#### SPHERA REANALYSIS CONFIGURATION:

In table 1 are reported the details of the reanalysis dataset SPHERA configuration. The setup differs from that of COSMO-2I operative model for the drivel model (i.e. ERA5) and the relative nesting modality, for the initial and boundary conditions, and for the data assimilation strategy employed over observations and the assimilated dataset itself. On the other hand, the resolution, domain, physical and dynamical schemes and the physiographic parameters are coherent with the operative model.

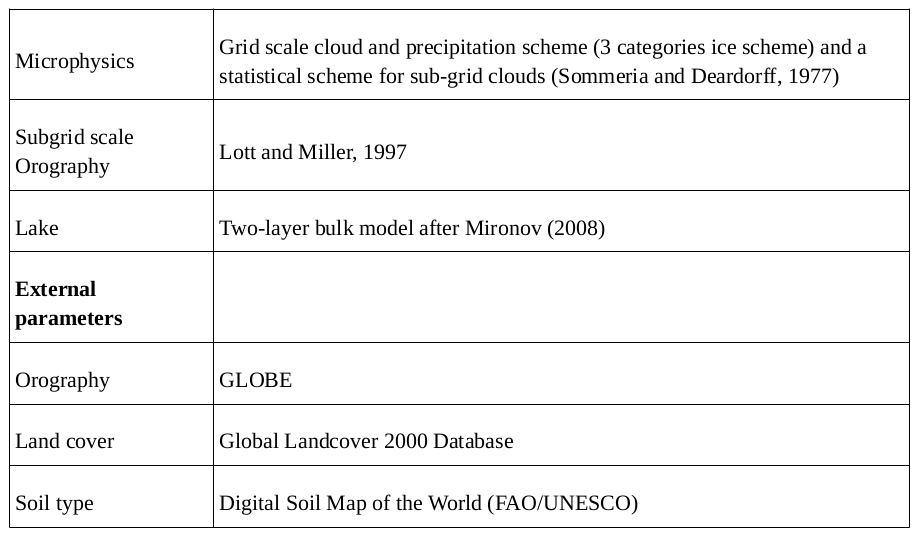
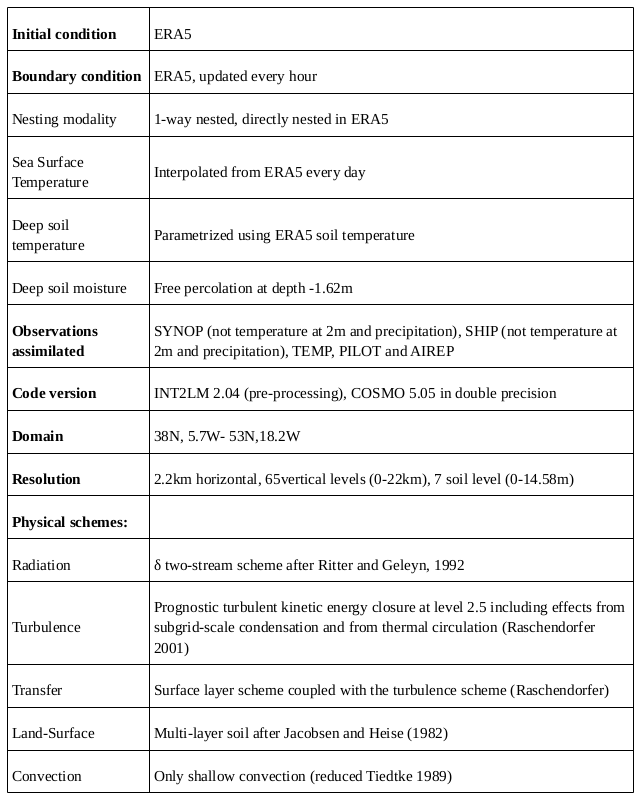


Table 1: SPHERA configuration

The reanalysis archive comprises:

* the 3-dimensional fields of temperature, wind components, specific humidity for different microphysical species, cloud coverage, pressure, turbulent kinetic energy and the Richardson number over COSMO vertical levels (65 vertical levels) at hourly frequency,
* the geopotential fields, wind components, temperature and specific humidity at some pressure levels (i.e. 300hPa, 500hPa, 700hPa, 850hPa, 925hPa, 950hPa, 1000 hPa) at hourly frequency,
* the wind component fields at sub-hourly frequency between 0 and 500 m above sea level,
* the 2-dimensional fields at the surface as fluxes, radiations, precipitation, runoff, diagnostic variables near-surface (temperature and humidity at 2m, wind components at 10m) at hourly frequency,
* the temperature and humidity fields in the soil at 7 depth levels at hourly frequency,
* the thunderstorm indices and other upper-air products (e.g. planetary boundary layer height, base and top cumulonimbus level heights, freezing level height, snow height,…) at hourly frequency

