**Development of precipitation fields for the COSMO-PL 2.5 km domain: processing of multi-source precipitation data, quality control, and quality-based combination**

Jan Szturc, Katarzyna Ośródka, Anna Jurczyk, Magdalena Pasierb

Institute of Meteorology and Water Management – National Research Institute, PL 01-673 Warsaw, Poland

**1. Introduction**

The work described below was carried out in the Institute of Meteorology and Water Management – National Research Institute (IMGW) as part of the EPOCS PT “Evaluate Personal Weather Station and Opportunistic Sensor Data CrowdSourcing”. The two subtasks were focused on adaptation of the precipitation estimation for assimilation into the COSMO-PL 2.5 km numerical weather prediction models.

In a first step, a RainGRS system for combination of multi-source precipitation data was updated for the enlarged COSMO-PL 2.5 km model domain (named as RainGRS+). This modification of the domain resulted in the need for including additional data, which were not employed when working on the smaller domain.

The emphasis was placed on telemetric data from non-professional rain gauge data networks and on quality control that will allow their corrections and effective detection and elimination of erroneous data (by RainGaugeQC system). Moreover, preliminary analyses were carried out on the possibility of using rainfall estimates obtained from processing information on the attenuation of commercial microwave link signals (CMLs) in precipitation to better reproduce a spatial variability of the precipitation field.

**2. Processing different precipitation data sources based on RaingGRS system**

*2.1. Development of RainGRS system for multi-source precipitation estimation in the COSMO-PL 2.8-km domain version (RainGRS+)*

At IMGW multi-source precipitation field estimates are produced by the RainGRS system (Jurczyk et al., 2020a), which combines data from:

* telemetric rain gauges,
* weather radar network from Poland (POLRAD) and from neighbouring countries: GE, CZ, SK, LT, and DK,
* satellite data (Meteosat) processed by EUMETSAT NWC-SAF software (Jurczyk et al., 2020).

The algorithm for combining rainfall data from different sources is based on a conditional merging introduced by Sinclair and Pegram (2005) that attempts to enhance the strengths of the individual inputs and reduce the impact of their weaknesses. This algorithm was modified by taking into account the quality information of the individual input data, attributed to them when performing their quality control.

The RainGRS+ is a version of RainGRS, which is working on larger domain required by the COSMO-PL 2.8 km model which leads to including new input data from:

* networks of non-professional telemetric rain gauges (e.g., Bárdossy et al., 2021),
* EUMETNET OPERA radar precipitation operational data (Saltikoff et al., 2019) for the area outside of Poland where the radar data are not available directly.

The temporal resolution of the RainGRS system is 10 min while the spatial resolution is 1 km x 1 km. The domain sizes of standard RainGRS and RainGRS+ are 900 km x 800 km and 1200 km x 1300 km, respectively (Fig. 1).

Obraz zawierający tekst, mapa

Opis wygenerowany automatycznie

Fig. 1. Domains of the systems used: RainGRS (internal) and RainGRS+ (external).

*2.2. Development of a new precipitation input based on signal attenuation on commercial microwave links (CMLs)*

The aim of the work was to develop a basis for the operational use of data from opportunistic measurements based on commercial microwave links (CMLs). The network contains more than 1,000 links within the territory of Poland, and is particularly dense in large urban areas, for which radar precipitation observations are subject to high uncertainty due to ground clutters and interferences with RLAN signals.

Obraz zawierający tekst, mapa, pismo odręczne

Opis wygenerowany automatycznie

Fig. 2. Study area (Opole Voivodeship) marked with pink borders (a); locations of CMLs   
and manual rain gauges (b).

It was analysed to what extent the precipitation derived from CML attenuation data is useful in estimation of the precipitation field with the high temporal and spatial resolution. Various types of precipitation data from July 19 to August 18, 2022 have been used in this study for the Opole Voivodeship (Fig. 2). Table 1 summarises the statistics of CML-based precipitation using data from the manual rain gauges as a reference. Preliminary results showed that precipitation data from the CMLs have a quality lower than radar and rain gauge (professional) measurements, but higher than satellite ones. The results of this investigation are in the process of being published by Pasierb et al. (2024).

