Assimilation of surface pressure observations from personal weather stations in AROME-France

Alan Demortier[a], Marc Mandement[a], Vivien Pourret[a], and Olivier Caumont[a,b]

[a]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
[b]Météo-France, Direction des opérations pour la prévision, Toulouse, France

Correspondence: Alan Demortier (alan.demortier@meteo.fr)

Abstract. Spatially dense surface pressure observations from personal weather stations (PWSs) are able to describe pressure patterns at the surface, such as those associated with convective events, in more detail than with standard weather stations (SWSs) only. In this study, the benefit of assimilating PWS observations with the 3DVar and the 3DEnVar data assimilation schemes of the AROME-France model is evaluated over a 1-month period and during a heavy precipitation event in the south of France. Observations of surface pressure from PWSs are bias-corrected, quality-controlled, and thinned with a spacing equal to the horizontal dimension of an AROME-France grid cell. Over France, almost half of the 55 187 available PWS observations are assimilated, which is 129 times more than the number of assimilated SWSs observations. Despite the limited advantages found from their assimilation with the 3DVar assimilation scheme, the 3DEnVar assimilation scheme shows systematic improvement and reduces by 10.3 % the root-mean-square departure in surface pressure between 1 h model forecasts and SWS observations over France. Significant improvement is observed over the first 9 h of the forecasts in mean sea level pressure. Finally, when PWS observations are assimilated with the 3DEnVar assimilation scheme, a surface pressure anomaly generated by a mesoscale convective system – observed by PWSs and not visible without them – is successfully assimilated. In that case, the forecasts of location and temporal evolution of the mesoscale convective system as well as rainfall are closer to the observations when PWS observations are assimilated.

1 Introduction

Amongst all the challenges to improve data assimilation (DA) for convection-permitting numerical weather prediction (NWP), the use of new observation types has emerged (Hu et al., 2022). Indeed, accurately estimating the initial state (analysis) of a convective-scale model requires the assimilation of spatially dense and temporally frequent observations (Gustafsson et al., 2018). A step forward in improving analyses could come from near-surface crowdsourced observations, e.g. from smartphones or personal weather stations (PWSs) (Hintz et al., 2019). Crowdsourced observations are much more numerous than observations of standard weather stations (SWSs) currently assimilated in NWP models, while having a more heterogeneous quality and less metadata (Nipen et al., 2020; Sgoff et al., 2022).

Clark et al. (2018) and Mandement and Caumont (2020, hereafter MC20) showed that PWS observations, when combined with SWS observations, reveal small-scale spatio-temporal variations in screen-level temperature, relative humidity, and surface

pressure associated with the life cycles of thunderstorms which are partly visible, or not visible at all, with SWS observations only. Amongst these three observed variables, MC20 showed that use of PWS allows approximately 100 times more surface pressure observations to be assimilated. These observations, when used to produce gridded analyses of mean sea level pressure (MSLP), are the observations that most reduce (of the three observed variables, during four stormy days) the root-mean-square error of these analyses compared to analyses using only SWSs. The shape and amplitude of surface pressure perturbations, which vary from one thunderstorm to another, are described by Madaus and Hakim (2016) using idealized simulations for the particular case of isolated thunderstorms. The interest in assimilating PWS observations into a convection-permitting NWP model has therefore emerged.

As well as varying on a fine scale with the passage of thunderstorms, surface pressure is a crucial variable in NWP models. Surface pressure observations, provide information not only near the surface, but over the entire air mass in a column of the atmosphere, and are less dependent on the surface characteristics (e.g., cities, vegetation; Ingleby, 2015), than other surface observations. Most global and limited-area models assimilate surface pressure observations over land and oceans from the World Meteorological Organization (WMO) Regional/Global Basic Observing Networks (WMO, 2021) exchanged through the WMO Global Telecommunication System (WMO, 2020). In the Met Office global model, Ingleby (2015) showed that surface data contribute to 16.9 % of the total observation impact of 24 h forecast (compared to 10.0 % for radiosondes and 9.2 % for aircraft). In the Météo-France global model ARPEGE, Chambon et al. (2023) showed that surface observations are crucial to reduce forecast errors: over Europe, the Performance Index 18 (IP18) forecast score (based on the arithmetic average of the geopotential, temperature, and wind root-mean-square errors against radiosoundings at 48 and 72 h range and normalized by the value of the year 2008) diminished from 6.51 to −1.63 when the surface observations are removed over a 6-month period (Chambon et al., 2023), i.e. being worse than 2008 score. Surface pressure errors are identified, amongst the other surface observations, as the main source of 24 h forecast errors in terms of a moist energy norm.

These last years, several studies (Madaus et al., 2014; McNicholas and Mass, 2018b; Hintz et al., 2019; Ridal et al., 2019) reported assimilation experiments of crowdsourced surface pressure observations from dense networks to capture local-scale features. Various techniques were designed depending on the kind of observations (e.g., smartphones or PWSs) and the assimilation system.

Among these studies, Madaus et al. (2014) assimilated between 1300 and 1700 pressure observations from PWSs over the Pacific Northwest in addition to the 140 SWSs commonly assimilated in NWP models. One-month assimilation experiments were done in the WRF model using an ensemble Kalman filter DA system. With appropriate bias correction and quality control, they showed significant improvements of 3 h forecasts of surface pressure, and lower but still significant improvements of 2 m-temperature and 10 m-wind meridian component. Furthermore, results showed a better timing of frontal passages (20 to 45 min of improvement) in the forecasts. Then, a few experiments have been done using smartphone observations. Hintz et al. (2019) showed the difficulties of the task due to quality control and bias correction issues because the altitude of the smartphone pressure observations were not accurately known (accentuated by the moving location of the sensors) and have used a 3DVar assimilation system. However, McNicholas and Mass (2018a) developed a machine learning quality control method for smartphone pressure observations. Once applied, quality-controlled observations were assimilated in an ensemble DA

system during two specific cases (McNicholas and Mass, 2018b). Results showed an improvement of the 1 h forecast on both 2 m temperature and humidity; and on the track and intensity of a coastal wind storm. In the same way for PWSs, Ridal et al. (2019), implemented and compared different quality control methods, using in particular a variational bias correction to remove the pressure bias from the observation time series. The results of the three-week assimilation experiments in HARMONIE-AROME, the limited-area NWP model developed within the ALADIN consortium (Bengtsson et al., 2017), which uses a 3DVar DA scheme, did not show any advantage. In summary, the assimilation of dense surface pressure observations from a few pairs of observation network – DA scheme has been tested, but there is a lack of information on the role that the DA scheme plays in making the most of dense PWS surface pressure observations.

Different DA schemes have been developed for AROME-France. AROME-France is the convection-permitting, limited-area model operational at Météo-France since 2008. AROME-France has a three-dimensional variational (3DVar) DA scheme launched hourly and assimilates, in addition to observations assimilated in ARPEGE, dedicated observations for the mesoscale within a 1 h assimilation window (Seity et al., 2011; Brousseau et al., 2016). The background error covariance matrix determines how the information from the observations is spread in the model space (horizontally and vertically) to make the analysis. In the 3DVar scheme, the background error covariance matrix is climatological, spatially homogeneous, and isotropic. Montmerle et al. (2018) and Michel and Brousseau (2021) developed an ensemble DA scheme, to replace the static matrix with a dynamic background error covariance matrix, improving the spread of the information from the observation. This scheme is called three-dimensional ensemble variational (3DEnVar).

This study intends to answer the following questions: does the assimilation of PWS surface pressure observations in the current AROME-France 3DVar data assimilation system improve weather analyses and forecasts? And what are the differences with the 3DEnVar system currently under development?

The remainder of this paper is organized as follows. Section 2 describes the crowdsourced surface pressure observations used and the different assimilation configurations available for AROME-France. The observation pre-processing and the DA experiments are described in Sect. 3. Section 4 presents the scores of one-month cycled assimilation experiments, and a case study of a heavy precipitation event (HPE) is analysed (Sect. 5). Finally, the advantages of the 3DEnVar system for the assimilation of dense surface pressure observations are discussed (Sect. 6).

2 Data, tools, and methodology

A study period of 1 month is chosen, ranging from 6 September 2021 00:00 UTC to 5 October 2021 23:00 UTC. In total, it represents 717 hours for each experiment (excluding two time steps during which PWS observations are missing). This period includes several rainy situations, such as precipitation events in the south-west of France on 8 September, in the south-east of France on 14 September, and over the north and south-east of France between 4 and 5 October.

2.1 Surface pressure observations

Pressure observations are mainly retrieved at the surface with weather stations. Two main surface networks are used, the standard weather stations and the personal weather stations.

2.1.1 Standard weather station observations

Standard weather stations include land surface synoptic stations (sending SYNOP weather reports through the WMO Global Telecommunication System, see World Meteorological Organization (2020)), the French extensive data acquisition and weather observation network (RADOME, Tardieu and Leroy, 2003) and stations located over sea on ships (sending SHIP reports). On average during the one-month study period, 1124 surface pressure observations per hour are available in the AROME-France domain coming from manual land SYNOP (10.0 %), automatic land SYNOP (58.6 %), French RADOME (17.9 %), manual SHIP (2.0 %) and automatic SHIP (11.4 %) reports. As France is a fraction of the AROME-France domain, only a part of these stations are located in France. Indeed, among these reports, 214 are located in France over land: 10.1 % are automatic land SYNOP and 89.9 % are French RADOME. Stations associated to these surface pressure observations are further called SWSs. Most of the scores will focus on SWSs located in France.

2.1.2 Personal weather stations observations

Amongst all types of Personal Weather Station (PWS) networks, the Netatmo network stands out being at the same time the largest in France and available in near-real time (MC20). The stations are privately owned, which explains why they are not spatially equally distributed (Fig. 1) i.e., they are denser in the cities.

Netatmo PWS is composed of outdoor and indoor modules. The indoor module measures the surface pressure. Observation times are not exactly regular: the mean time step between two observations is approximately 5 min, and the company updates its servers only every 10 min. The owner of the station can perform on-demand measurements, which shift the following observation times. Also, the evolution of the surface pressure related to the passage of thunderstorms is of the order of a few minutes. For all these reasons, to limit time lags, the observations are linearly interpolated at fixed times (i.e., round hours) using the two closest measurements around the interpolation time. If the two closest measurements are separated by a period of 700 s or more, the observation is removed.

The absolute accuracy of the pressure sensor of the Netatmo PWS given by the manufacturer is ±1 hPa (Netatmo, 2023). In 2018, the sensor was an STMicroelectronics LPS25HB, which has the same absolute accuracy indicated by Netatmo and a relative accuracy of ±0.2 hPa after calibration for standard atmospheric surface conditions (STMicroelectronics, 2016). The exact model of the sensor may change regularly, however the absolute accuracy provided by Netatmo has not changed since 2019 (MC20), suggesting that their characteristics remain similar. Once the PWS is installed and running, pressure sensor calibration is possible but is not performed by default, which can lead to systematic errors. One of the main sources of surface pressure bias is errors in the station altitude, caused in part by uncorrected GPS heights (Pauley and Ingleby, 2022). When PWS MSLP observations were compared to SWS analyses taken as reference, MC20 showed that the first (Q25) and the third (Q75)
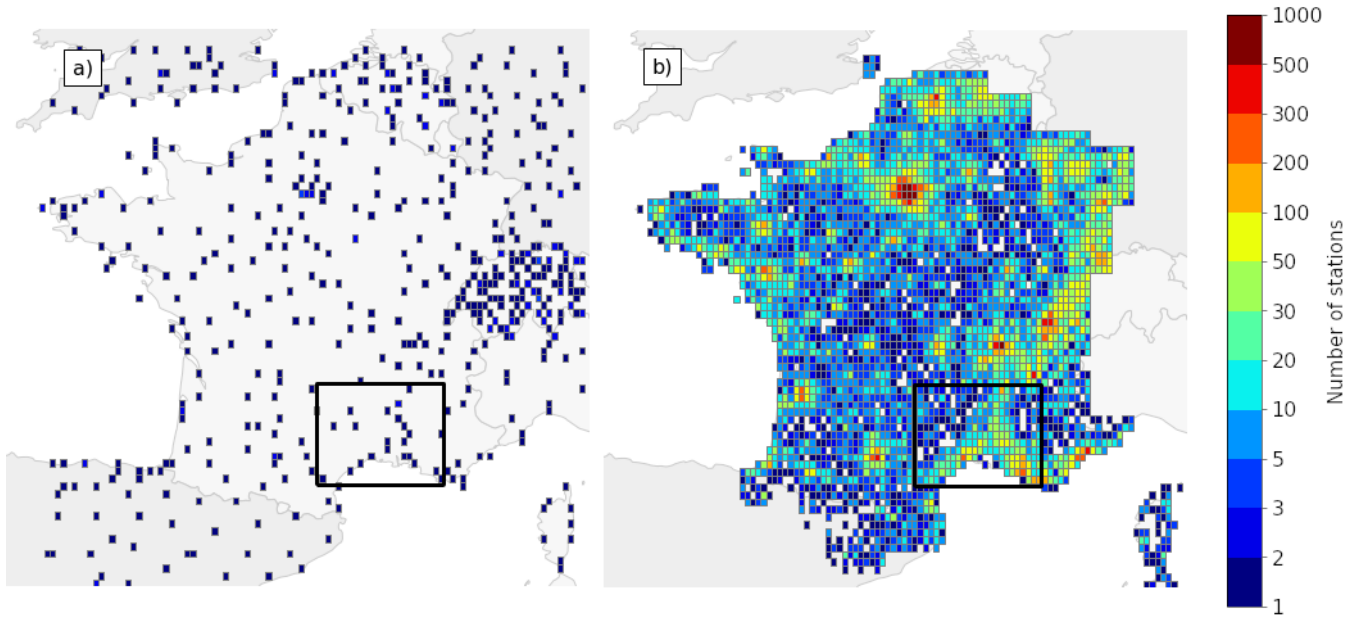
Figure 1. Number of (a) SWSs and (b) PWSs providing at least one observation during the one-month study period over France. Observation counts are binned into approximately $0.15° \times 0.1°$ bins. The black rectangle delimits the domain used for the case studied in Sect. 5.

quartiles of the systematic errors were equal to $-4\,\text{hPa}$ and $2\,\text{hPa}$ respectively. The supplementary errors result from lagged-time pressure measurements to pressure change, and inaccurate or unavailable metadata. These constraints will be taken into consideration in the pre-processing method.