**EXTRACTED VARIABLES FROM SPHERA:**

The available data extracted from SPHERA archive in .grib format (ecmwf) are located at imk-tss-risk under /hail1/climate-model-data/reanalysis/SPHERA, and are relative for a period of 3 years (2016-2018) for April to October. They include:

|  |  |
| --- | --- |
| **VARIABLE NAME** | **DESCRIPTION** |
| 2d | Dewpoint temperature at 2m (K) |
| 2t | Temperature at 2m (K) |
| 500hPa/  z  t  r  q  u  v | @500hPa Geopotential (m2/s2)  Temperature (K)  Relative humidity (%)  Specific humidity (kg/kg)  u-component of wind (m/s)  v-component of wind (m/s) |
| 700hPa/ (t,q,u,v,  wz  w | @700hPa  Vertical velocity (geometric)  Vertical velocity (pressure) |
| 850hPa/ (t,r,q) | (see above) |
| 950hPa/ (u,v) | (see above) |
| BAS\_TOP\_CON  BAS\_CON  TOP\_CON | Cloud base index (vertical level)  Cloud top index (vertical level) |
| CAPE/  CAPE\_CON  CAPE\_MU  CAPE\_ML | Convective availab pot energy (J/kg)  Most unstable CAPE (start level of most unst parcel)  Mixed layer CAPE (start level of mixed layer parcel) |
| HZEROCL | Level of 0°C isotherm (geometric height (m)) above mean sea level |
| SLI | Surface lifted index (K) |
| TMAX\_2M/ | Daily maximum temperature at 2m (K) |
| TMIN\_2M/ | Daily minimum temperature at 2m (K) |
| Hum\_hPa/ (r,q) | Relative and specific humidity at various pressure levels (250,500,600,700,850,925,950,1000 hPa) |
| 10uv | Horizontal wind speed components @10m |

Furthermore Daniele D’Alessandro (ARPAE) developed a machine learning algorithm to detect the COSMO-based convective proxies most suitable for the description of severe thunderstorms (not specifically for hail). The four extracted proxies are calculated from SPHERA dataset, not at its original resolution but averaged over boxes with horizontal grid spacing of 1/5° and 1/10° (i.e. approximately 20 and 10 km respectively) for the period of three years (2016-2018), and available at imk-tss-risk under /hail4/giordano/hail\_obs/DAlessandro\_indices\_gridded, and include:

|  |  |
| --- | --- |
| **VARIABLE NAME** | **DESCRIZIONE** |
| %VV700 | Percentage of grid points (from the original reanalysis resolution) in the area where vertical velocity @700hPa in isobaric coordinates (omega, Pa/s) is lower than -0.5 Pa/s (in isobaric coordinates negative values indicate ascending motions!) (typical for significant vertical ascending motions (reference?)). It quantifies vertical velocity strength (if model represents convection in right location high values of %VV700 should be expected).  This parameter does not make sense if calculated over the original high-res grid (as it is calculated from the fraction of grid points in a larger area than a single grid box exceeding a certain value) |
| AvvGeoPot500 | Gopotential advection @500hPa averaged over the box in 12 hours: absolute horizontal geopotential advection (U\*dZ/dx + V\*dZ/dy) \* 12 hours (not sure if this index represents what it was originally designed for(?), but from the analysis of D’Alessandro it was the index that worked better!(not so sure)). Used to quantify both temperature variations (i.e. average virtual temperature of the layer representing the average troposphere) as well as the passage of depressionary waves: if the geopotential height decreases in the 12 hours before an event this might indicate potential atmospheric instability for convection occurrence. Typical values (not for hail but for thunderstorm): values drop around **5-10 dam/12h** |
| Kindex | K = **T(850 mb) + Td(850 mb) - T(500 mb) - DD(700 mb) [K]**  Measure of atmospheric instability (thunderstorm potential)based on the vertical temperature lapse rate, and the amount and vertical extent of low-level moisture in the atmosphere. Typical values **>30K** (not for hail but for severe thunderstorm), but varies depending on season |
| LI | Surface lifted index LI=**T(@500hpa environment) - T(@500hpa parcel)**  evaluate the instability overall the entire atmospheric column. Typical values  **< -5°C** (not for hail but for severe thunderstorms, depending on season and model) (prescribed as direct output from SPHERA). Anything below 0°C is relevant for convection, for strong events reach -10°C. Parameter related to the estabilishment of instability conditions BEFORE the convective event may take place -> must be considered in pre-convective conditions, 1-2 hours before the storm formation (during storm LI would tend to decrease as instability is already estabilished and temperature difference decreases) |
| CAPE (MU or ML) | Convective available potential energy, CAPE is proportional to maximum updraft speed -> more CAPE = larger hail. Direct output in two formulations:   * Most unstable: when parcel lifted from the profile yielding the largest CAPE (useful to consider elevated convection) * Mixed layer: when considering a parcel lifting from an average height on the layer near the surface (lowest 100hPa), usually smaller than MU (but can be larger if a shallow inversion is present)   (with COSMO 2.8km in assimilation mode gaps in CAPE (i.e. going to 0) have been detected when convective event has started due to the explicit representation of convection, so this must be considered with care also in my case having high-res model! -> look for CAPE gaps in event after convective initiation and development, before hailfall!). Typical values (not for hail but for severe thunderstorms): **>1000J/kg** in summer, **>100-200 J/Kg** in other seasons (but depending on model and cape formulation)  Investigate optimal threshold for CAPE for hailstorm in Italy! |
| H0 | Height of the 0°C isotherm **above mean sea level**. (When below the sea level is defined as -999 (substituted with “rmiss”), also when in the same averaged grid (10km) all H0<0m -> Val =0.340282347E+39  Investigate optimal threshold for H0 for hailstorm in Italy! |
| DLS | Deep layer shear calculated as the vector difference between the horizontal wind speed at 500hPa and 10m (parameter that in US is very much important for describing hail/severe convective events, while in Europe much less, probably due to the much more complex topography, and has shown no added skill when combined with LI in a study (of Michael), BUT still it needs to be considered when considering hail renalysis proxies. When DLS is larger, meaning larger wind gradient within the vertical of the thunderstorm, the updraft is likely to be more tilted and the recirculation for hail growth is more likely. Typical values: **>10-20m/s** (not for hail but for severe thunderstorms)  Investigate optimal threshold for DLS for hailstorm in Italy! |

All these proxies are available for the moment for the period 2016-2018 April-October on a 10km grid extended to 49°N.

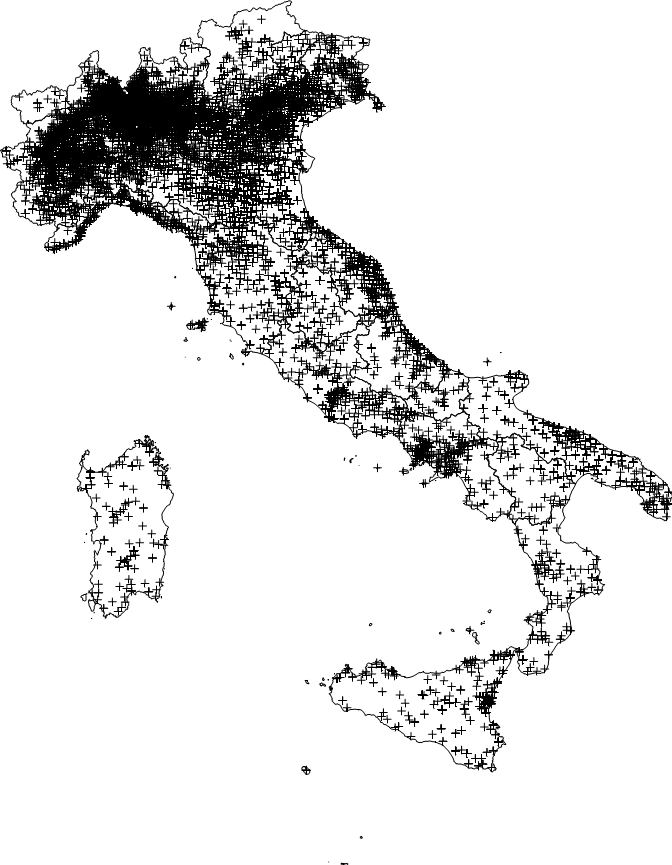
**POSSIBLE PROBLEM:** SPHERA indices are NOT VALIDATED!!! Should this be a problem??

