful use in the application areas contemplated here.
More technical details can be found in the Quaderno di meteorologia aperta No. 2 "Reanalysis or retrospective analysis in meteorology." published by Radarmeteo in 2020.

With the widespread application of reanalysis technology, the need has emerged to emphasise certain specific aspects that could induce even significant numerical differences.

2.2.1. Management of reanalysis *datasets* updating cycles

With the ongoing evolution in technology (increased computational capacity) and analysis methodologies (new data validation techniques, more accurate models, new observation sources, etc.) updating cycles of the historical datasets are run so as to gradually achieve greater precision and representativeness. These cycles are run every few years and thus they do not occur frequently. However, they are nevertheless part of common, well consolidated processes in meteorology, affecting both datasets processed by the main world meteorological centres (NOAA, ECMWF, etc.) and those produced by local entities seeking to improve their reanalysis products. This is obtained by reducing — either statistically or through completely new processing — the uncertainties associated with the weather-climate variable estimates. Updating the dataset means that the numerical values associated with the variables in the new database will generally be different from those reported in the original dataset, even those covering the same geographic location and time frame, thus making the information ambiguous.

Mitigation action: appropriate dating of the metadata makes it possible to trace the versions of the datasets used throughout the various decision-making processes. The end users are notified when new versions of the reanalysis datasets are released so they can consider whether and how they might be acquired and used, also in light of the benefits that might be gained. In many situations, if the reanalysis dataset has been used to feed internal models, it may be deemed appropriate to re-run the computational process with the updated database, particularly if the models employ artificial intelligence techniques that need to be “re-trained” on the new version of the dataset to assimilate the new statistics.

When reanalysis datasets are used as evidence, such as in the insurance field, the fact that the new dataset has different values from those originally used in the decision-making process can generate contrasts. Take, for example, the case of a claim referring to a trigger defined on the value of a particular meteorological variable that is to be reached and/or exceeded. If, in the original data base, the value of the variable was lower than the trigger, it would not give rise to compensation, while if in the new dataset this value was exceeded, litigation might ensue. To avoid these situations, the contractual documentation accepted by the parties must specify which version of the dataset is to be used, and they must still keep track of all changes that have occurred.

2.2.2. Managing differences via pseudo reanalysis datasets

The datasets construction characteristics (technique employed, sources used, spatial resolution, etc.) affect the weather data valuation. This means that comparisons between datasets of different natures may lead to numerical differences for a given parameter — even with the same geographic coordinate and the same time instant — thus generating a dichotomy that is not easy for the end user to resolve, as he/she is unable to make a choice towards the most representative dataset, i.e. the one best suited to the specific context. This happens with some frequency when the comparison is made between a reanalysis product and a pseudo-reanalysis product, i.e., when the historical data processing technique is not reanalysis, but instead originates from different sources — such as outputs from forecasting models, direct use of satellite or radar measurements, interpolation of station data, and so on — all of which deliver products whose quantitative output is generally lower, if not downright poor.

Mitigation action: a clear technical description of the product makes it possible to trace the methodology used to build the dataset and the values of the key metrics used to assess the mean uncertainties associated with the meteorological variables. In fact, pseudo reanalysis techniques, especially for some parameters (e.g. rain), provide products of rather poor quality, not comparable with reanalytical processing. Even so, there are general areas of application for which low-cost, but also high-uncertainty, information is sufficient. As a consequence, these types cannot be used in contractual contexts, however, if they are, their objective limitations should be made explicit.

3. DEFINITION OF CONVENTIONALITY IN DIFFERENT APPLICATION CONTEXTS

Conventionality is a fundamental aspect as it must identify an acceptable synthesis of the representativeness of the meteorological data as such, even with respect to the technologies used to collect it. Added to this is the fact that a) meteorology does not technically lend itself, beyond a certain limit, to providing comparable quantitative data, b) the phenomena are in continuous change, c) different meteorological variables can be measured with different levels of precision, d) measurements draw on the specific point remotely, and finally

e) while the index-based approach is quite onerous, it does not attribute great weight to the data.

However, all of the factors mentioned enjoy generally accepted tolerances, except for the specific situation where the value is decisive with respect to a trigger. It is therefore necessary to precisely define the conventional aspects and actions that allow for appropriate use for operational requirements.

3.1. Definition of nationally representative resolution

The question of space-time dataset resolution is the result of a complex process synthesizing not only the specific functional requirements and the supporting technologies but also the pragmatic and heuristic component that can make it reasonably shared on the basis of a substantial wealth of experience derived from widespread use. It must, therefore, be consistent with at least the following aspects:

- availability of observational data and their spatial coverage;
- space-time scales for the meteorological phenomena to be represented with adequate accuracy;
- spatial properties;
- application context;
- cost of computational resources.

In general terms, in Italy, the historical depth of an ordered collection of weather data and the current observation set allow the data — both historical and near-real time and forecast — to be discretized over a 1 km^2 pixel.
The widespread application experience gained in recent years confirms the correctness and agreement on this choice. More detailed definitions can be explored in scientific and experimental settings, however, they might be prove difficult to apply and would, in any case, be limited by the presence of very dense local networks and thus not generalisable.
This topic, too, is discussed in detail in Quaderno di meteorologia aperta No. 1 “Characteristics and representativeness of precision meteorology in the Italian national context” published by Radarmeteo in 2020.