2.2 AROME-France assimilation system

AROME-France is the Météo-France operational limited-area model (Seity et al., 2011; Brousseau et al., 2016). The model is convection-permitting at $1.3\,\text{km}$ horizontal grid spacing with 90 vertical hybrid pressure levels. It does not rely on the hydrostatic assumption and fully resolves elastic Euler equations. Physical parametrizations come from the Meso-NH research model (Lac et al., 2018). At the surface, AROME-France is coupled with the SURFEX model (Masson et al., 2013). The model is regularly updated with new features, creating new development cycles. In the current study, the cycle 48t1 of AROME-France is used. Hereafter, AROME will systematically refer to AROME-France.

The DA scheme combines the active observations (see below for the meaning of active) with the background (for AROME the $1\,\text{h}$ forecast starting from the previous analysis), to produce x_a the analysis state of the atmosphere, obtained by the minimization of the cost function J:

$$J\left(\boldsymbol{x}\right) = \frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{x}_b\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\boldsymbol{x} - \boldsymbol{x}_b\right) + \frac{1}{2}\left[\boldsymbol{y}_o - \mathcal{H}\left(\boldsymbol{x}\right)\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\boldsymbol{y}_o - \mathcal{H}\left(\boldsymbol{x}\right)\right],$$ (1)

in which, $\boldsymbol{x}_b$ is the background state, $\boldsymbol{y}_o$ the observation vector, $\mathcal{H}$ the nonlinear observation operator, and $\mathbf{B}$ and $\mathbf{R}$ are the background and the observation error covariance matrix, respectively. $\mathbf{B}$ approximates the error of the model at the observation point, propagates the information from the observation in both horizontal and vertical directions and influences other variables. To reduce the computational cost, the formulation used to find the solution from J is incremental i.e., using δx the difference between the atmosphere state and the one from the background (Courtier et al., 1998), at full model resolution. If $\mathbf{B}$ is computed at a lower resolution than the model, interpolation operators from low to high resolution are applied.

The difference between an observation and its corresponding value in the background, called observation minus background (OmB), provides a comparison between the observations and the model 1 h forecast. The OmB and the observation minus analysis (OmA) can be calculated using observation operators. However, the OmB is rather preferred in score calculation, due to correlation that analysis has with observations.

AROME uses a 3DVar data assimilation scheme, with an hourly data assimilation cycle, to keep as close as possible to the true state of the atmosphere (Brousseau et al., 2016; Gustafsson et al., 2018). The 3DVar analysis done at time t_0 considers all observations between $t_0 - 30\,\mathrm{min}$ and $t_0 + 30\,\mathrm{min}$ to be valid at t_0. To avoid redundancy issues, only the observation closest to t_0 is taken into account for stationary stations (SWS except ships) providing multiple observations during the assimilation window. Only PWS observations valid at t_0 are used, following the method described in Sect. 2.1.2. All available observations are submitted to a quality control procedure called screening, which removes the observations deemed questionable. One of the screening steps compares the observation vector $\boldsymbol{y}_o$ to the background state $\boldsymbol{x}_b$ using the observation operator $\mathcal{H}$, and keeps observations such as $\|\boldsymbol{y}_o - \mathcal{H}(\boldsymbol{x_b})\| < 5\sigma$ with $\sigma = \sqrt{\sigma_o^2 + \sigma_b^2}\sigma_{\mathrm{coef}}$, given σ_o and σ_b, the standard deviation of observation and background errors, respectively, and $\sigma_{\mathrm{coef}} = 0.9^{-2}$ a supplementary factor which artificially increases σ allowing to keep more observations. Two other screening steps that are important for mass observations (surface pressure and geopotential height) include checks for missing station altitude, and compliance with WMO observation standards: if the reported observation level is far from the true station level, σ_o is inflated (more details in Sect. 4.4.1 of ECMWF, 2023). Height differences between stations and the ground level of the model, which can be large in mountainous areas, are not managed by the screening but rather by the observation operator described in Sect. 2.2.1. Observations rejected by the screening are not used further.

After this step, the observations kept are called active observations. This procedure matters, removing suspicious observations, which helps the convergence of the minimization algorithm and limits the spin-up period. However, some observations might be rejected when they should not because of an inaccurate background. Additional quality control is tested in Sect. 3 to reject observations depending on the background departure of their time series rather than the background departure at one specific time.

The 3DVar and 3DEnVar share the same formulation, with the only difference that $\mathbf{B}$ is static in the 3DVar and becomes $\mathbf{B}_e$, an ensemble error covariance matrix, in the 3DEnVar (Michel and Brousseau, 2021). The matrix $\mathbf{B}_e$ is estimated for each analysis from an Ensemble Data Assimilation (EDA) system composed of 50 members at 3.2 km horizontal grid spacing and with 90 vertical levels (Brousseau et al., 2016). Despite the fact that the coefficients of the $\mathbf{B}_e$ matrix are given at 1.3 km of resolution, the structure of the increments is still at the spatial resolution of the coarser grid. The advantage of the matrix $\mathbf{B}_e$ is to have a flow-dependent background error statistics updated at each analysis, which allows the model to take better

advantage of the observation. However, in order to remove the sampling noise of the model error covariance estimate, a spatial localization is applied (Hamill et al., 2001), which limits the propagation of the OmB to shorter distances. The localized matrix $\hat{\mathbf{B}}_e$ is obtained making an element-to-element product (known as a Hadamard–Schur product) between $\mathbf{B}_e$ and a correlation matrix $\mathbf{L}$ of the same dimension (Eq. 2).

175 $$\hat{\mathbf{B}}_e = \mathbf{B}_e \circ \mathbf{L}, \tag{2}$$

The correlation functions in the $\mathbf{L}$ matrix are quasi-Gaussian with horizontal and vertical localizations (Montmerle et al., 2018). The specified length of localization corresponds to the half width at half maximum (HWHM) of the quasi-Gaussian function. The tuning of the horizontal localization has been found to be optimal using a linear function from 25 km at the ground to 150 km at the highest model level (Michel and Brousseau, 2021). The vertical localization is set to about 0.15 hPa (Michel and
180 Brousseau, 2021).

2.2.1 Assimilating geopotential at the level of the observed station pressure

In AROME, the geopotential is equal to gz with $g = 9.806\,65\,\mathrm{m\,s^{-2}}$ the standard acceleration of gravity and z the altitude. For each observation of surface pressure (p_s) at the altitude of the station's barometer (z_s), the geopotential equal to gz_s is assimilated at the vertical level corresponding to p_s. For geopotential observations, σ_o is set to $8g$. The model geopotential at
185 the location of the observation is retrieved using an observation operator described by Courtier et al. (1998).

Surface pressure and geopotential variables are both used in this study to refer to the same observations. The surface pressure will be used when studying the observations, whereas the corresponding geopotential at the level of the observed station pressure will be the variable used when looking at assimilation statistics.

The propagation of the OmB from a surface pressure observation differs between the 3DVar and the 3DEnVar DA systems.
190 If we consider that a typical horizontal scale in the surface pressure analysis is given by the distance at which the amplitude of a surface pressure analysis increment divided by its maximum value is below 1 %, then, this scale is approximately 200 km for the 3DVar scheme and 40 km for the 3DEnVar scheme (Fig. A1). In 3DEnVar this typical horizontal scale can slightly vary depending on the meteorological situation.

2.3 Evaluation tools

195 When assimilating new observations, the associated $\mathbf{R}$ matrix has to be specified. The allocated part of the $\mathbf{R}$ matrix for PWSs will have the same characteristics as the one for the SWSs, i.e. it is assumed to be diagonal, and the diagonal values are the same. This assumption requires verifying that observation errors are not correlated and, if they are, using techniques such as thinning to reduce the negative impact of this correlation. The effect of different σ_o between PWS and SWS observations could be investigated in a sensitivity study.
200 Desroziers et al. (2005) proposed a posterior diagnostic to estimate the observation error covariance matrix $\mathbf{R}$, which can be estimated by $E(\mathbf{OmA}[\mathbf{OmB}]^T)$. This method has some limitations. It assumes that the background and the observation error covariance matrices are accurately specified and uncorrelated with a sufficiently large and representative sample of forecast

error statistics (Pourret et al., 2022). However, the **B** matrix is not totally estimated or stored but specified through operators in the 3DVar scheme (Seity et al., 2011; Brousseau et al., 2014). Some limitations will be discussed further.

The estimation of the correlation between observations errors matters. Liu and Rabier (2003) showed that increasing the number of observations with uncorrelated errors generally increases the quality of the analysis and the forecast. However, if the observations have correlated errors greater than a threshold (e.g., 0.2 as used by Liu and Rabier, 2003), it decreases the quality of the analysis and the forecast. To prevent observation errors from being spatially correlated, a thinning technique which removes observations at an optimal spatial resolution can be applied (Benáček et al., 2016). Thinning is done for observing systems providing high spatial resolution observations which have correlated observation errors, such as radar (Wattrelot et al., 2012), satellites (Dando et al., 2007), or aircraft (Pourret et al., 2021).

This method is not used for SWS, but given the high spatial density of the PWS (Fig. 1b), is tested further. To find this optimal spatial resolution, the Desroziers posterior diagnostic can be extended to a method estimating the spatial correlation of observation error (Mile et al., 2019). First, for each distance bin, from 0 to 30 km at 1 km intervals, it finds equidistant pairs of stations. Then for each pair (i, j) of stations in each bin, the covariate $\mathrm{OmA}_i.\mathrm{OmB}_j$ is calculated. Finally, the average of the covariates is normalized by the average of the first covariate, when the distance is zero, which corresponds to the actual Desroziers diagnostic. This diagnostic gives an estimation of the spatial correlation errors, which can be used to estimate a minimum thinning length where the correlation errors are significantly low. It has the same limitation as the Desroziers diagnostic previously described.

2.4 Design of the experiments

In order to evaluate the benefits of the new observations, the Observing System Experiments (OSE) framework is used to compare a reference experiment with an experiment where new observations are assimilated (e.g., Pourret et al., 2022). OSEs allow us to compare the background (i.e., 1 h forecast) or the analysis of both experiments to observations. In AROME the examination can be both at the surface with observations from stations and at different levels of the atmosphere with observations from radiosondes, aircraft, and satellites.

Sect. 3 describes an experiment monitoring the raw pressure observations from the PWSs in a non-cycled 3DVar DA scheme. This monitoring experiment will be used to compute both the bias correction and the quality control of the PWS observations. Then, the Sect. 4 will focus on long cycled experiments with the 3DVar and the 3DEnVar assimilation schemes. Finally, the Sect. 5 will use long, cycled experiments, and an extra experiment assimilating the new observations only at 07:00 UTC on 14 September 2021. All the experiments are described in Table 1. The experiments without assimilating PWS observations will also be called control experiments.

3 Pre-processing algorithm

All PWS observations are subject to a pre-processing algorithm before assimilation composed of a bias correction, quality control, and spatial and temporal thinning. The number of observations remaining after each step of the process is provided in

Table 1. Overview of the experiments. 3DVar, 3DVarP, 3DEnVar and 3DEnVarP experiments are launched on 6 September 2021 00:00 UTC. The Monitor experiment is launched on 6 August 2021 00:00 UTC (one month earlier). The 3DEnVarP7UTC experiment is launched on 14 September 2021 07:00 UTC. BC means bias correction and QC means quality control.

Experiments	Duration	Cycling	Data assimilation scheme	Use of PWS surface pressure observations
Monitor	2 months	No	3DVar	Raw observations monitored
3DVar			3DVar	No
3DVarP	1 month	Yes		Assimilated after BC, QC, and thinning
3DEnVar			3DEnVar	No
3DEnVarP				Assimilated after BC, QC, and thinning
3DEnVarP7UTC	1 h	No	3DEnVar	Assimilated after BC, QC, and thinning

235   Table 2. Starting from raw observations, they are interpolated to round hours, using the method described in Sect. 2.1.2, which removes some observations. Stations with identical coordinates are assumed to have not been properly set up by the owner, leading to automatic placement based on their wireless IP address (Meier et al., 2017). These stations are further removed.