15/3/22 + 5/4/22

Prepared SPHERA data of CAPE, HZEROCL, uv@10m e uv@500hPa (2016-2018, April-October) for gridding to 10km, must calculate DLS and then grid all hourly data to obtain csv tables (asked Davide for help in calculating DLS). (DONE)  
  
**OBSERVATIONAL DATASETS**

The available observational datasets include:

* **UNIPOL** insurance company claims for request of refund for damages to vehicles and buildings related to hail occurrence, covering Italy (see map), with spatial resolution related to the centroids of the cities (not exact location), covering 2017-2020, with non-uniform temporal resolution (reported with UTC and also non-UTC format) having an estimated uncertainty of +- 3 hours. Detected higher density claims in northern Italy (Piedmont, Lombardy and Veneto regions), lower in other areas but strongly depending on the amount of people insured with UNIPOL or with other companies. A total of 120.000 observations is present in the dataset. No information on the entity of the damage (in €) or to the hail size reported.



* **ESSL** dataset: for 1995-2020 a number of 2468 hail reports with QC0+> (plausibility check passed)
* **LAMPINET** lightning dataset, covering Italy and extending from 2014-up to now, information on the total number of lightning strikes with 15 min temporal resolution (reference:https://link.springer.com/chapter/10.1007/978-1-4020-9079-0\_6).
* EUCLID lightning dataset for 2001-2020, covering central Europe, not entirety of Italy, detailed info divided between C-G and C-C
* German insurance company(ies) claims: refer to Katharina
* **OT** detection data from Kris Bedka (with new algorithm he will send us data for 2016-2018): data at 4km horiz. Resolution, extending from lat[30.02,64.98] lon[-9.98,29.98]

**LJ INDEX:**

All these three dataset where considered in a preliminar investigation with the aim to produce a hourly lightning jump index related to the presence of hail to validate lightning information against UNIPOL and ESSL reports. To do so a spatial aggregation over boxes having various horizontal resolution was operated and retained only for the grids of 1/5° and 1/10° (i.e. approiximately 20 and 10 km respectively).

Valentina performed some sensitivity analysis tests of the LJ index (must be re done for a paper in a more systematic way), starting from the raw punctual 15-min aggregated LAMPINET lightning data, vs UNIPOL insurance dataset, based on various parameters:

* Grid resolution (10,20,50,100 km -> retained 20km (too fine 10km, losing too many events, while 50-100 too coarse for the purpose)
* Number of grid points: 1 or 1+8 around it -> 9 grid points (significantly higher number of events catched)
* Number of minimum lightning strikes to define a LJ (0,1,5 -> retained 1)
* Number accumulated lightning strikes to define a LJ (5,10,20,30 -> retained 20)

Temporal aggregation performed over 1 hour (so 4 temporal steps of 15 mins aggregated together) and the LJ value is attributed to the last instant of the hour in question (i.e. a LJ issued at 17 means that it covers the minutes 17:00 to 17:59) **GIUSTO??? CHIEDERE BENE A VALENTINA PER ESSERE SICURI!!!!**

**I THINK IT IS ACTUALLY THE CONTRARY: LJ issued at 12 means from 11 to 12 (from the qualitative matching with OT occurrence it seems like that) ASK TO BE SURE!!!**nuova conferma del fatto dall’analisi dell’evento del 14/7/2017: eswd reports alle 19.30, matching dell’LJ con l’issue delle 20, quindi incluso tra le 19 e le 20

**E sono in formato UTC o UTC+1 ????**

Available the LAMPINET daily index used to produce hail risk maps for unipol (for now), for now for the grid of 20km for the period 2015-2020 in csv format.

Q to Valentina: temporal aggregation of LJ in one hour: does the 1-20 lightning strikes rate apply on 15 minutes intervals or to 1 hour? And if in 15 mins intervals, how data are aggregated in 1 hour data (like in 4 15 mins intervals we have 2 positive LJ and 2 negative, how data in this case reflect on hourly basis?)

**Must understand well if LJ developed with LAMPINET respect the definition accepted in the literature after Schultz et al. 2011, 2017, used in several other studies (e.g. Farnell and RIgo 2020, Nisi et al. 2020)!! MUst meet with Valentina!**