3.2. Buffers

In the vicinity of triggers, meteorological data cannot take on an absolute discriminating value, since the space-time dimension of the event, which is adequately captured as a whole, cannot be superimposed in static terms on a precise territorial scale, weather phenomena being constantly evolving. Here again, reference must be made to the specific need for which a given dataset is prepared; for example, while a wind speed of 13 m/s or 14 m/s does not significantly alter the output of a decision support system (e.g. wind productivity model, DSS in agriculture, etc.), instead, a 14 m/s trigger would, for example, change insurance options

for which this latter value could lead to damage compensation, but this would not be the case for the lower wind speed values.
Below are some methods for handling uncertainties based also on the nature of the uncertainties themselves.

3.2.1. Trigger spatial discontinuity

This aspect may be due to different causes. The first originates in the fact that the processed meteorological data is discretized into a regular grid consisting of cells with sides measuring 1km. Each of these cells is assigned the most representative value of the meteorological variable for the underlying geographic area. The breakdown of cells is made according to the DEM (Digital Elevation Model), and whether or not a cell reaches a trigger value may depend on the technical characteristics of the DEM used.
Above all, however, it should be considered that the spatial boundaries of meteorological phenomena are extremely unstable, often discontinuous, and that the precision with which they are monitored depends on the instrument precision and sampling density; thus it follows that the phenomena cannot be confined following clear-cut lines of separation.

Mitigation action: it is appropriate to conventionally define a transition area — denoted as 1 km — to act as a buffer zone between the area where the phenomenon is present and where it is absent or where different intensities occur.
The spatial buffer acts in two ways:

- in the case of trigger evaluation (boolean value TRUE/FALSE), the cells that, according to the original dataset, do not reach the condition required for definition of the adversity, but which are nonetheless adjacent to cells where the trigger value has been exceeded, are conventionally included in the area affected by the adversity;
- when querying the numeric variable data associated with a specific cell, to compare it with the trigger, a value is assigned representing the maximum intensity of the analysed adversity found in the area defined between the cell itself and the adjacent cells.

The previous two points define the typical modus operandi in application of the spatial buffer. However, there are some particular, specific cases for which the rules for applying the buffer area have been adapted to the characteristics of the meteorological variable being considered. More specifically these are:

- wind gust: for the “strong wind” adversity whose reference variable is the maximum daily wind gust speed, the buffer area is not 1 km but 3 km. This is because the spatial sampling density for the variable is lower than that for the other variables; in fact, wind measurement relies only on anemometer sensors, with more sparse spatial distribution than the thermometric and pluviometric networks. The reduced capillarity of wind measurements is also emphasized by the extreme spatial variability of the gust phenomenon, for which intensity peaks cover very small spatial scales, further motivating adoption of a larger spatial buffer;

- temperature: no spatial buffer is applied for the variable “temperature” because, in complex areas where elevation varies rapidly from one point to another within the territory — such as in a narrow valley floors — application of the buffer could assign temperature values which are unrepresentative;
- precipitation: the variable “precipitation” is assigned a spatial buffer of 1 km only for evaluation of triggers related to events that run their course in a very short time (1-3 hours), i.e., to verify adversities related to such phenomena as cloudbursts. This buffer is not applied to precipitation values on which other operations are performed (e.g. summations on daily precipitation), as this would lead to an overestimation.

3.2.2. Data uncertainty

For the meteorological variable processed and provided as part of the service, each value is associated with an overall uncertainty which arises from two main components:

- uncertainty relating to instrumental measurement;
- processing-related uncertainty (largely dependent on the density of the measurement network).

It can be estimated as an average value using specific dataset validation techniques; however, such definition is managed in the back office and does not achieve external visibility.

Uncertainty, therefore, is a quantity that varies according to the physical variable under consideration since it not only depends on the instrumental uncertainty but also on measurement point density: the higher the observation density, the lower the uncertainty inherent to the value of the final processed data. Since the observation density can vary between areas within the territory of interest, the uncertainty varies accordingly; however, to make management of this information easier, and in situations where observation density is rather uniform over the territory, the uncertainty can be considered spatially constant.

Mitigation action: conventionally, depending on the specific application domain, management of data uncertainty can lead to the adoption of values other than the best estimate (i.e., that provided by the dataset as the “true value”). In agricultural risk management, the average uncertainty can be added to the figure to define a value above which there is reasonable certainty that the event has not occurred or, in the energy field, provide the worst-case scenario to prepare all systems to manage it with an adequate margin of safety.
Thus, this buffer is dependent on the margin of uncertainty for estimation of the variable. For example, in the assessment of frost events (adversity defined for areas where the minimum temperature fell below 0°C), given that the average uncertainty on the estimated temperature value is 1°C, a buffer equal to 1°C is conventionally added (in effect, therefore,

a trigger for frost is considered to be +1°C or, as an equivalent operation, 1°C is subtracted from the temperature value obtained from the original dataset). Thus, in this case, all grid cells with values between 0°C and +1°C are conventionally included in the adversity area and thus could, within the margins of uncertainty, have been affected by the adversity.