3.1   Bias correction and quality control

Figure 2a shows one-month time series of surface pressure observations from 30 PWSs close to the city of Toulouse (i.e., less
240   than 0.05 ° from the city centre), located in the south of France. Most of the time series exhibit similar pressure variations, and large differences in the mean value. However, the area being relatively flat (i.e., less than 150 m from lowest to highest point in this area), it is not realistic to have a systematic difference up to 25 hPa (approx. 250 m), i.e., a first glance of the different systematic errors at each station location. Then, it has to be confirmed with long-term statistics, which can be done with the monitoring experiment. In this experiment, which assimilated the raw pressure observations from the PWSs, the observations
245   are compared to the model background (i.e. the one-hour forecast from the operational AROME model), giving the OmB for the whole set of observations. The 30 PWS time series have systematic biases from $-500\,\mathrm{m^2\,s^{-2}}$ to $1200\,\mathrm{m^2\,s^{-2}}$ in terms of geopotential (Figure 2b). More globally, Figure 3 shows the whole set of geopotential OmB during the pre-processing. The raw data exhibit the biases from $\pm 1000\,\mathrm{m^2\,s^{-2}}$ (looking at the whiskers, representing the percentile range of the data, i.e. from 5th to 95th percentiles). Therefore, a systematic bias correction is crucial.
250   To implement a geopotential bias correction for PWS data based on AROME, the previously described monitoring experiment with a non-cycled 3DVar DA scheme has been used. PWS individual biases are then obtained by averaging the geopotential OmB for each station over a rolling time window of 1 month to prevent the correction from depending on the meteorological situation and for robust statistics. This bias is then subtracted from the actual observation which, at the end, only changes the altitude of the station in the assimilation scheme. PWSs with missing observations for more than half the rolling
255   time window are eliminated to ensure the robustness of the bias. This bias correction reduces the bias range to $\pm 15\,\mathrm{m^2\,s^{-2}}$ and removes 4.6 % more stations while keeping numerous ones with heterogeneous quality (Fig. 3b). Contrary to SWSs, some

Table 2. Average number of PWS surface pressure observations following the pre-processing process between 6 September 2021 00:00 UTC and 5 October 2021 23:00 UTC (717 time steps). Each step of the process is described in the text. Std refers to standard deviation of the OmBs last rolling month (see text).

	Raw	Preparation		BC	QC		
		interpolation	Same lat-lon	>1/2 of the period	Station has moved	>1/4 during the last 48 h	std > 60 m^2 s^{-2}
PWS remaining	55187	52334	50757	48275	47748	47564	46800
% of total PWS	100 %	94.8 %	92.0 %	87.4 %	87.0 %	86.2 %	84.8 %

PWSs can suddenly move, and their spatial coordinates or altitude metadata change. Some other stations also show sudden pressure change without an abnormal change in their metadata (Fig. 2a). That is why a supplementary quality control has been tested. First, PWSs that have moved from more than 100 m on the horizontal and 1 m on the vertical are discarded. That represents 0.4 % more stations removed. Also, an interruption in the PWS observation time series is suspected to be an early sign of erroneous observations: PWS with missing observations for more than a quarter of the last 48 h are eliminated. Finally, PWSs whose geopotential observations deviate too much from AROME background, i.e. with a standard deviation of their last month's geopotential OmB above, $60 \, \mathrm{m^2 \, s^{-2}}$ are removed. This empirical threshold removes the tail of the OmB distribution (i.e., 1.4 % of the observations) corresponding to unrealistic observations, while keeping physical patterns. In the end, 84.8 % of the stations are kept, and the range of the geopotential OmB is under the SWS ones (Figs. 3c, d).

3.2 Thinning

After the bias correction and the quality control, the number of PWS observations within a 1.3 km horizontal grid cell of AROME ranges from 0 to 69, 69 being reached in the centre of Paris. In order to keep a minimal weight to the model's background and other observations at such grid points where several PWS geopotential observations are available, it is chosen to thin PWS observations at 1.3 km, the model's horizontal grid spacing, for the whole studied period, independently of the assimilation scheme. Among PWS observations available in a grid point, one is randomly drawn. Such thinning, not usually applied for independent surface observations, is generally applied in DA to minimize the risk of having correlated observation errors. For surface pressure, correlated errors could come from instrument systematic bias or correlation of the model altitude error at the subgrid scale. To verify that PWS geopotential observation errors after thinning are not correlated, a spatial Desroziers diagnostic (Sect. 2.3) is computed for two analyses. The analysis have been chosen at midnight and at mid-afternoon.

Figure 4 presents this diagnostic for pre-processed pressure observations from PWSs. From the first bin (1 km) there is limited spatial correlation of observation error. This diagnostic shows small variability between the two assimilation schemes and the hour of study. It confirms that, if a thinning radius is chosen, there is no need to increase the thinning radius above 1.3 km to mitigate the issue of correlated errors. The effect of assimilation without thinning or shorter thinning lengths is not tested. After thinning, there is an average of 27 889 PWS surface pressure observations per hour assimilated, which represents 51 % of the initial amount of PWS observations.
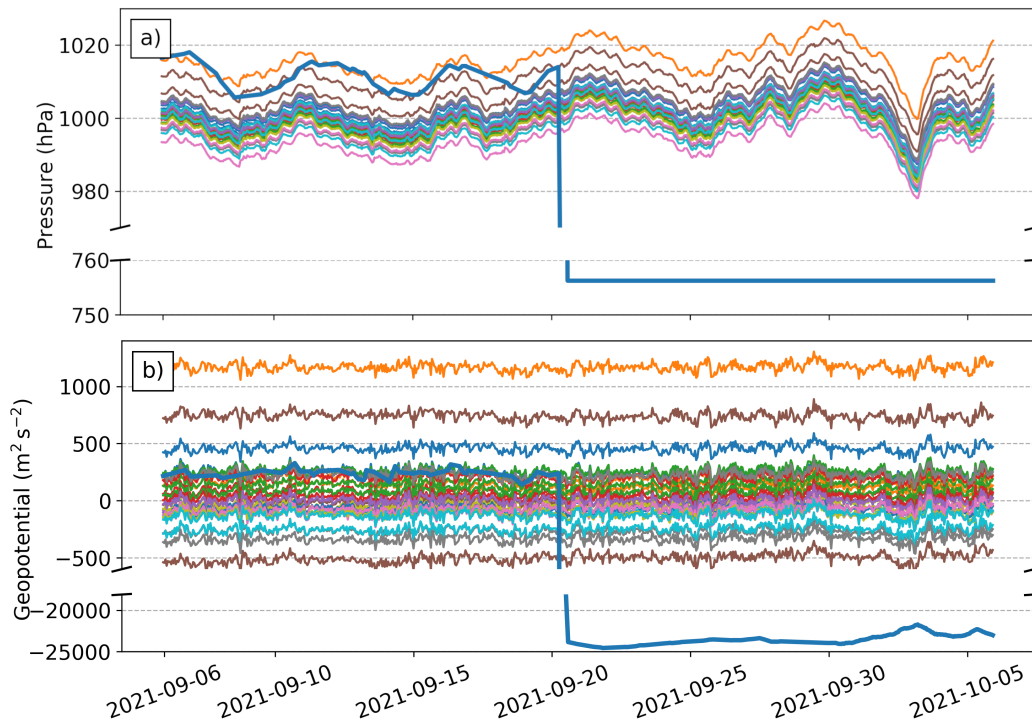
Figure 2. Time series of (a) surface pressure observations and (b) corresponding geopotential OmBs from 30 PWSs around Toulouse and one anomalous PWS during the study period from 6 September to 5 October 2021. The background is from the Monitor experiment.

4 Results of 1 month assimilation experiments

When statistical values (STAT) such as standard deviation or root-mean-square (RMS) of time series are compared between experiments, their relative evolution in an experiment (XP) with respect to another (CTRL) is given by:

$$\Delta\text{STAT} = \frac{\text{STAT}_{\text{XP}} - \text{STAT}_{\text{CTRL}}}{\text{STAT}_{\text{CTRL}}} \tag{3}$$

Special attention will be paid when the CTRL statistical value is close or equal to 0, in which case the relative evolution is not defined. The significance of the values is determined using a bootstrap method, where the RMS is computed from the time series sampled with replacement 1000 times. The 1000 RMS values form a distribution, from which 95 % confidence intervals are estimated using the "percentile" method (Virtanen et al., 2020).

The OmBs from the four experiments described in Table 1 are compared in this section. When pressure observations from PWSs were added to the AROME assimilation cycles, no convergence problems were reported.
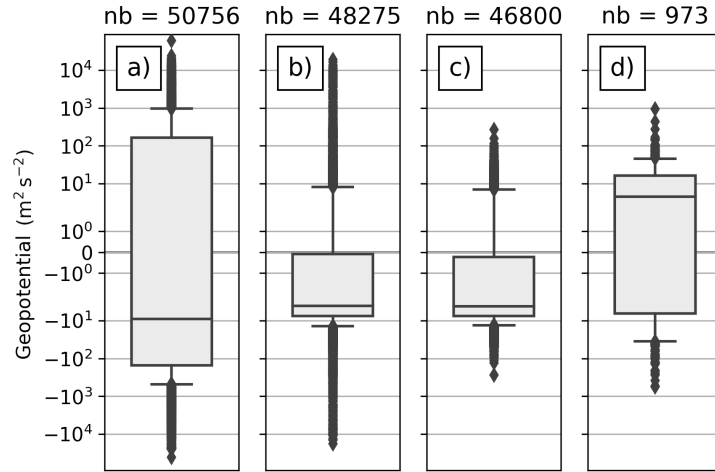
Figure 3. Box plots of geopotential OmB during the one-month study period. Geopotential observations originate from (a) PWS raw data, (b) PWS bias-corrected data, (c) PWS assimilated (bias-corrected and quality-controlled), and (d) SWS assimilated. The background fields are coming from the non-cycled monitoring experiment. Box-plot titles indicate the hourly average number of stations. The bar across the box represents the median, and its borders represent the 25th and 75th percentiles. The box-plot whiskers indicate the 5th and 95th percentiles of the data. Outliers are plotted as separate dots.
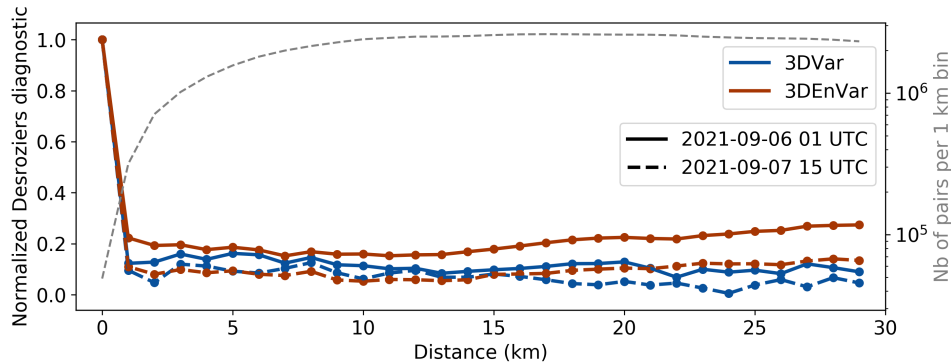
Figure 4. Spatial Desroziers diagnostic of PWS surface pressure observations. The colours represent the assimilation scheme, with the 3DVar, in blue, and 3DEnVar, in brown. The diagnostic is done at (plain lines) 01:00 UTC on 5 September 2021 and (dashed lines) at 15:00 UTC on 6 September 2021. In grey, the number of pairs of equidistant stations (in 1 km bins).

Table 3. Mean and standard deviation of geopotential OmA and OmB (m^2 s^{-2}) and Desroziers diagnostic ($\widetilde{\sigma_O}$ in m^2 s^{-2}) for SWSs located in France and PWSs over the one-month study period, which represent respectively 212 and, 27889 observations per hour in average over the 717 hours of the experiments.

	Monitor		3DVar		3DVarP				3DEnVar		3DEnVarP			
	PWS		SWS		SWS		PWS		SWS		SWS		PWS	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
OmA	n.a.	n.a.	4.7	26.5	11.9	25.4	−0.1	14.6	12.8	38.0	12.2	24.8	0.2	15.9
OmB	-2.6	36.3	11.4	43.3	16.7	41.7	3.9	48.9	15.6	46.4	15.3	41.3	2.6	49.2
$\widetilde{\sigma_O}$	n.a.		31.2		27.7		15.7		43.6		29.3		19.7	

4.1 Impact on geopotential analyses and 1 h forecasts

Statistics of geopotential OmA and OmB for SWSs and PWSs gathered in Table 3 reveal the ability of the assimilation schemes to take advantage of the PWS geopotential observations.