Table 1. Results of verification of CML-based precipitation and interpolated telemetric rain gauges , corrected radar , satellite , and operational RainGRS () data, using manual rain gauge data as a reference, daily accumulations (in mm) from July 19 – August 18, 2022.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estimate** | **BIAS** (mm) | **RMSE** (mm) | **RRSE** (--) | **CC** (--) |
|  | -1.27 | 4.37 | 0.71 | 0.854 |
|  | -0.92 | 2.91 | 0.48 | 0.905 |
|  | -0.05 | 2.14 | 0.35 | 0.938 |
|  | -1.77 | 6.02 | 0.96 | 0.486 |
|  | -0.35 | 2.04 | 0.34 | 0.952 |

**3. Development and automatic quality control methods for rain gauge data based on the RainGaugeQC algorithms**

Data from non-professional rain gauge networks may constitute additional data which allow a better representation of the precipitation field with high spatial resolution, but generally they are subject to higher uncertainty than professional data. Therefore, they require more rigorous quality control for their correction and eliminate erroneous measurements.

*3.1. RainGaugeQC system*

The algorithms of the RainGaugeQC system for real-time quality control (QC) of rain gauge data from telemetric stations, which is used operationally in the IMGW, are described in detail in the publication by Ośródka et al. (2022). Within the framework of the COSMO PT EPOCS project, the RainGaugeQC has been extended with sub-algorithms, which have been designed especially taking into account the specific characteristics of non-professional weather stations. This category includes rain gauges such as private weather stations and other gauges not maintained and supervised by a national meteorological or hydrological service.

Table 1 summarises all the algorithms and sub-algorithms of the RainGaugeQC system including those developed and implemented within the EPOCS project, namely TCC/3 and TCC/4.

Table 1. List of algorithms for QC of rain gauge data designed for the RainGaugeQC system. The green colour indicates the sub-algorithms designed in the frame of PT EPOCS project for QC of data from non-professional stations, including private weather stations.

|  |  |  |
| --- | --- | --- |
| **Abbr.** | **Algorithm** | **Sub-algorithms** |
| GEC | Gross Error Check |  |
| RC | Range Check |  |
| RCC | Radar Conformity Check | 1) Detection of incorrect „no precipitation” data |
| 2) Detection of false precipitation reports |
| TCC | Temporal Consistency Check | 1) Detection of blocked sensors |
| 2) Comparison of two sensors |
| 3) Time series comparison with adjusted weather radar data |
| 4) Bias correction with adjusted weather radar data |
| SCC | Spatial Consistency Check | 1) Detection of outliers from the local vicinity |

*3.2. New TCC/3 sub-algorithm: assessing the consistency of time series of rain gauge and radar data*

The TCC/3 algorithm is designed to eliminate erroneous rain gauge measurements from non-professional networks by analysing the correlation with weather radar observations on long time series.

For the calculation of the correlation coefficient, pairs of rain gauge () and radar () data are taken if both are not equal “no data”, and at least one of the values is greater than 0.025 mm, and their quality index (*QI*) is at least 0.8 for and 0.7 for .

“Long” and “short” data series are analysed:

* for “long” series comprising 10 days using 10-min temporal resolution, a number of pairs and correlation coefficient are calculated,
* for “short”, covering 5 days, a number of and correlation coefficient are calculated.

The number of non-precipitation data pairs is determined from data pairs where both values are less than 0.025 mm.

The procedure for assessing data quality is carried out by checking the following quantities applying preset thresholds:

1. precipitation accumulation in 10 days,
2. the number of data in 5 days and the correlation coefficient ,
3. the number of data in 10 days and the correlation coefficient ,
4. the number of non-precipitation data .

Finally it is checked if there were not enough measurements in analysed period or the correlation coefficient with radar data could not be determined for 5 and 10 days respectively.

For a particular measurement these conditions are checked sequentially and, depending on their result, further conditions are checked or the quality index is reduced accordingly. In the case of conditions 2 and 3, the quality index of this measurement is reduced on the basis of the correlation coefficient by the value: .