3.2.3. Metadata-related uncertainty

The metadata used in processing are also affected by uncertainty due, in particular, to resolution of the grid adopted and the ability to correctly categorize a portion of land through remote sensing systems (generally, land-cover surveys are carried out with satellite surveys, which, like all measurements, have a certain margin of uncertainty).
Consequently, categorization of the land (land-cover) at a 1km resolution could lead to the classification of its portions into a specific class, even though mixed situations actually exist. For example, a cell classified as urban might actually be 70% urban and 30% agricultural. Cell classification affects the value of the associated meteorological variable (e.g. higher temperature in the urban area). Any farms found in cells not categorized as agricultural could be included in trigger-free zones even though the adversity had actually occurred in neighbouring agricultural areas.

Mitigation action: uncertainty about metadata, such as land categorisation, can be mitigated in the case of trigger assessments, by using spatial buffers. In fact, the problem of incorrect land categorisation is encountered in transition zones (e.g. from urban area to agricultural area), therefore a spatial buffer of 1 km can include portions of the territory with correct categorisation and thus representative data. In the case of querying the numerical data, the logic that can be adopted is to query the grid cell with correct categorisation lying closest to the location or area of interest.
The same uncertainty mitigation methodology can also be applied to other types of metadata, such as altitude.

4. DIVERSITY OF VALUES ON A GIVEN GEOGRAPHIC UNIT

As a consequence of an important assumption of operational meteorology — that each dataset is functional and representative only for the specific use for which it is intended, i.e., for a certain area and for a certain scale and time depth — it may occur that, for a given area of interest and a given event, different processing produces divergent meteorological parameter values. That divergence will be reasonably small, but if it serves as a trigger quantifying an event, the difference takes on a key role on any insurance contractual obligation, especially in a market that is increasingly moving toward solutions of a parametric nature (index-based policies). But that’s not all. Such events undermine the credibility of the “objective” component of the system, weakening the role of the oracle, i.e., those providing the weather data used. For several reasons, a particular situation has arisen in Italy. More specifically, this is the fact that the vast majority of the agricultural insurance portfolio takes as reference the same weather service provider, and this holds for both the insurance companies and the insured farmers and their associations. It should be specified, however, that that provider does not have its own monitoring systems, instead it collects and processes data made available in open formats by the public and private enterprises that manage the main meteorological networks (National Department of Civil Protection, regional ARPAs, Air Force, Land Reclamation Authorities, Utility Companies, non-profit associations, etc.), parties constituting the so-called observational set. Moreover, from these it derives the datasets required to meet market needs, according to tested, public methods and technologies. This specific situation brings undoubted advantages to the overall system, since it has led to a substantial zeroing of weather data-related litigation. However, it requires absolute clarity and transparency, which are built extra-contractually with operational consistency and long-term credibility, factors which can be undermined by misalignments, even though technically justifiable.

At the original 1 km resolution, the reanalysis dataset serves as the unambiguous source of all historical data; therefore, its direct use would not give rise to ambiguity because, given a coordinate pair (latitude and longitude) and a time instant, the value of the meteorological variable of interest most representative of the 1 km cell in which the specific area of interest falls is uniquely defined.

Differences in the values for a given geographic area and a given instant in time may be found using different versions of the same reanalysis dataset or by comparing data present in datasets of different natures, as covered in previous chapters.

Given the same space-time conditions, another context that can give rise to differences in values is that of spatial aggregations based on the original reanalysis dataset, i.e. when "second-level datasets" are produced.

The purpose of this chapter is to highlight how different methods of spatial aggregation can lead to meteorological variables being valued differently even with the same geographic and temporal units, and what mitigating actions are needed to avoid possible ambiguities in the use of those datasets.

4.1. Second-level datasets derived from spatial aggregation

Often, the use of the original dataset is not consistent with the scope, and therefore appropriate spatial aggregations need to be made to:

- reduce the number of data (e.g. regridding to apply grids with lower resolution than the original one);
- standardise the weather data reference system to the particular field of application, which often operates by spatial units (e.g. municipality, province, zip code, watersheds, etc.).

The following sections provide a description of the different spatial aggregation processes and the related mitigation actions that must be deployed to handle any differences in variable valuations introduced by these procedures.

4.1.1. Regridding and the HYPER-GRID ID® system

Although preferable for achieving maximum data representativeness, use of the original 1 km resolution reanalysis dataset may be inefficient for some activities that may not require particularly high spatial detail. In fact, for some applications, it is preferable to lighten the processing chain by favouring computational efficiency because use of lower resolution grids — with smaller cell counts — guarantees much shorter computation times.
The operation of transforming a high-resolution grid into a low-resolution one is called regridding. This operation may be performed using several methods, all of which share the fact that they take the value of a representative weather variable from the original cells and associate it with the target cell, given that the original cells are smaller, therefore have higher resolution and fall into the target grid. (Fig. 1).