In the 3DVar experiment, the OmA geopotential mean for SWSs is equal to 4.7 m^2 s^{-2}, which shows a little bias between AROME analyses and the SWS geopotential observations over France. Over the whole domain, it decreases to 2.0 m^2 s^{-2} (not shown). Whereas, in the 3DEnVar experiment, the OmA geopotential mean for SWS is higher and equal to 12.8 m^2 s^{-2}. The geopotential standard deviation OmA increases by 43 % from 26.5 to 38.0 m^2 s^{-2} between 3DVar and 3DEnVar. This can be due to the finer localization given to geopotential SWS observations in 3DEnVar than in 3DVar. In 3DVarP and 3DEnVarP, experiments in which PWS geopotential observations are assimilated, the OmA geopotential means for PWSs are closer to zero (respectively −0.1 m^2 s^{-2} and 0.2 m^2 s^{-2}). This result was expected given the prior PWS bias correction using AROME's OmB statistics over a rolling month.

Between 3DVar and 3DVarP, the OmB geopotential mean for SWSs increases, while the geopotential standard deviation OmB for SWSs decreases by 4 %. Time series of geopotential mean and standard deviation OmB shows the same trends every day (Fig. 5a).

Between 3DEnVar and 3DEnVarP, both the OmB geopotential mean and standard deviation for SWSs decrease (i.e. −12 % for the standard deviation). Time series of geopotential standard deviation OmB shows the same trends every day (Fig. 5b), while 3DEnVarP is not systematically closer to 0 than 3DEnVar for the geopotential mean OmB. Times series also show that improvements in OmB statistics (mean and standard deviation) of 3DEnVarP compared to 3DEnVar are larger on days when 3DEnVar OmB statistics are furthest from zero, which corresponds to days of unsettled weather, such as on 8 and 14 September and 5 October 2021.

In the framework of these experiments, considering SWS observations as the reference, the assimilation of PWS observations in a 3DVar scheme has a mixed impact: it increases OmA and OmB geopotential mean but decreases OmA and OmB geopotential standard deviation. The large increase in OmA geopotential mean when PWS observations are assimilated could

315 be explained by the PWS bias correction method, which uses OmB statistics, and thus contains model or other observation biases (as SWS ones). As 129 times more PWS observations than SWS observations are assimilated, with the same prescribed observation error, the 3DVar long range PWS increments overwhelm the SWS ones, which questions the relative number of observations in each subset and its impact in different analysis methods. On the contrary, the assimilation of PWS observations in a 3DEnVar scheme systematically improves all statistics of both analyses and 1 h forecasts.

320 The Desroziers diagnostic ($\widetilde{\sigma_O}$), described in Sect. 2.3, calculated from OmA and OmB statistics, encompass these results. For all experiments, values of this diagnostic estimating the observation error range from 15.7 to 43.6 $m^2\,s^{-2}$ (Table 3), which is only between 20 and 56 % of the prescribed value in AROME (78.4532 $m^2\,s^{-2}$). For SWSs, the Desroziers diagnostic increases by 40 % from 31.2 to 43.6 $m^2\,s^{-2}$ between 3DVar and 3DEnVar experiments, which shows a high sensitivity to the choice of the assimilation scheme. Also, the Desroziers diagnostic varies for SWSs when PWSs are assimilated: it decreases by 12 %

325 between 3DVar and 3DVarP and by 48 % between 3DEnVar and 3DEnVarP. Finally, the PWS bias correction explains why the Desroziers diagnostic is lower for PWSs than for SWSs in 3DVarP and 3DEnVarP experiments, which could be wrongly interpreted as PWS geopotential observations are of better quality than SWS geopotential observations (see Appendix B). These findings show some limitations of this diagnostic, which should be used with caution, especially when its assumptions such as that the average OmB and OmA are close to zero are not fulfilled.
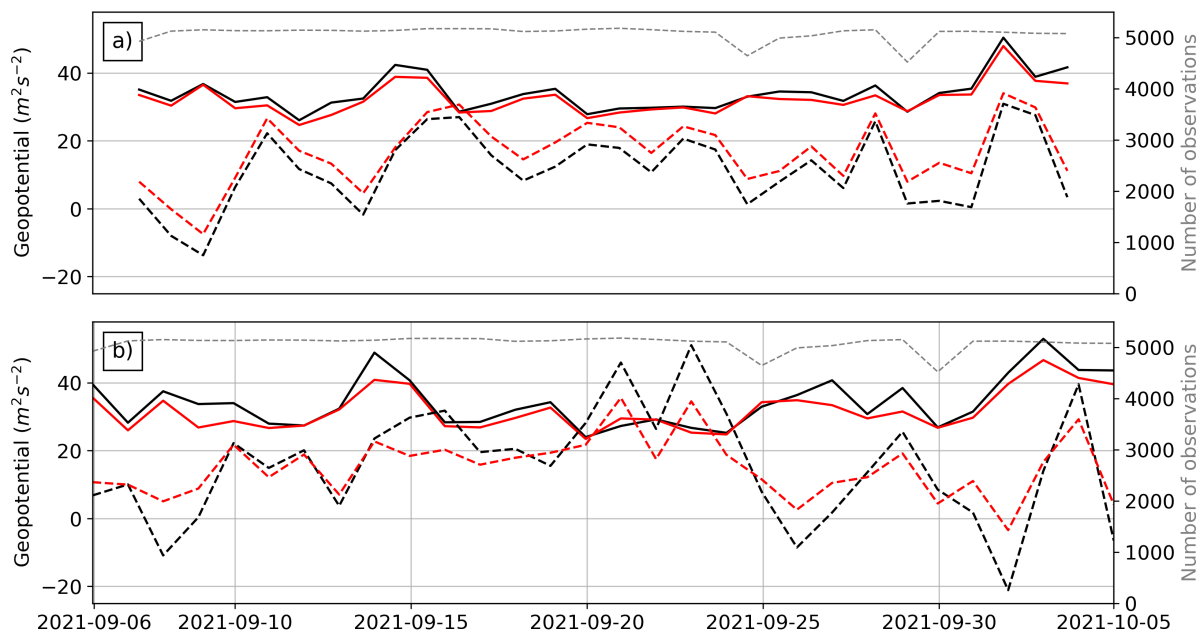
Figure 5. Daily average geopotential (dashed line) mean and (plain line) standard deviation OmB for (a, black) 3DVar, (a, red) 3DVarP, (b, black) 3DEnVar, and (b, red) 3DEnVarP experiments over the one-month study period. Only observations available for all experiments are taken. The grey dashed line is the daily number of observations.

Table 4. ΔRMS OmB of surface pressure, 2 m temperature, 2 m relative humidity and 10 m zonal and meridian wind for SWSs over the one-month study period. Negative values of an experiment (XP) w.r.t. another (CTRL) indicate that 1 h forecasts of XP are closer than CTRL to SWS observations. Significant values are in bold (see text for details).

	Surface pressure		Temperature		Relative humidity		Zonal wind (U)		Meridional wind (V)	
	AROME domain	France	AROME domain	France	AROME domain	France	AROME domain	France	AROME domain	France
3DEnVar w.r.t. 3DVar	**8.0 %**	**9.3 %**	**2.2 %**	**1.2 %**	**3.6 %**	**−1.1 %**	**2.0 %**	**1.1 %**	**2.0 %**	**1.3 %**
3DVarP w.r.t. 3DVar	0.8 %	0.7 %	0.0 %	−0.8 %	0.0 %	0.0 %	−1.4 %	**−1.6 %**	−1.4 %	**−1.5 %**
3DEnVarP w.r.t. 3DEnVar	**−3.0 %**	**−10.3 %**	**0.6 %**	**0.8 %**	**0.8 %**	**1.3 %**	0.2 %	0.0 %	-0.2 %	-0.2 %

4.2 Impact on 1 h forecasts of surface pressure, temperature, relative humidity and wind

To show how the background (1 h forecast) deviates from SWS observations, the RMS OmB of a given variable observed near the ground by SWSs is calculated. This score is presented in Table 4.

The comparison of 3DEnVar with respect to 3DVar shows an 8.0 % increase in RMS OmB of surface pressure for SWSs over the AROME domain and 9.3 % over France. It indicates that the 3DEnVar background deviates more than the 3DVar background from SWS observations of surface pressure. For 2 m temperature and 10 m zonal and meridional wind, increases in RMS OmB range from 1.1 % to 2.0 %. Only the RMS OmB of 2 m relative humidity decreases by 1.1 % for SWSs over France but still increases by 3.6 % over the AROME domain.

The comparison of 3DVarP with respect to 3DVar shows slightly positive ΔRMS OmB of surface pressure for SWSs over France (0.7 %) and the AROME domain (0.8 %), showing a slight degradation, but not statistically significant. Despite this, a significant decrease of −1.6 % and −1.5 % of RMS OmB of the zonal and meridional wind for SWSs located in France are found.

The comparison of 3DEnVarP with respect to 3DEnVar shows significantly negative ΔRMS OmB of surface pressure for SWSs, reaching −10.3 % over France and −3.0 % over the AROME domain. Small but significant degradation from 0.8 % to 1.3 % is found on 2 m temperature and 2 m relative humidity.

These experiments show that PWS assimilation improves significantly the quality of the surface pressure background when a 3DEnVar scheme is used, and partly compensates for the significant degradation associated with the transition to a 3DEnVar system from the 3DVar. Even if they are statistically significant, only a little improvement of the background is found in the AROME 3DVar scheme in terms of wind when PWS geopotential observations are assimilated, with a neutral impact on surface pressure.

4.3 Impact on the forecasts up to 30 h range

For 3DVar, 3DVarP, 3DEnVar, and 3DEnVarP experiments, forecasts starting from the analyses at 00:00, 06:00, 12:00 and 18:00 UTC were run each day during the one-month study period. Statistics of observation minus forecast (OmF) and their
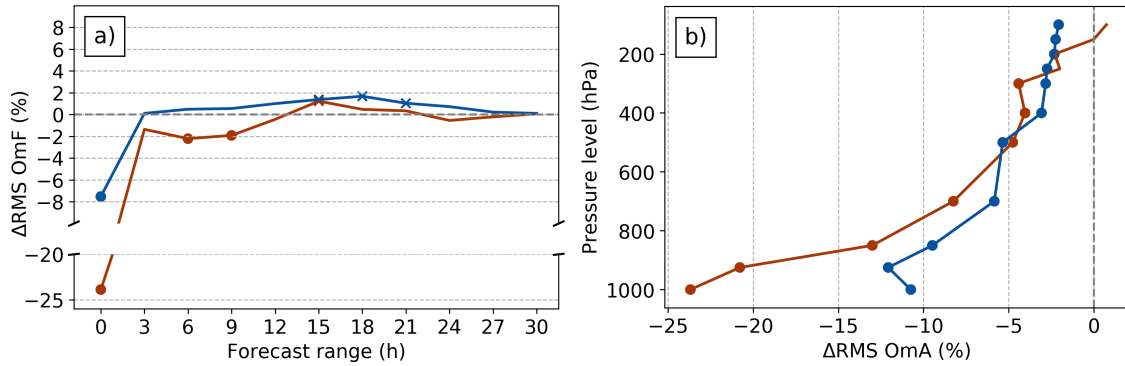
Figure 6. (a) ΔRMS OmF of MSLP for SWSs over France between 0 and 30 h forecast range and (b) ΔRMS OmA of geopotential for radiosoundings launched over France for (blue) 3DVarP w.r.t. 3DVar and (brown) 3DEnVarP w.r.t. 3DEnVar. Points (respectively crosses) indicate improvement (resp. degradation) of XP w.r.t. CTRL at 90 % statistical significance level.

evolution between experiments (as indicated in Eq. 3) are computed. By definition, OmF at 0 h time step is equal to OmA but OmF at 1 h time step is not equal to OmB because of the use of the incremental analysis update technique in AROME forecasts
355 (Brousseau et al., 2016).

At the analysis (0 h), ΔRMS OmF of MSLP for 3DVarP w.r.t. 3DVar is negative and equal to −7.5 %, then rapidly increases during the first hours of the forecast and is positive after only 3 h of forecast (Fig. 6a). For 3DEnVarP w.r.t. 3DEnVar, ΔRMS OmF of MSLP is negative and equal to −24 % at the analysis, also rapidly increases during the first hours of the forecast, but remains negative before 12 h of forecast. This shows that the benefit of the assimilation of PWS observations quickly disappears
360 in 3DVarP, while it remains in the 3DEnVarP experiment. Small but significant degradation is noticeable between 15 to 21 h of forecast for 3DVarP w.r.t. 3DVar, which is less pronounced and not significant for 3DEnVarP w.r.t. 3DEnVar. The influence of the assimilation of PWS observations on other variables near the ground is less pronounced, and no significant results are seen (not shown).

Impact at different levels of the atmosphere can be assessed with the help of radiosoundings (Fig. 6b). Because PWS ob-
365 servations are only assimilated over France or near its borders, only 5 French radiosoundings launched twice a day are taken into consideration. Over the one-month study period, it represents between 220 (at 1000 hPa) and 272 observations. Negative ΔRMS OmA of geopotential near the ground and for the whole troposphere is seen, reaching at 1000 hPa −11 % for 3DVarP w.r.t. 3DVar and −24 % for 3DEnVarP w.r.t 3DEnVar. This shows a substantial improvement of geopotential in the analysis throughout the troposphere when PWS observations are assimilated. Regarding other variables, improvement is seen on the
370 wind heterogeneously for the whole troposphere for 3DVarP w.r.t. 3DVar experiment (not shown), and no significant change could be seen for other radiosounding parameters.