*3.3. New TCC/4 sub-algorithm: unbiasing non-professional data*

The TCC/4 algorithm is applied for unbiasing rain gauge measurements obtained from non-professional networks based on their quantitative comparison with weather radar observations. For the most recent 10 days using a 10-min temporal resolution, precipitation accumulations are calculated from rain gauge-radar pairs for which both measurements are not equal “no data” and have a quality index of at least 0.8 for and 0.7 for :

and

Unbiasing is carried out and quality index of the data is reduced:

1. If or are equal to “no data”, then the quality index ( is reduced by and unbiasing is not performed.
2. If or are equal to zero, then the quality index ( is reduced by and unbiasing is not performed.
3. The *bias* limited to a factor of 4 is calculated from:

Then the unbiasing is carried out with the formula:

1. The similarity of the accumulations of the two quantities and is checked by means of the similarity function introduced by the algorithm authors (Ośródka et al., 2024). If the accumulations and are similar, then depending on the value, the quality index of a given measurement is reduced accordingly. On the other hand, if the accumulations i are not similar, then depending on the value of , the value of the quality index of a given measurement is also reduced, but to a larger extent.

*3.4. Example of RainGaugeQC system performance*

The RainGaugeQC is implemented operationally at IMGW, in the Centre of Meteorological Modelling. In Fig. 3 an example of the RainGaugeQC performance is depicted: false heavy precipitation in the data applying standard RaingGaugeQC marked by red circle was reduced by means of the improved version with new sub-algorithms TCC/3 and TCC/4. The erroneous data is evident on longer accumulations, however on 10-min precipitation accumulations, for which the RainGaugeQC system is operationally employed, it is more difficult to detect malfunctioning rain gauges.

Obraz zawierający tekst, zrzut ekranu, mapa

Opis wygenerowany automatycznieObraz zawierający tekst, mapa

Opis wygenerowany automatycznie

Fig. 3. An example of the RainGaugeQC system performance: impact of reduction of the false high precipitation values observed by one of the non-professional rain gauges on RainGRS+ estimate. Daily accumulations (in mm) calculated from 10-min products applying standard (left) and improved (right) version of RainGaugeQC. RainGRS+ domain, data from 2023-09-22.

**Acknowledgement**

The work was carried out as part of the Priority Task EPOCS „Evaluate Personal Weather Station and Opportunistic Sensor Data CrowdSourcing” during 2023-2024 in the frame of subtasks: 2.1 “Development and testing automatic QC methods based on the RainGaugeQC algorithms developed at IMGW-PIB” and 3.1 “Processing different rainfall data sources (private rain gauges, commercial microwave links, sewer/water service stations, etc.)”.

**References**

Bárdossy, A., Seidel, J., El Hachem, A., 2021. The use of personal weather station observation for improving precipitation estimation and interpolation. *Hydrology and Earth System Sciences*, **25**, 583–601, <https://doi.org/10.5194/hess-25-583-2021>.

de Vos, L. W., Leijnse, H., Overeem, A., Uijlenhoet, R., 2019. Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. Geophysical Research Letters, 46, 8820–8829, <https://doi.org/10.1029/2019GL083731>.

Jurczyk, A., J. Szturc, I. Otop, K. Ośródka, and P. Struzik. 2020. Quality-based combination of multi-source precipitation data. *Remote Sensing*, **12**, 1709a, [https://doi.org/10.3390/rs12111709.](https://doi.org/10.3390/rs12111709)

Ośródka, K., I. Otop, and J. Szturc, 2022. Automatic quality control of telemetric rain gauge data providing quantitative quality information (RainGaugeQC), *Atmospheric Measurement Techniques* 15 (19): 5581–5597. <https://doi.org/10.5194/amt-15-5581-2022>.

Ośródka, K. and J. Szturc, 2024. Adaptation of RainGaugeQC algorithms for quality control of rain gauge data to non-professional or hybrid measurement networks (RainGauge-2), (in preparation).

Pasierb M., Baldysz Z., Szturc J., Nykiel G., Jurczyk A., Ośródka K., Figurski M., Wojtczak M., Wojtkowski C., 2024. Application of commercial microwave links (CMLs) attenuation for quantitative estimation of precipitation, *Meteorological Applications*, (in review).

Saltikoff, E., Haase, G., Delobbe, L., Gaussiat, N., Martet, M., Idziorek, D., Leijnse, H., Novák, P., Lukach, M., and Stephan, K., 2019. OPERA the radar project, *Atmosphere*, **10**, 320, <https://doi.org/10.3390/atmos10060320>.

Sinclair, S. and Pegram, G., 2005. Combining radar and rain gauge rainfall estimates using conditional merging*, Atmospheric Science Letters*, **6**, 19–22, <https://doi.org/10.1002/asl.85>.