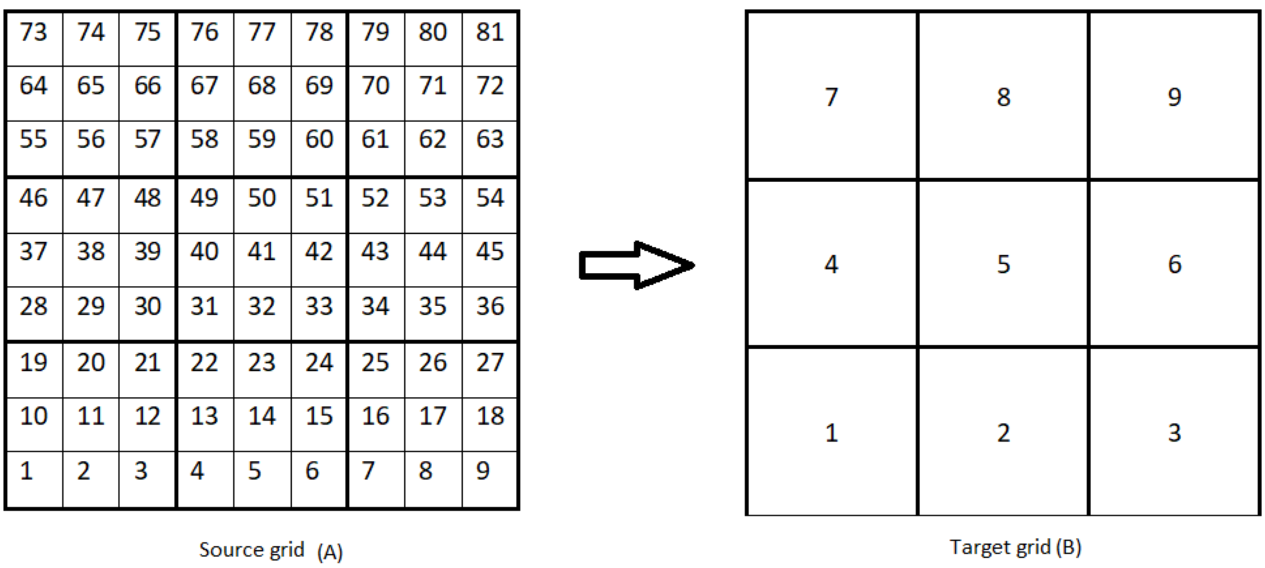

Fig. 1 Regridding

In the example shown in Fig. 1, where the spatial resolution of the original grid A is 3 times greater than that of the final grid B, the number of cells is reduced by a factor of 9; representing that same geographic area with the high-resolution grid requires 81 cells, while with the low-resolution grid, 9 cells are sufficient. The reduction factor for the number of cells — in this case 9 — affects the time required to process the contained data, therefore the processing of grid B is 9 times faster than that of grid A. Assuming a processing chain that takes one hour to process grid B, it would take as much as nine hours to process the data at the original resolution; this means that 9 times more computing power would be required to keep the processing time unchanged, with a proportional increase in computing infrastructure costs. However, the use of grids with different resolutions may result in a problem of ambiguity in the value of the weather variable for the same geographic location. In principle, regridding the field of a meteorological variable — from a high resolution grid to a lower resolution grid — results in averaging the original values present in the cells that are now represented by a single cell.
In the example in Figure 2, the precipitation field represented by 9 cells at 1 km resolution, covering a certain geographic area, may exhibit very high spatial variability in case of convective precipitation. If a regridding operation is performed on a final 3 km resolution grid, the values of the original 9 cells are essentially averaged to reconstruct the most representative data for the final cell. If one were interested in the precipitation value of a specific location, it would differ depending on which grid is used: querying the high-resolution grid, the location assumes the original value of the 1 km per side cell into which it falls, i.e., 19 mm; instead, querying the low-resolution grid, the location assumes the average value of the original 9 cells used to reconstruct the datum for the final grid cell, i.e., 6 mm (Fig. 2).

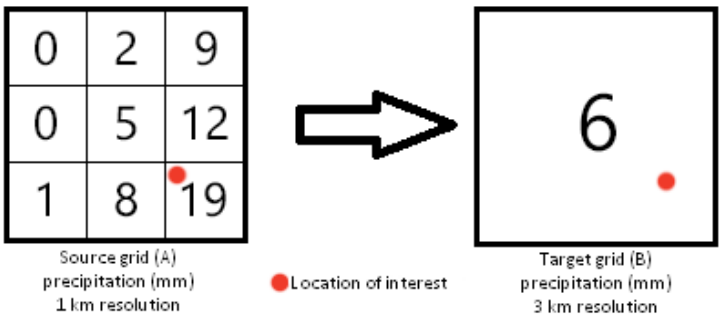

Fig. 2 Regridding of precipitation data from a high resolution grid to a low resolution grid. In the original grid, the precipitation value associated with the location of interest is 19 mm; in the final grid it is 6 mm.

Mitigation action: in datasets used in decision-making processes, proper metadata is necessary to uniquely define the grid to which they belong and, therefore, the characteristics of the data used. In order to provide each cell in a grid with an identification code so that it can be uniquely linked to the dataset to which it belongs and its geolocation, Hypermeteo adopts the HYPER-GRID ID® system, which combines the cell's identification code not only with the weather data but also with a whole series of indicators and other information useful for analysing the specific features of the area of interest enclosed in the cell. By defining a standardized and shared logic, HYPER-GRID ID represents the key that enables the unique