AROME-France being the finer-scale operational model available over France, one of its added values lies in the forecasting of local precipitation events. One common verification score called AROME index (also called IP16) described in Amodei et al. (2015) synthesizes regional Brier skill scores (BSS_NO) of wind gusts and 6 h-accumulated rainfall up to 24 h of forecast. It
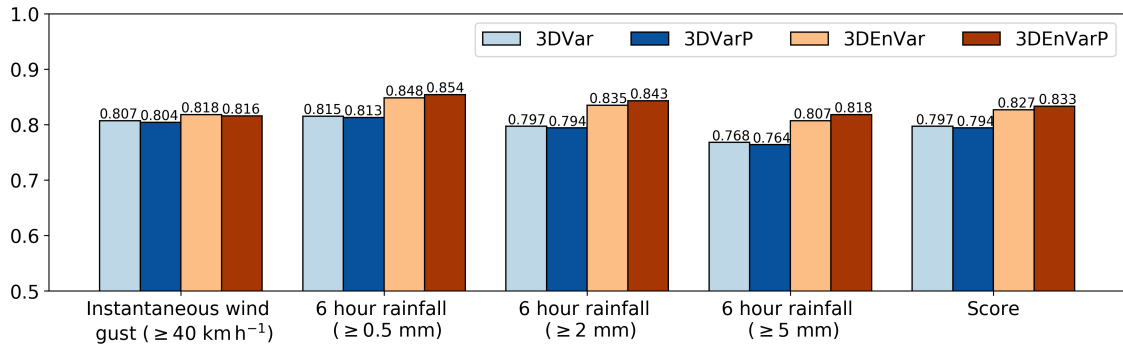
Figure 7. IP16 score of forecasts launched at 00:00 UTC during the one-month study period except the first day (7 September to 5 October 2021) averaged over 4 forecast times (6 h, 12 h, 18 h, 24 h) given by horizontal bars for 3DVar (light blue), 3DVarP (navy blue), 3DEnVar (beige), 3DEnVarP (brown) experiments. All columns correspond to regional Brier skill scores (with a spatial window of 50 km) contributing to IP16: instantaneous wind gusts above $40 \, \mathrm{km \, h^{-1}}$, 6 h-accumulated rainfall above or equal to 0.5 mm, 2 mm, and 5 mm. Higher is better.

is computed for the different experiments and is shown in Figure 7. IP16 increases from 0.797 to 0.827 between 3DVar and 3DEnVar, which is a 3.8 % improvement associated with the change of the DA scheme. The assimilation of PWS geopotential observations decreases IP16 by −0.3 % in 3DVarP with respect to 3DVar, and increases by 0.7 % in 3DEnVarP with respect to 3DEnVar.

When looking in detail at BSS_NO composing the IP16, the largest improvement of BSS in 3DEnVarP w.r.t. 3DEnVar is found for the highest threshold of accumulated rainfall: BSS_NO increases by 0.7 % for the 0.5 mm threshold against 1.3 % for the 5 mm threshold. Improvements in BSS_NO are found particularly between 6 to 18 h of the forecast, and vanish after 24 h (not shown). Results concerning the 14 September event are shown in the next section.

5 Results on the 14 September 2021 heavy precipitation event

5.1 Presentation of the case

The Mediterranean region is regularly affected by heavy precipitation events (HPEs, Doswell et al., 1998; Nuissier et al., 2020; Caumont et al., 2021). On 14 September 2021, a mesoscale convective system (MCS) remained quasi-stationary for 8 h, giving up to 260 mm of rain in 3 h between 07:15 and 10:15 UTC in the city of Saint-Dionisy, in the Gard department (south of France). The particular organization of mesoscale vertical circulations, clouds, hydrometeors, and precipitation inside MCSs can lead to the formation of surface pressure structures which can be observed from the ground.

When PWS surface pressure observations — quality-controlled by MC20 algorithm — and SWS surface pressure observations are combined, surface pressure structures appear near and below the MCS (Fig. 8). At 07:00 UTC a low-pressure anomaly in front of the precipitation system is observed, which could be associated with updrafts, increased relative humidity, and increased temperature; and at the rear high-pressure anomaly, potentially due to downdrafts. These surface pressure anomalies
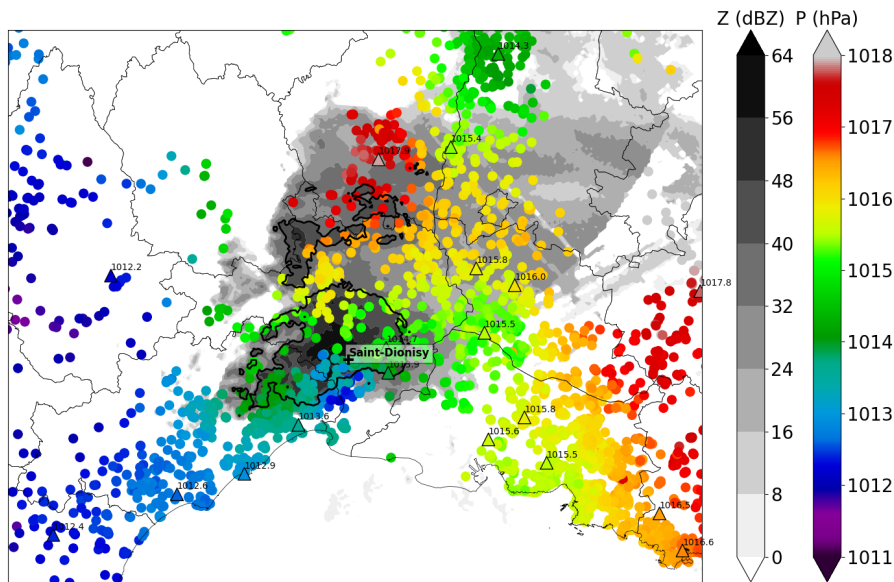
Figure 8. MSLP observations of SWS and PWS networks at 07:00 UTC on 14 September. SWSs are indicated by coloured triangles with black contours, and PWSs by coloured circles. The instantaneous (500 or 250 ms) wind gust during the last minute is shown with barbs (in kt). Two values separated by a slash indicate respectively from left to right the instantaneous wind gust during the last minute and during the last 10 min in km h^{-1}. Base reflectivity (Z) in grey colours indicates thunderstorm activity and location. Reflectivities over 40 dBZ are illustrated with bold black edges.

could not be seen with SWS observations only: therefore, this section examines the impact of assimilating PWS observations
395   of pressure in the area where these anomalies are observed.

5.2 Impact of PWS observations assimilation on surface pressure increments

Figure 9 shows increments of surface pressure for experiments assimilating PWS observations of geopotential (3DVarP, 3DEn-VarP) or not (3DVar, 3DEnVar). Because the four experiments are cycled and starting from much earlier, they have a different background at 07:00 UTC. Since the four experiments start from a different background, this comparison is intended to illus-
400   trate the spatial structure of surface pressure increments and, for each observation, the associated change in analysis departure with respect to background departure. Other observations are assimilated in this area at 07:00 UTC, coming from SWSs, from radars and from an aircraft flying towards Marseille-Provence airport (not shown).

Increments in experiments using a 3DVar DA scheme (Fig. 9a, b) are smoother than increments in experiments using a 3DEnVar DA scheme (Fig. 9c, d). These differences are due to the differences in background error covariance matrices between
405   both schemes. In the 3DVar experiment (Fig. 9a), all pressure observations from SWSs are consistent with the increment signs, which indicate that with their large analysis increment, they are more predominant than other assimilated observations. This
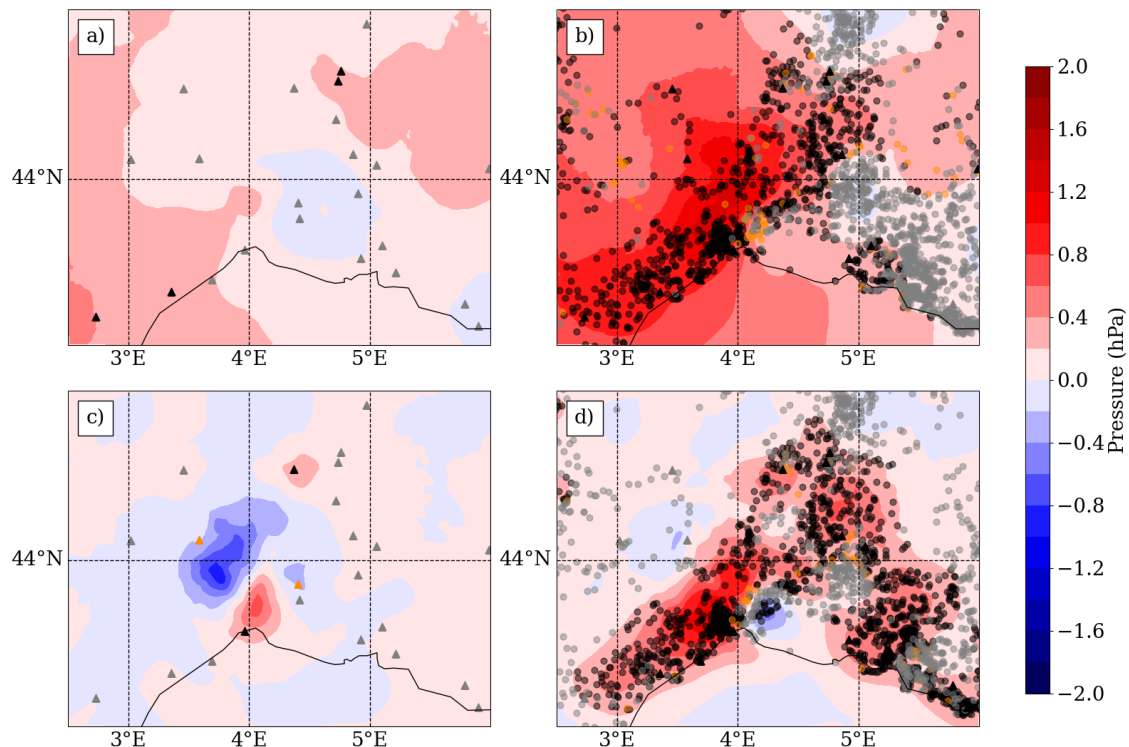
Figure 9. Surface pressure increments (analysis minus background) in (a) 3DVar, (b) 3DVarP, (c) 3DEnVar, and (d) 3DEnVarP experiments on 14 September 2021 at 07:00 UTC. Locations of the stations whose observations of surface pressure are assimilated are indicated by triangles for SWSs and circles for PWSs. Colours of triangles and circles indicate for each surface pressure observation the change in analysis departure (OmA) with respect to background departure (OmB): reduced in black such as $|OmA| - |OmB| < -0.2\,\mathrm{hPa}$, increased in orange such as $|OmA| - |OmB| > 0.2\,\mathrm{hPa}$ or neutral in grey such as $||OmA| - |OmB|| < 0.2\,\mathrm{hPa}$.

is not the case for the 3DEnVar experiment (Fig. 9c), where anomalies appear close to the centre of the MCS. They are not consistent with the 2 closest SWS surface pressure observations. To know from which observations these anomalies were coming, a supplementary experiment has been done rejecting the radar data which is the other type of observation assimilated close to the anomalies; this effectively removes the anomalies (not shown). This confirms that surface pressure observations have a lower impact than other assimilated observations (here the radars) in this specific case with the 3DEnVar scheme.

When assimilating PWS observations, the two assimilation schemes react differently. For the 3DVarP (Fig. 9b), all the positive increments from PWSs add up to form a large positive anomaly. A small group of observations, close to the MCS, are not consistent with this anomaly, which shows the difficulties that the 3DVar DA scheme has to take advantage of the local scale information. However, the 3DEnVarP experiment (Fig. 9d) effectively achieves to push the model according to the OmBs brought by PWS observations. The number of PWS observations assimilated compensate for the small localization given to the surface observations in the 3DEnVar DA scheme. Only a few stations remain at the boundary between positive and
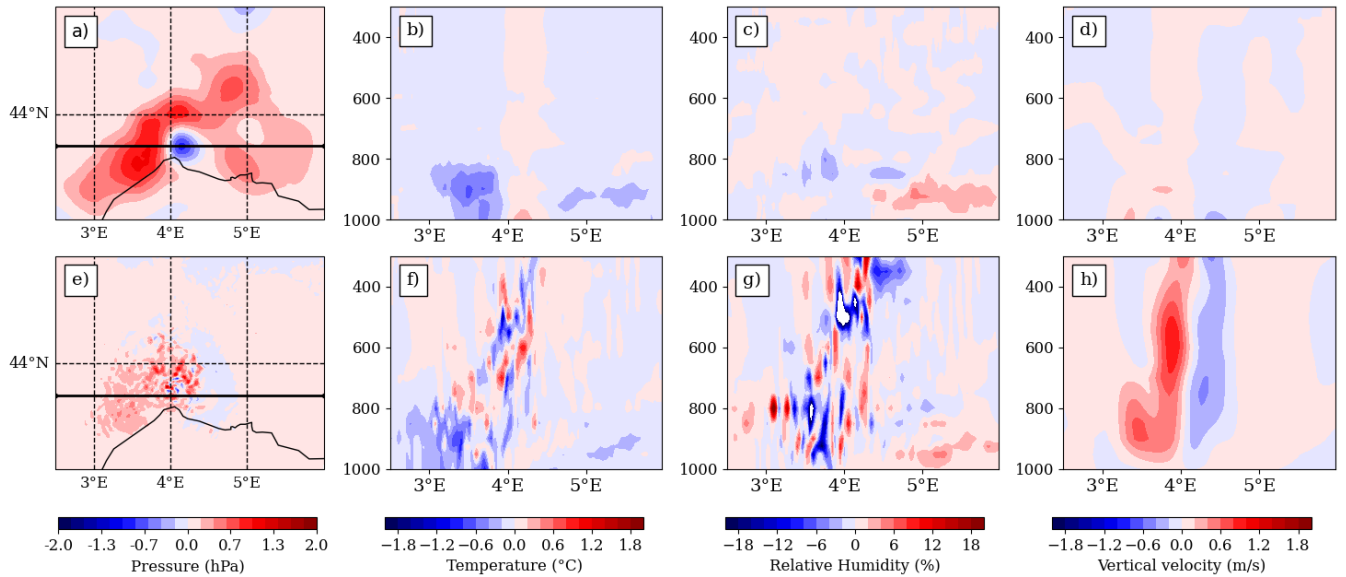
Figure 10. Differences between 3DEnVarP7UTC and 3DEnVar experiments of (a, e) surface pressure, (b, f) temperature, (c, g) relative humidity and (d, h) vertical velocity (a–d) at the analysis of 07:00 UTC and (e–h) at 50 min forecast time. (b–d, f–h) Differences along the vertical cross-section indicated by the thick black line in (a) in pressure coordinates (hPa).

negative anomalies, which do not correspond to the increment. As 3DEnVarP is the only experiment able to reproduce the local anomaly of surface pressure observed by PWSs in the 07:00 UTC analysis, the following section will compare experiments
420   with a 3DEnVar DA scheme.