link between the weather-climate data and the data of the specific application domain (e.g. contracts, company files, policy certificates, industrial assets, etc.).
For example, in the field of risk management, each item insured may be automatically matched to an identification code (ID) based on its territorial location and thus can be associated with the data flow deemed most representative. All at-risk items whose geographical coordinates fall within a given HYPER-GRID ID® are characterised by the same level of risk and the same weather and climate parameter values used by the application or service. For example, in index-based policies or decision support services (DSS), risk indices and adversity measurement values on which the application or service is based, can be made available and constantly updated for each ID.
The HYPER-GRID ID® logic makes it possible to associate several virtual weather stations belonging to different datasets (i.e. data type and geographical resolution) with each object, thus ensuring total transparency and unique information regarding the source of all data streams (Fig. 3 and 4).
Knowledge of the identifier of the cell where the area of interest lies, used as an input key in digital systems for querying datasets made for the purpose (web apps, management applications, etc.) enables the user to rapidly, constantly monitor the changing weather parameters associated with the identifier itself, thus facilitating bilateral transparency of the data within a contractual relationship (e.g. between the insurance company and the insured).
Moreover, in the context of a contractual relationship, whether or not involving insurance, in which the value of the weather parameter is determining for a a contractual obligation to arise, the HYPER-GRID ID® system makes it easy to manage cases where the area of interest combined with the asset/service covered by the contract falls within two or more cells of the grid under consideration identified by different codes. In such cases, the parties could choose the cell identifier to be contractually referenced on an interactive map, in advance and in a conventional manner.

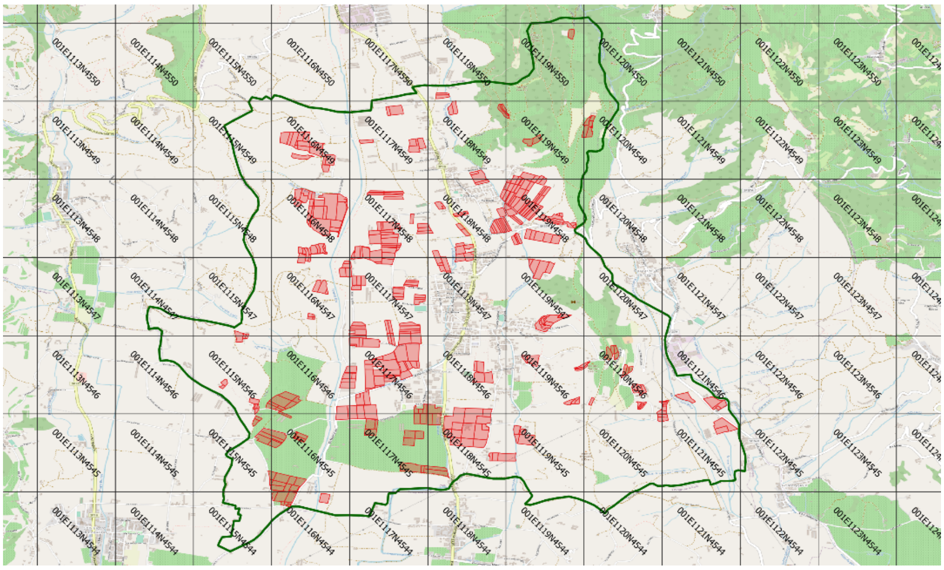

Fig. 3 Representation of the HYPER-GRID ID® system over a municipal area: each cell is identified by a unique code that makes it possible to trace the characteristics of the source dataset.

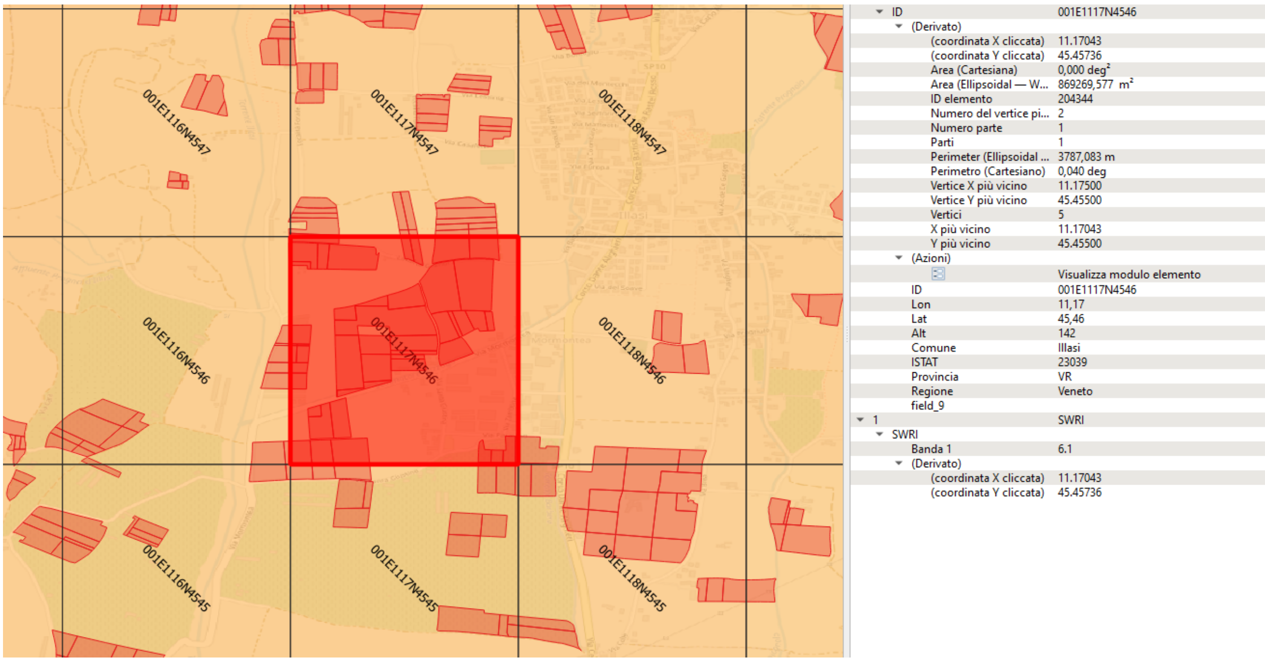

Fig. 4 Example of metadata provided by the HYPER-GRID ID® system (e.g. coordinates, altitude, municipality, ISTAT code, province, etc.).