5.3   Impact on forecasts from a single hour PWS assimilation

Given the pressure increments added by the PWSs with the 3DEnVar scheme, the influence on other variables is explored. In that way, an experiment using the 3DEnVar DA scheme and assimilating PWS observations only at 07:00 UTC was realized (Fig. 10). It is called hereafter 3DEnVarP7UTC. In the AROME DA system, the surface pressure, temperature, specific hu-
425   midity, and the two components of the wind are analysed every hour, while the other prognostic model fields, such as vertical divergence, are only updated during the forecast steps of the assimilation cycles. In consequence, with the help of the 3DEn-Var background covariance error matrix, the PWS geopotential observations induce surface pressure, temperature, and relative humidity changes close to the surface. Indeed, above the positive surface pressure increments, between 3 and 4 °E, a negative temperature difference implying a denser air mass appears (Fig. 10b, c). Vertical velocity, being a diagnostic field, is not up-
430   dated at the analysis and only small anomalies appear due to changes in surface pressure and thus pressure of all grid points on the vertical (Fig. 10d). Regarding the evolution of these anomalies, after a few time steps of forecast, the AROME model balances the information provided by the surface pressure increments into other prognostic fields, and uses them to create new meteorological structures. The surface pressure differences in Fig. 10e have a smaller scale than the analysis increments in

Fig. 10a. Near 4 °E, after 50 min of forecast, a dipole of positive and negative vertical velocity differences appears (Fig. 10h).

435 At that time, multiple fine-scale differences, up to 1 °C in temperature and up to 20 % in relative humidity shows that the location of convective cells inside the MCS is modified. This particular event shows that changes in surface pressure have an effect on the other variables in the model which can rapidly extend all the way to the top of the troposphere. As such storms occur rapidly and over a small geographical area compared with the area of the model domain, the modifications of temperature and relative humidity seen in this figure are barely visible on assimilation experiment scores over a time period of 1 month shown

440 in Sect. 4.

5.4 Impact on predicted rainfall

The MCS remained quasi stationary in the southern part of the Gard department which caused heavy rainfall during the event. The rainfall hourly time series analysed by ANTILOPE, the Météo-France operational precipitation analysis combining radar and rain gauge observations at 1 km grid spacing (Laurantin, 2008), shows the chronology of the event (Fig. 11d). Rainfall

445 started over the area at 00:00 UTC, maximum hourly rainfall continuously increased until 09:00 UTC, and decreased after 09:00 UTC. Between 08:00 and 09:00 UTC, ANTILOPE analysis (Fig. 11c) shows two convective cells, the southern being the most intense with an hourly maximum rainfall accumulation of 144 mm. The corresponding 9 h forecasts starting from the 3DEnVar and 3DEnVarP analyses are shown in Fig. 11a and Fig. 11b. The 3DEnVar experiment forecasts rainfall over the entire Gard department and even north of it, whereas the 3DEnVarP experiment is closer to ANTILOPE for both the spatial extension

450 and the location of the rainfall maximum in the south of the Gard department. Looking at the maximum rainfall accumulation, the 3DEnVarP experiment with 96 mm is closer to the observation (144 mm) than the 3DEnVar experiment, with 65 mm. It can be assumed that the vertical velocity anomalies previously observed could generate more intense precipitation rates. Fractions Skill Scores (FSS, Mittermaier, 2018) show no advantage of 3DEnVarP w.r.t. 3DEnVar for small thresholds of precipitation (e.g., 1 mm). However, over 5 mm thresholds, the FSS is significantly better for the 3DEnVarP experiment (not shown). This

455 means a better location and intensity of the hourly maximum rainfall accumulation of precipitation at this time. In addition, the time series of the hourly rainfall accumulation of the 3DEnVarP experiment is closer to ANTILOPE than the 3DEnVar experiment.

Finally, what forecasters are interested in is the anticipation of the event. Then, the added value of the 12 h forecast launched from the 3DEnVarP analysis at 00:00 UTC on 14 September 2021 with respect to the 12 h forecast launched from the 3DEn-

460 Var analysis at the same time is explored. The maximum rainfall accumulation observed during this period is 320 mm. The maximum rainfall accumulation of the 12 h forecast launched from the 3DEnVarP analysis is 221 mm, which is 22 % higher than the 3DEnVar one (180 mm). Looking at the FSS, small but significant degradation is seen at the 2 mm threshold, whereas significant improvement can be seen at higher thresholds, especially at 50 and 100 mm (Fig. 12). In this case, the FSS of the 12 h forecast at 100 mm threshold is significantly improved at a window size of 1° in latitude and longitude. This could be

465 explained by more realistic analyses a few hours before the forecast launch, thanks to the assimilation of PWS surface pressure observations.
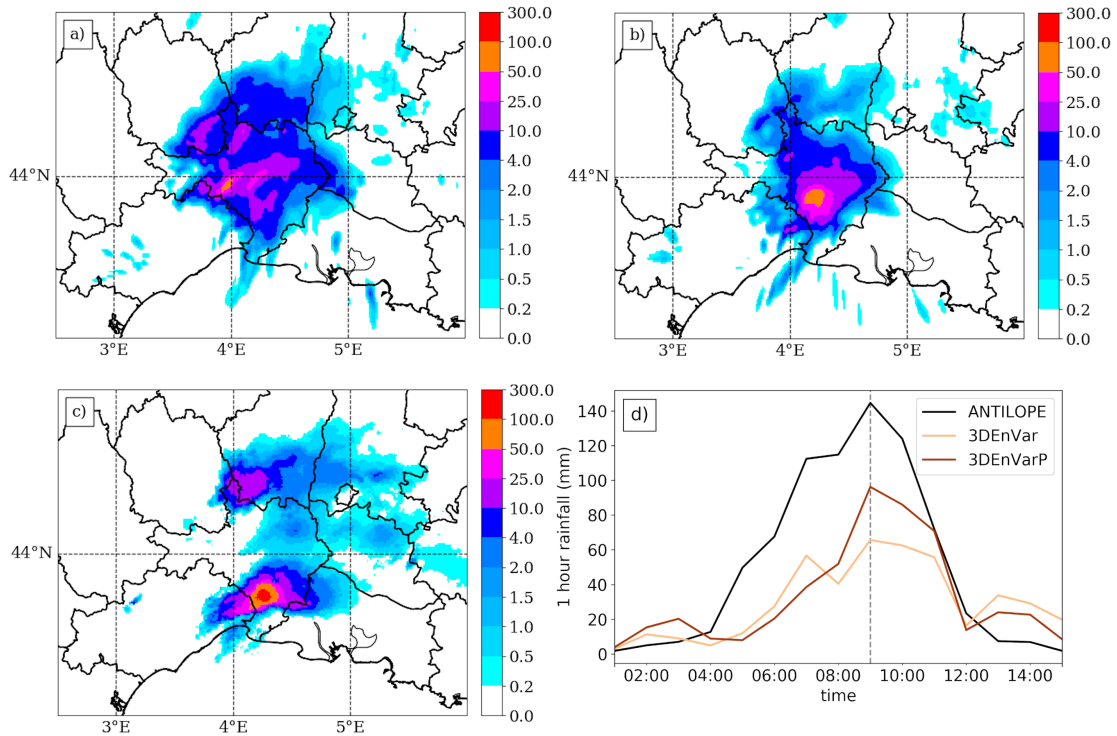
Figure 11. Rainfall accumulation (mm) on 14 September 2021 between 08:00 and 09:00 UTC of forecasts starting from the (a) 3DEnVar and (b) 3DEnVarP analyses at 08:00 UTC on 14 September 2021 and (c) analysed by ANTILOPE. (d) Time series of hourly maximum rainfall over the area of forecasts starting from the (light brown) 3DEnVar and (brown) 3DEnVarP analyses and (black) analysed by ANTILOPE.

6   Discussion and conclusion

This study has examined the impact of the assimilation of spatially dense PWS surface pressure observations with the 3DVar and 3DEnVar data assimilation schemes of AROME-France in one-month cycled experiments. A procedure has been presented in order to correct and select the observation dataset for purpose of assimilation.

The usability of surface pressure observations is related to the accuracy of their altitude metadata. When assimilated in a non-cycled AROME 3DVar experiment (the Monitor experiment), abnormal systematic differences in the OmB statistics were shown in the pressure time series of 30 individual stations around Toulouse. The model topography is different from the true one, and the OmB statistics combine the error from the observation and the error from the model. The choice was made to correct the altitude of the PWS stations to the altitude they would have in the model. The station dependent bias has been found by averaging the OmBs from the Monitor experiment during one rolling month. The corrected observations have been used for both 3DVar and 3DEnVar experiments. The bias correction substantially reduces the range of individual geopotential bias from $\pm 1000\,\mathrm{m^2\,s^{-2}}$ to $\pm 15\,\mathrm{m^2\,s^{-2}}$.
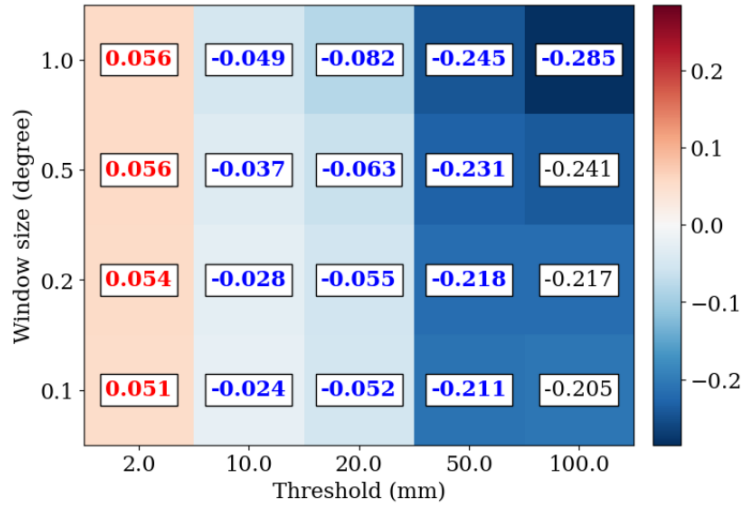
Figure 12. Difference in FSS of 12 h precipitation forecasts starting from the 3DEnVar and 3DEnVarP analyses at 00:00 UTC on 14 September 2021. Difference is computed as FSS(3DEnVarP)−FSS(3DEnVar). FSS is computed for 2, 10, 20, 50, and 100 mm thresholds and spatial neighbourhoods of 0.1, 0.2, 0.5, and 1° in latitude and longitude. Significant positive (resp. negative) values are in red (resp. blue).

Then, unrealistic pressure observations were filtered. This was done by removing observations when their time series were too sparse; and when the standard deviation of the one-month geopotential departure between the PWS and the model was larger than $60\,\mathrm{m^2\,s^{-2}}$. This method ensures that observations belonging to time series with abnormal variations, due in particular to sudden changes in altitude or miscalibration, are removed.

The last pre-processing step was a thinning. With the use of the spatial Desroziers diagnostic, the set of observations was found to have limited spatial correlation of observation error. PWS observations were thinned by keeping one observation every 1.3 km in order to avoid (1) giving too much influence to PWSs observations when they are dense compared to conventional observing systems, which can make local minimization of the cost function difficult and ultimately deteriorate the forecast (Eq. 1) and (2) propagating possible correlated observation errors in space, even if their existence has not been proven. After the preparation of the data, the bias correction, the quality control, and the thinning, there are an average of 27 578 pressure observations per hour ready to be assimilated, which represents 51 % of the initial observations.