4.1.2. Aggregations by spatial units and unique datasets

The "natural" geographic reference system for meteorological data — i.e., grids whose cell positioning is defined by coordinate pairs (latitude and longitude) — often proves difficult to interpret in the context of risk management and, more generally, in cases where the systems used to locate assets are those of an administrative nature (e.g. street addresses, cadastral references, municipalities, provinces, regions, etc.).
The set of these spatial units, which define the georeferencing system for the field of application, can be divided into two categories:

- point spatial units: those that have a spatial dimension which is smaller than the grid cell; in this case the location of interest/cell association is unique (e.g. each street address is associated with one and only one grid cell because the size of the building identified by the address is often several orders of magnitude smaller even than the cell size of higher resolution grids);
- area spatial units: those that have a spatial dimension that is larger than the grid cell; in this case the location of interest/cell association is no longer unique since there is a plurality of cells within the area underlying the territorial unit of interest.

Therefore definition of a point or area georeferencing system is strictly dependent on the spatial resolution of the data grid to be used. For example, a municipal area can be seen as an area spatial unit if a grid with a 1 km resolution is being used (thus, numerically, the

number of cells falling within the municipal area is equal to the square kilometre extension of the municipality itself) while
it can be considered as a point spatial unit if the dataset resolution is very coarse (e.g. 50 km) in which case the municipality falls within a single cell and uniquely takes on its value.
Translating from the territorial unit reference system to the grid reference system can sometimes prove difficult, even when dealing with "point" references (e.g. street addresses and cadastral parcels), for which one only has to identify the grid cell in which the point of interest falls. In fact, procedures are required to "translate" the geographical information for the application into the grid reference system (latitude and longitude). For example, for street addresses there are numerous web services that enable one to extract the coordinate pair for each, while for cadastral data, vector databases that geolocate parcels are only now becoming available.
When dealing with area spatial units, in order to associate the weather-climate information "characteristic" of that particular location, post-processing of the data on a grid must be introduced. Since there is a plurality of cells underlying the area of interest, the procedure adopted involves aggregation of the data according to specific rules. Even with the same spatial unit name (e.g. municipality name), these different aggregation logics may result in different valuations of the meteorological variables.
Spatial aggregation (e.g. by municipality, province, basin, etc.) may differ for several reasons:

- Different aggregation techniques: the logic used to obtain a single representative figure for a certain spatial unit, may differ depending on the purposes. The most commonly used techniques are, for example, arithmetic mean, weighted mean, extraction of a particular percentile (e.g. median, first quartile, third quartile, etc.), extraction of the minimum or maximum value.
- Filters: to obtain a more representative figure for the field of application, it may be necessary to exclude from the aggregation those grid cells that, while falling within the spatial unit in question, are not of specific interest and thus their weather-climate values could distort the final value. A typical example of a filter is that regarding altitude: using a DEM (Digital Elevation Model), it is possible to associate each cell in a grid with an altitude value and, depending on the field of application, exclude those cells belonging to unrepresentative altitude intervals. This may be the case for a municipality whose territory is distributed over areas including valley floor, or even coastal areas, and major elevations.
- Buffer: sometimes it may be useful to enlarge the area underlying the spatial unit to include grid cells that would, in fact, lie outside the boundaries. Including these cells in the aggregation process contributes to a different valorisation of the meteorological variables associated with the spatial unit when compared with simple aggregation of the cells within the territory.
- Updates of territorial units: periodically changes in the extensions and boundaries of some territorial units can occur, such as in the aggregation between municipalities. This leads to changes in the set of cells underlying the spatial unit and to lack of temporal homogeneity in the historical series, which must then be recalculated.

To make it clearer how all these cases can lead to different valuations and give rise to ambiguity, the example of a municipal temperature dataset is proposed. If the dataset is intended for use in agriculture, it may prove necessary to exclude all cells that are at altitudes above the elevation-limit for crop growth, thus ensuring that the aggregated data (e.g. average temperatures) is not affected by the values of cells that are not of interest for the specific case. In addition, again with a view to obtaining more representative data for the agricultural land, a filter based on land cover classifications can be applied to exclude from the aggregation all grid cells representing portions of land not used in agricultural production (e.g. urban areas, forests, inland waters, uncultivated areas).
The following is a comparison of the results obtained from a municipality-level aggregation without filters and from an aggregation applying the agricultural-type filters mentioned above (agricultural type land cover, elevation < 1200m) for a municipal area where the environments (urban and mountain areas) vary significantly.