In order to improve the pre-processing, there are several possibilities. Firstly, with regard to the bias correction, which can induce a slight bias of $-2.6\,\mathrm{m^2\,s^{-2}}$ on average. The bias correction uses the one-hour forecast of the Monitor experiment, which may be biased. Investigations over a longer period are needed to fully understand this bias. Another limitation of this method may lie in the consistency of assimilating into a 3DEnVar experiment PWS observations that have been bias-corrected thanks to OmB time series from the non-cycled 3DVar experiment. Other results may be expected using a bias correction which uses OmB time series from a non-cycled 3DEnVar experiment. In an operational context, a VarBC method could be a convenient

and more consistent solution to routinely correct the bias of PWS observations, while being aware of the need to maintain sufficient anchor observations from SWSs (Eyre, 2016).

Concerning the quality control (QC), the method used here sometimes suffers from not removing a wrong PWS surface pressure observation whose time series undergo a small but abnormal pressure variation (e.g. change of spatial position without any change in metadata). If such abnormal pressure variation persists over time, then the monthly rolling standard deviation of the OmBs will keep growing hour by hour, and it will eventually exceed the rejection threshold. The QC will take some time to remove the wrong stations. More sophisticated QC could be done to remove erroneous observations faster and more accurately, with the help of the neighbours' variations. Indeed, a simple buddy check (as used by Vasiljevic et al., 2006) could help, while keeping in mind its difficulties to make the difference between erroneous variations and local physical patterns (as the one seen during the 14 September HPE). Finally, a more advanced technique, taking into account a temporal window to remove stations that are unrealistic during a few time steps (as MC20 method) could be beneficial. In the end, the supplementary QC presented here removed 1.6 % more observations, which is the same order of magnitude as the leave-one-out cross validation method presented by MC20. More investigation could be done on the utility of such method, moreover, if a superobbing method is tested in place of the thinning one. The necessity of the thinning procedure is not questioned in this study. A dense observational network can be detrimental in the case of correlated observations errors and cause convergence problems and imbalanced fields (Pourret et al., 2021; Brousseau et al., 2016). Therefore, we applied a thinning at 1.3 km scale which reduced those effects. However, the spatial Desroziers diagnostics has shown small observation correlation errors for the first 1 km distance bin. Experiments without thinning could be done in the future to understand how the assimilation scheme reacts when assimilating pressure observations from PWS in dense observations areas, like in the cities.

Making one-month cycled experiments has allowed us to see the overall impact of the assimilation of dense PWS surface pressure observations on analyses and forecasts of the AROME NWP system. When looking at statistics of geopotential OmB for SWS observations, the RMS OmB does not diminish when PWS observations are assimilated in the 3DVar DA scheme, due to systematic increases in the mean geopotential OmB for SWS observations. When removing the bias from the geopotential time series of individual SWSs (Appendix B), as it has been done previously for PWSs, the mean and the standard deviation of the OmB for SWSs are not degraded when assimilating PWS geopotential observations. This indicates that new methods to detect and reduce observational biases of SWS geopotential observations, bearing in mind that these anchoring observations are the link to reality, could allow to take advantage of PWS observations with AROME current 3DVar DA scheme. On the contrary, when assimilating the new observations with the 3DEnVar DA scheme, the OmB RMS diminish up to −10.3 % for geopotential observations from SWSs situated in France.

To extend this study, the impact of assimilating PWS surface pressure observations on the quality of analyses could be studied with the Degrees of Freedom for Signal diagnostic (Chapnik et al., 2006). If the added value of assimilating surface pressure observation appears in local areas, further research should be conducted there.

Finally, the impact of the assimilation of a dense network of surface pressure observations related to a high precipitation event has been investigated. The event has been chosen due to its high intensity of 260 mm of rainfall in 3 h observed in the southern part of the Gard department and its associated local pressure variation. While the 3DVar has difficulties in reproducing

the local pressure features, the 3DEnVar was found to create reduced length scale surface pressure increments at the location where the PWS network observed the anomalies. This reduction of length scale from 200 km in 3DVar to 40 km in 3DEnVar for a single observation is illustrated in Appendix A. The surface pressure increment is able to modify temperature, relative humidity and vertical velocity fields 50 min after the analysis. Then, the impact on rainfall has been studied by comparing forecasts to quantitative precipitation estimation products merging observations. The forecast launched from the 3DEnVar experiment assimilating PWS observations approximately 8 h before the most intense rainfall, improves both the location and the intensity of the rainfall maximum compared to a forecast launched from the 3DEnVar experiment without assimilating PWS observations. The maximum rainfall accumulation is increased by 22 % in the better forecast, which is closer to the observations. Still, this maximum rainfall accumulation remains 30 % lower from the observed one, underlining the need to further improve both analyses and the AROME model. Further research on other case studies is needed to confirm the positive results.

For operational use, the flows and storage space for this data, as well as the bias correction algorithms and quality control, must be fully implemented in the model's operational environment. Discussions may be held in the meantime to pool some of this work with other national meteorological services. AROME-France's 3DEnVar DA system is currently being tested at Météo-France. The assimilation of PWS observations could therefore be included in a subsequent AROME-France test chain, controlled over several months.

Appendix A: Single-observation assimilation experiments

Single-observations assimilation experiments are performed to describe the ability of a DA scheme to propagate the OmB from an observation. In Figure A1, one observation of geopotential in Blagnac has been assimilated, and all the other observations have been removed. The increment is quasi isotropic in 3DVar (Fig. A1a), while slightly northward in 3DEnVar (Fig. A1b) at this time. The horizontal propagation of the increment on a longitudinal cross-section at the observation latitude is given by Fig. A1c,d. With the 3DVar scheme, the propagation of the increment goes up to 200 km, i.e where the increment is reduced by 99 % of the OmB. With the 3DEnVar scheme, with the localization at 25 km, the propagation goes up to 40 km.

Appendix B: Sensitivity of the Desroziers diagnostic to a geopotential bias correction

In a NWP system, because SWSs observations are, as said by Ingleby (2015), our link to reality, they are not bias-corrected. Thus, the mean geopotential OmA for SWSs ranges between 4.7 and $11.9\,\mathrm{m^2\,s^{-2}}$ for 3DVar and 3DVarP experiments, respectively. In this article, however, PWS observations are bias-corrected before the assimilation. As a consequence, the mean geopotential OmA for PWSs is $-0.1\,\mathrm{m^2\,s^{-2}}$ in the 3DVarP experiment (Table 3) and the Desroziers diagnostic is lower for PWSs than for SWSs. To study the sensitivity of the bias correction on this diagnostic, an experiment in which SWS are bias-corrected was realized.
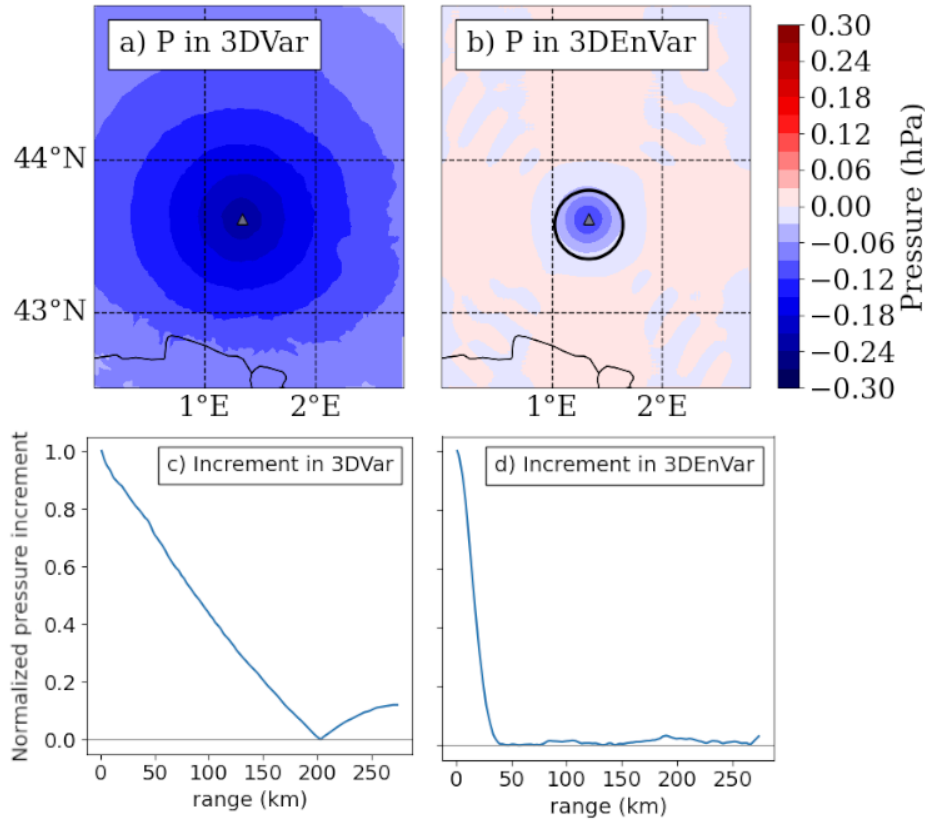
Figure A1. (a, b) Analysis increment of surface pressure resulting from the assimilation of a single observation of pressure at 1 m height in Blagnac, France and (c, d) amplitude of the longitudinal surface pressure analysis increment divided by its maximum value called normalized pressure increment (a, c) in an idealized 3DVar experiment and (b, d) in an idealized 3DEnVar experiment starting from the same guess. The black circle indicates a 25 km distance around the observation location.

In this experiment called 3DVarP8corr (Table B1), SWS geopotential observations are assimilated after being corrected by subtracting the mean geopotential OmB computed on all time steps of one rolling month given by the Monitor experiment, similarly as for PWSs. The 3DVarP8corr experiment was run until the statistics, given in Table B2, were found to be stable over time, which took 8 days.

In 3DVarP8corr, the mean OmA for SWSs become close to zero (i.e. $-0.2\,\mathrm{m^2\,s^{-2}}$) as wished. Even if the mean OmB for SWS are slightly diverging from the OmA statistics, the estimated Desroziers diagnostic significantly decreases from $25.0\,\mathrm{m^2\,s^{-2}}$ for 3DVarP8 to $10.7\,\mathrm{m^2\,s^{-2}}$ for 3DVarP8corr. The PWS bias correction explains why the Desroziers diagnostic is lower for PWSs than for SWSs in 3DVarP and 3DEnVarP experiments. This result confirms that the diagnostic cannot be used to compare the quality of two observing systems whose observations have not been corrected similarly.

Table B1. Overview of supplementary experiments run during 8 days between 6 September 2021 00:00 UTC and 14 September 2021 00:00 UTC.

Experiments	Duration	Cycling	DA scheme	Use of PWS surface pressure observations	Use of SWS surface pressure observations
3DVarP8	8 d	Yes	3DVar	Assimilated after BC, QC and thinning	Assimilated without BC
3DVarP8corr					Assimilated after BC

Table B2. As Table 3 but for 3DVarP8 and 3DVarP8corr experiments. SWSs (resp. PWSs) provide 182 (resp. 27 997) observations per hour over 215 time steps.

	3DVarP8				3DVarP8corr			
	SWS		PWS		SWS		PWS	
	mean	std	mean	std	mean	std	mean	std
OmA	11.4	21.6	−0.1	13.0	−0.2	8.8	0.0	12.9
OmB	10.4	40.2	−0.6	34.6	−1.9	35.1	−0.8	34.8
$\widetilde{\sigma_O}$	25.0		13.9		10.7		13.9	

Author contributions. Alan Demortier: Formal investigations, methodology, analysis; resources; writing – original draft; writing – review and editing. Marc Mandement: Formal analysis; resources; writing – review and editing. Vivien Pourret: Formal analysis; software; validation; supervision; writing – review and editing. Olivier Caumont: validation; project administration; supervision; validation; writing – review and editing.

Competing interests. The authors declare that they have no conflict of interest.

Code availability. The code used for the assimilation experiments in AROME-France is owned by the members of the ACCORD consortium. This agreement allows each member of the consortium to license the shared ACCORD codes to academic institutions in their home countries for non-commercial research. Access to ACCORD codes, and codes used for the figures, can be obtained by contacting the corresponding author.

Data availability. World Meteorological Organization essential weather station data (SYNOP and SHIP reports), the radar mosaic at a 15 min time step, and AROME analyses and forecasts are available freely, online, in real time at https://donneespubliques.meteofrance.fr/ (last access: 15 June 2023); archives of RADOME data, radar mosaic data, and AROME analyses and forecasts are available offline by making a request via this website. Personal weather station data are provided by and property of Netatmo and cannot be shared according to their

licence: https://dev.netatmo.com/legal (last access: 15 June 2023). ANTILOPE QPE and the results of data assimilation experiments are available on request from the corresponding author solely for non-commercial research purposes and for reasonable data volumes.

585 *Acknowledgements.* This work was supported by the French National program Les Enveloppes Fluides et l'Environnement (LEFE), project ASMA.

References

Amodei, M., Sanchez, I., and Stein, J.: Verification of the French operational high-resolution model AROME with the regional Brier probability score, Meteorological Applications, 22, 731–745, https://doi.org/10.1002/met.1510, 2015.

590 Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K.-I., Lenderink, G., Niemelä, S., Nielsen, K. P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D. S., Subias, A., Tijm, S., Toll, V., Yang, X., and Ødegaard Køltzow, M.: The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System, Monthly Weather Review, 145, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1, 2017.

Benáček, P., Brožková, R., Bučánek, A., and Trojáková, A.: Tuning of 3D-Var and ALADIN-CZ system for aircraft data and assimilation,
595 https://www.rclace.eu/media/files/Data_Assimilation/2016/rep_Patrik_ModeS.pdf (last access: 15 June 2023), 2016.

Brousseau, P., Desroziers, G., Bouttier, F., and Chapnik, B.: A posteriori diagnostics of the impact of observations on the AROME-France convective-scale data assimilation system, Quarterly Journal of the Royal Meteorological Society, 140, 982–994, https://doi.org/10.1002/qj.2179, 2014.

Brousseau, P., Seity, Y., Ricard, D., and Léger, J.: Improvement of the forecast of convective activity from the AROME-France system,
600 Quarterly Journal of the Royal Meteorological Society, 142, 2231–2243, https://doi.org/10.1002/qj.2822, 2016.

Caumont, O., Mandement, M., Bouttier, F., Eeckman, J., Lebeaupin Brossier, C., Lovat, A., Nuissier, O., and Laurantin, O.: The heavy precipitation event of 14–15 October 2018 in the Aude catchment: a meteorological study based on operational numerical weather prediction systems and standard and personal observations, Natural Hazards and Earth System Sciences, 21, 1135–1157, https://doi.org/10.5194/nhess-21-1135-2021, 2021.

605 Chambon, P., Mahfouf, J.-F., Audouin, O., Birman, C., Fourrié, N., Loo, C., Martet, M., Moll, P., Payan, C., Pourret, V., and Raspaud, D.: Global Observing System Experiments within the Météo-France 4D-Var Data Assimilation System, Monthly Weather Review, 151, 127–143, https://doi.org/10.1175/MWR-D-22-0087.1, 2023.

Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O.: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting, Quarterly Journal of the Royal Meteorological Society, 132, 543–565, https://doi.org/10.1256/qj.04.102, 2006.

610 Clark, M. R., Webb, J. D. C., and Kirk, P. J.: Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations, Meteorological Applications, 25, 472–492, https://doi.org/10.1002/met.1715, 2018.

Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., and Pailleux, J.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, Quarterly Journal of the Royal Meteorological Society, 124, 1783–1807, https://doi.org/10.1002/qj.49712455002, 1998.

615 Dando, M. L., Thorpe, A. J., and Eyre, J. R.: The impact of targeted satellite observations on numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 133, 1945–1957, https://doi.org/10.1002/qj.176, 2007.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/10.1256/qj.05.108, 2005.

Doswell, III, C. A., Ramis, C., Romero, R., and Alonso, S.: A Diagnostic Study of Three Heavy Precipitation Episodes in the Western
620 Mediterranean Region, Weather and Forecasting, 13, 102–124, https://doi.org/10.1175/1520-0434(1998)013<0102:ADSOTH>2.0.CO;2, 1998.

ECMWF: IFS Documentation CY48R1 - Part I: Observations, 1, ECMWF, https://doi.org/10.21957/0f360ba4ca, 2023.

Eyre, J. R.: Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties, Quarterly Journal of the Royal Meteorological Society, 142, 2284–2291, https://doi.org/10.1002/qj.2819, 2016.

625 Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissmann, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., Mile, M., Hamdi, R., Lindskog, M., Barkmeijer, J., Dahlbom, M., Macpherson, B., Ballard, S., Inverarity, G., Carley, J., Alexander, C., Dowell, D., Liu, S., Ikuta, Y., and Fujita, T.: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres, Quarterly Journal of the Royal Meteorological Society, 144, 1218–1256, https://doi.org/10.1002/qj.3179, 2018.

Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensem-
630 ble Kalman Filter, Monthly Weather Review, 129, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2, 2001.

Hintz, K. S., Vedel, H., and Kaas, E.: Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation, Meteorological Applications, 26, 733–746, https://doi.org/10.1002/met.1805, 2019.

Hu, G., Dance, S. L., Bannister, R. N., Chipilski, H. G., Guillet, O., Macpherson, B., Weissmann, M., and Yussouf, N.: Progress, challenges,
635 and future steps in data assimilation for convection-permitting numerical weather prediction: Report on the virtual meeting held on 10 and 12 November 2021, Atmospheric Science Letters, https://doi.org/10.1002/asl.1130, 2022.

Ingleby, B.: Global assimilation of air temperature, humidity, wind and pressure from surface stations, Quarterly Journal of the Royal Meteorological Society, 141, 504–517, https://doi.org/10.1002/qj.2372, 2015.

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., Leriche, M., Barthe, C., Aouizerats, B., Augros, C., Aumond, P.,
640 Auguste, F., Bechtold, P., Berthet, S., Bielli, S., Bosseur, F., Caumont, O., Cohard, J.-M., Colin, J., Couvreux, F., Cuxart, J., Delautier, G., Dauhut, T., Ducrocq, V., Filippi, J.-B., Gazen, D., Geoffroy, O., Gheusi, F., Honnert, R., Lafore, J.-P., Brossier, C. L., Libois, Q., Lunet, T., Mari, C., Maric, T., Mascart, P., Mogé, M., Molinié, G., Nuissier, O., Pantillon, F., Peyrillé, P., Pergaud, J., Perraud, E., Pianezze, J., Redelsperger, J.-L., Ricard, D., Richard, E., Riette, S., Rodier, Q., Schoetter, R., Seyfried, L., Stein, J., Suhre, K., Taufour, M., Thouron, O., Turner, S., Verrelle, A., Vié, B., Visentin, F., Vionnet, V., and Wautelet, P.: Overview of the Meso-NH model version 5.4 and its
645 applications, Geoscientific Model Development, 11, 1929–1969, https://doi.org/10.5194/gmd-11-1929-2018, 2018.

Laurantin, O.: ANTILOPE: Hourly rainfall analysis merging radar and rain gauge data, in: Proceedings of the International Symposium on Weather Radar and Hydrology, pp. 2–8, 2008.

Liu, Z.-Q. and Rabier, F.: The potential of high-density observations for numerical weather prediction: A study with simulated observations, Quarterly Journal of the Royal Meteorological Society, 129, 3013–3035, https://doi.org/10.1256/qj.02.170, 2003.

650 Madaus, L. E. and Hakim, G. J.: Observable Surface Anomalies Preceding Simulated Isolated Convective Initiation, Monthly Weather Review, 144, 2265–2284, https://doi.org/10.1175/MWR-D-15-0332.1, 2016.

Madaus, L. E., Hakim, G. J., and Mass, C. F.: Utility of Dense Pressure Observations for Improving Mesoscale Analyses and Forecasts, Monthly Weather Review, 142, 2398–2413, https://doi.org/10.1175/MWR-D-13-00269.1, 2014.

Mandement, M. and Caumont, O.: Contribution of personal weather stations to the observation of deep-convection features near the ground,
655 Natural Hazards and Earth System Sciences, 20, 299–322, https://doi.org/10.5194/nhess-20-299-2020, 2020.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The

660   SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

McNicholas, C. and Mass, C. F.: Smartphone Pressure Collection and Bias Correction Using Machine Learning, Journal of Atmospheric and Oceanic Technology, 35, 523–540, https://doi.org/10.1175/jtech-d-17-0096.1, 2018a.

McNicholas, C. and Mass, C. F.: Impacts of Assimilating Smartphone Pressure Observations on Forecast Skill during Two Case Studies in

665   the Pacific Northwest, Weather and Forecasting, 33, 1375–1396, https://doi.org/10.1175/waf-d-18-0085.1, 2018b.

Meier, F., Fenner, D., Grassmann, T., Otto, M., and Scherer, D.: Crowdsourcing air temperature from citizen weather stations for urban climate research, Urban Climate, 19, 170–191, https://doi.org/10.1016/j.uclim.2017.01.006, 2017.

Michel, Y. and Brousseau, P.: A Square-Root, Dual-Resolution 3DEnVar for the AROME Model: Formulation and Evaluation on a Summer-time Convective Period, Monthly Weather Review, 149, 3135–3153, https://doi.org/10.1175/MWR-D-21-0026.1, 2021.

670   Mile, M., Benáček, P., and Rózsa, S.: The use of GNSS zenith total delays in operational AROME/Hungary 3D-Var over a central European domain, Atmospheric Measurement Techniques, 12, 1569–1579, https://doi.org/10.5194/amt-12-1569-2019, 2019.

Mittermaier, M. P.: A "Meta" Analysis of the Fractions Skill Score: The Limiting Case and Implications for Aggregation, Monthly Weather Review, 149, 3491–3504, https://doi.org/10.1175/MWR-D-18-0106.1, 2018.

Montmerle, T., Michel, Y., Arbogast, E., Ménétrier, B., and Brousseau, P.: A 3D ensemble variational data assimilation scheme for the

675   limited-area AROME model: Formulation and preliminary results, Quarterly Journal of the Royal Meteorological Society, 144, 2196–2215, https://doi.org/10.1002/qj.3334, 2018.

Netatmo: Smart weather station technical specifications, https://www.netatmo.com/en-gb/smart-weather-station/specifications (last access: 29 November 2023), 2023.

Nipen, T. N., Seierstad, I. A., Lussana, C., Kristiansen, J., and Hov, Ø.: Adopting Citizen Observations in Operational Weather Prediction,

680   Bulletin of the American Meteorological Society, 101, E43–E57, https://doi.org/10.1175/BAMS-D-18-0237.1, 2020.

Nuissier, O., Duffourg, F., Martinet, M., Ducrocq, V., and Lac, C.: Hectometric-scale simulations of a Mediterranean heavy-precipitation event during the Hydrological cycle in the Mediterranean Experiment (HyMeX) first Special Observation Period (SOP1), Atmospheric Chemistry and Physics, 20, 14 649–14 667, https://doi.org/10.5194/acp-20-14649-2020, 2020.

Pauley, P. M. and Ingleby, B.: Assimilation of In-Situ Observations, pp. 293–371, Springer International Publishing, Cham,

685   https://doi.org/10.1007/978-3-030-77722-7_12, 2022.

Pourret, V., Mahfouf, J.-F., Guidard, V., Moll, P., Doerenbecher, A., and Piguet, B.: Variational bias correction for Mode-S aircraft derived winds, Tellus A: Dynamic Meteorology and Oceanography, 73, 1886 808, https://doi.org/10.1080/16000870.2021.1886808, 2021.

Pourret, V., Šavli, M., Mahfouf, J.-F., Raspaud, D., Doerenbecher, A., Bénichou, H., and Payan, C.: Operational assimilation of Aeolus winds in the Météo-France global NWP model ARPEGE, Quarterly Journal of the Royal Meteorological Society, 148, 2652–2671,

690   https://doi.org/10.1002/qj.4329, 2022.

Ridal, M., Bojarova, J., Hagelin, S., Medeiros, P., Gregow, E., and Stappers, R.: Study of forecast quality improvement (magnitude and duration), relative to using a reference set of observations, from assimilation of novel observations, https://www.ecmwf.int/sites/default/files/special_projects/2020/spnolabo-2020-finalreport.pdf (last access: 15 June 2023), 2019.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France Convective-Scale

695   Operational Model, Monthly Weather Review, 139, 976–991, https://doi.org/10.1175/2010MWR3425.1, 2011.

Sgoff, C., Acevedo, W., Paschalidi, Z., Ulbrich, S., Bauernschubert, E., Kratzsch, T., and Potthast, R.: Assimilation of crowd-sourced surface observations over Germany in a regional weather prediction system, Quarterly Journal of the Royal Meteorological Society, 148, 1752–1767, https://doi.org/10.1002/qj.4276, 2022.

STMicroelectronics: MEMS pressure sensor: 260-1260 hPa absolute digital output barometer, https://www.st.com/en/mems-and-sensors/lps25hb.html#documentation (last access: 15 June 2023), rev. 4, 2016.

Tardieu, J. and Leroy, M.: Radome, le réseau temps réel d'observation au sol de Météo-France, La Météorologie, pp. 40–43, https://doi.org/10.4267/2042/36262, 2003.

Vasiljevic, D., Andersson, E., Isaksen, L., and Garcia-Mendez, A.: Surface pressure bias correction in data assimilation, Tech. rep., ECMWF, https://doi.org/10.21957/UV295RFMX5, 2006.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Wattrelot, E., Montmerle, T., and Guerrero, C. G.: Evolution of the assimilation of radar data in the AROME model at convective scale, http://www.meteo.fr/cic/meetings/2012/ERAD/extended_abs/NWP_401_ext_abs.pdf (last access: 15 June 2023), 2012.

World Meteorological Organization: Manual on the Global Telecommunication System: Annex III to the WMO Technical Regulations (WMO-No. 386), https://library.wmo.int/doc_num.php?explnum_id=10469 (last access: 15 June 2023), 2020.

World Meteorological Organization: Manual on the WMO Integrated Global Observing System: Annex VIII to the WMO Technical Regulations (WMO-No. 1160), https://community.wmo.int/en/activity-areas/wigos/gbon (last access: 31 October 2023), 2021.