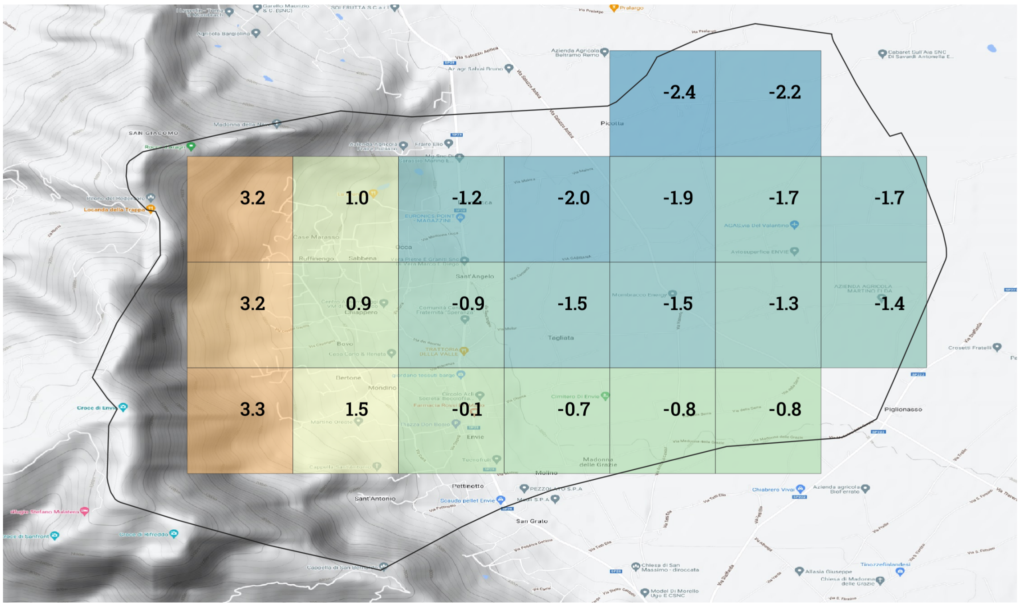

Fig. 5 Minimum daily temperature values for all grid cells falling within the municipality of Envie on 23 March 2022.

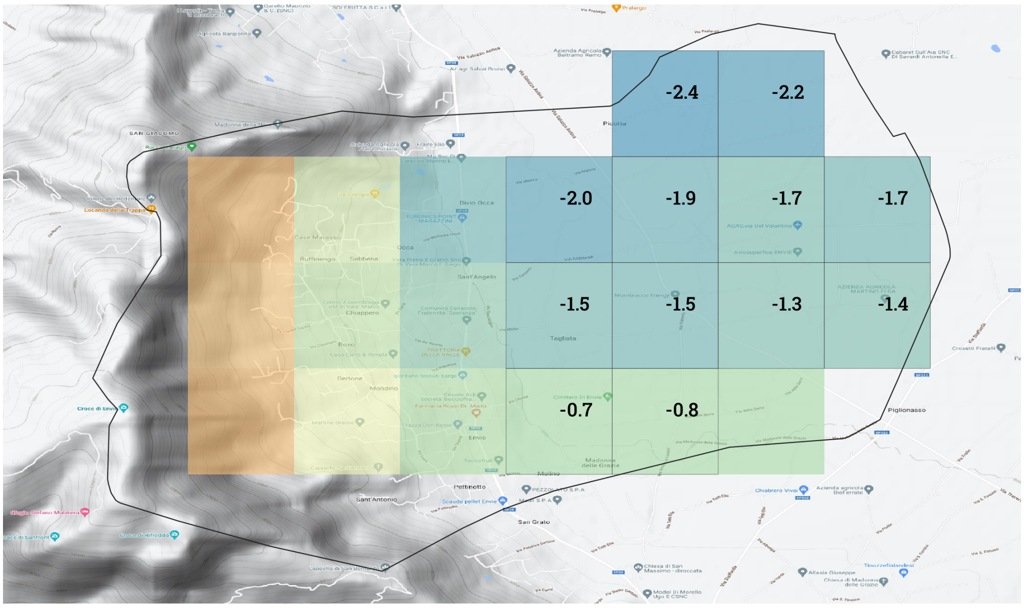

Fig. 6 Minimum daily temperature values related to grid cells for agricultural use below 1200m falling within the municipality of Envie dated 23 March 2022.

Municipality	Minimum value	Average value	Maximum value
Envie (all cells)	-2.4 °C	-0.4 °C	+3.3 °C
Envie (agricultural cells)	-2.4 °C	-1.6 °C	-0.7 °C

Tab. 1 Aggregate dataset values for minimum daily temperature for the municipality of Envie on 23 March 2022 depending on the filter applied: first row, no filter, second row, agricultural land cover filter and altitude <1200m.

The example highlights how different aggregation logic can lead to completely different aggregate values for a given spatial unit. In the specific case, the presence of thermal inversion results in higher temperatures on the grid cells for the mountain area which, when no filters are applied (first row of Table 1), shows a significant increase in mean (+1.2°C) and maximum (+4.0°C) values as compared to the aggregate dataset using filters for elevation and land cover.

Mitigation action: Adoption of a unique dataset for each application area, with the aim of:

- maximizing meteorological data representativeness for the specific context;
- providing unique values so that there is no room for contractual ambiguity;

these objectives are achieved by clearly defining the:

- data aggregation logic, so as to enable reconstruction of the procedure that made it possible to obtain those particular variable values from the original dataset applied on the grid.
- Spatial units used; when updates are made to the geographical base (e.g. merging of municipalities), the historic series related to those territorial units should also be updated to homogenize the data from a time perspective.

Below are some unique datasets used operationally for specific usage areas and the main aggregation logic applied to make the final data more representative.

DATASET	TYPICAL SPATIAL UNITS	AGGREGATION TECHNIQUES	FILTERS	BUFFER
AGRICULTURAL *DATASET* – AGRICULTURAL INSURANCE	Municipalities and provinces	Arithmetic mean, extraction of spatial max. and min. values	Agricultural type land cover, elevation below 1200m	None to 3 km, depending on variable considered
PROPERTY *DATASET* - PROPERTY INSURANCE	Municipalities and provinces	Arithmetic mean, extraction of spatial maximums and minimums	Urban type land cover	None
ENERGY CONSUMPTION *DATASET*	Municipalities, provinces, regions	Population density-weighted average of weather-climate variables	None	None
WIND *DATASET*	Municipalities and provinces	Wind farm density-weighted average of weather-climate variables	None	None
PHOTOVOLTAIC *DATASET*	Municipalities and provinces	PV plant density-weighted average of weather-climate variables	None	None
HYDROELECTRIC POWER *DATASET*	Watersheds	Arithmetic mean, extraction of spatial maximums and minimums	None	None

Tab. 2 Examples of unique datasets developed for different fields of application.

5. BIBLIOGRAPHY

[1] https://www.radarmeteo.com/wp-content/uploads/2022/02/Quaderno_Meteorologia_Aperta1_Rappresentativita%CC%80_dati_meteo-1.pdf

[2] https://www.radarmeteo.com/wp-content/uploads/2022/07/Quaderno_Meteorologia_Aperta2_Rianalisi-2.pdf

[3] https://www.radarmeteo.com/wp-content/uploads/2022/07/Quaderno_Meteorologia_Aperta-4_Lemergenza-climatica-al-tempo-del-digitale.pdf

[4] WMO. (1993). Siting and Exposure of Meteorological Instruments (J. Ehinger). Instruments and Observing Methods Report No. 55 (WMO/TD-No. 589). Geneva.

[5] WMO. (2000). *Representativeness, Data Gaps and Uncertainties in Climate Observations* WMO-TD No. 977. Geneva.

[6] WMO. (2008). *Guide to Hydrological Practices* (WMO-No. 168), Volume I. Geneva.

[7] WMO. (2011). *Guide to Climatological Practices* (WMO-No. 100,). Geneva.

[8] 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.

[9] Ramsey, C.A., Hewitt, A.D. (2005). *A methodology for assessing sample representativeness*. Environmental Forensics 6(1) 71-75.

[10] Sinclair, S., Pegram, G. (2005). *Combining radar and rain gauge rainfall estimates using conditional merging*, Atmospheric Science Letters, 6, 19-22.

[11] Thorne, P., and Vose, R. S. (2010). *Reanalyses suitable for characterizing long-term trends*. Bull. Amer. Meteor. Soc., 91, 353–361.

[12] Parker, W.S. (2016). *Reanalyses and Observations: What's the Difference?* Bulletin of the American Meteorological Society 97 (9): 1565-1572.

[13] 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.

[14] Conradt, S., R. Finger, and M. Spörri, (2015). *Flexible weather index-based insurance design. Climate Risk Manage*. 10, 106–117.

[15] Daniel K., Riishojgaard L. P., Eyre J., Varley R. A., *The Value of Surface-based Meteorological Observation Data*, The World Bank, World Meteorological Organization, British Crown, Met Office, 2021.

THE AUTHORS

Former Inspector of the State Forestry Corps-Ministry of Agriculture and Forestry, 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, Director of the Environmental and Hydrological Monitoring Centre of the European Union in Asunción (Paraguay) and Head of Latin America mission for meteorological projects. Author of more than 30 technical-scientific publications. Lecturer at EU, FAO, IILA courses for Latin American government officials; lecturer at university, postgraduate and master courses, as well as for corporates and unions.
Currently Chairman of Hypermeteo and CEO of Radarmeteo.

Massimo Crespi

Expert in meteorology applied to risk assessment and head of agricultural and property risk management at Radarmeteo Srl. Co-founder and head of data analysis at Hypermeteo Srl.
Author of technical-scientific publications on retrospective analysis (reanalysis) weather-climatological technologies and their application.

Gianluca Ferrari

Involved in the risk management field for over 30 years, he has worked for major Italian and foreign insurance companies and is an expert in the analysis and management of agricultural risks. Since 2015, he has been a consultant to Asnacodi Italia handling relations with the insurance world, management of Mutual Funds, and digitization and process innovation. He has participated in European projects for the use of satellite data (PREMIA) in the study of new index-based policies and for the integration of weather-climate datasets into the management of atmospheric risks in agriculture. He is also currently Senior Advisor to ISMEA for the AgriCat Mutual Fund trial project.

Paolo Spagna

Radarmeteo
Professional weather services

PMI innovativa
Via IV Novembre, 117
35020 – Due Carrare (PD)
radarmeteo.com
info@radarmeteo.com

Hypermeteo
Climate & weather grids

Startup innovativa
Via Nazionale, 181
00184 – Roma (RM)
hypermeteo.com
info@hypermeteo.com