
Validation of AltiKa Matching Pursuit Rain Flag

Tournadre Jean ^{1,*}, Poisson J. C. ², Steunou N. ³, Picard B. ²

¹ IFREMER, Lab Oceanog Spatiale, F-29280 Plouzane, France.

² Space Oceanog Div, Collecte Localisat Satellite, St Agne, France.

³ Ctr Natl Etud Spatiales, F-31055 Toulouse, France.

* Corresponding author : Jean Tournadre, email address : jean.tournadre@ifremer.fr

Abstract :

The major drawback of Ka band, operating frequency of the AltiKa altimeter on board SARAL, is its sensitivity to atmospheric liquid water. Even light rain or heavy clouds can strongly attenuate the signal and distort the signal leading to erroneous geophysical parameters estimates. A good detection of the samples affected by atmospheric liquid water is crucial. As AltiKa operates at a single frequency, a new technique based on the detection by a Matching Pursuit algorithm of short scale variations of the slope of the echo waveform plateau has been developed and implemented prelaunch in the ground segment. As the parameterization of the detection algorithm was defined using Jason-1 data, the parameters were re-estimated during the cal-val phase, during which the algorithm was also updated. The measured sensor signal-to-noise ratio is significantly better than planned, the data loss due to attenuation by rain is significantly smaller than expected (<0.1%). For cycles 2 to 9, the flag detects about 9% of 1Hz data, 5.5% as rainy and 3.5 % as backscatter bloom (or sigma0 bloom). The results of the flagging process are compared to independent rain data from microwave radiometers to evaluate its performances in term of detection and false alarms.

1 Introduction

AltiKa on board SARAL, launched in February 2013, is an altimeter developed by the French Centre National d'Etudes Spatiales (CNES). It is a wide-band Ka-band altimeter (35.75 GHz, 500 MHz bandwidth). It is the first oceanography altimeter to operate at such a high frequency. This unique technical characteristic of the instrument offers improvements both in terms of spatial and vertical resolution (Vincent et al., 2006). This improved accuracy should lead to improved observation of ice, coastal areas, inland waters and wave height.

The one major drawback of Ka band is its high sensitivity to atmospheric liquid water (both rain and clouds), ten times larger than that at Ku band used by past altimeters. Studies of Ka band altimeter have shown that even light rain or heavy clouds can strongly attenuate the radar signal, distort the altimeter echo waveform and hamper the retrieval of geophysical parameters (Tournadre, 1999; Tournadre et al., 2009a). It is thus necessary to detect and flag the samples potentially affected by atmospheric liquid water. The same problem was encountered for Ku-band altimeters, such as Topex, Jason-1 or Envisat, for medium and heavy rain. For these dual frequency altimeters rain flags based on the differential attenuation of the

main (Ku-band) and secondary channels (C-band for Topex and Jason-1 and S-band for Envisat) by rain droplets, were defined and are currently used operationally (Tournadre and Morland, 1997; Tournadre, 2004; Tran et al., 2005; Quartly et al., 1996). AltiKa being a single frequency altimeter and this kind of simple and efficient flag could obviously not be utilized. A new rain flag based on the analysis of the altimeter measurement alone has thus been developed for AltiKa. Based on past experience with Ku-band altimeter data and on the modeling of AltiKa echo waveform in presence of cloud and rain that showed that rain cells or clouds are characterized by sharp coherent along-track variations of the slope of the echo waveform plateau (characterized by the off-nadir angle estimate) (Tournadre, 1998; Tournadre et al., 2009b), a method of identification, detection and localization of these along-track transient features has been developed. The flagging algorithm is based on the analysis of the short scale variations of the off-nadir angle estimate by a Matching Pursuit (MP) algorithm. MP allows the decomposition of a signal into a few salient features or atoms chosen from a dictionary of elementary functions defined by the wavelet decomposition of the signal (Tournadre et al., 2009b). The flag has been tested pre-launch on simulated Ka-band waveforms as well as on real Jason-1 Ku-band rain-data affected. The tests on Jason-1 data were also used to define the configuration parameters of the pre-launch MP algorithm: the number of atoms and the mean noise level of the off-nadir angle. Using this parameterization, the pre-launch algorithm did not perform well. Taking advantage of the first reprocessing of AltiKa Geophysical Data Record (GDR) products in the beginning of 2014, the parameters have been updated and combined with some minor changes in the algorithm to account for the AltiKa ground segment characteristics. The updated algorithm is used to flag the first 9 cycles of AltiKa and the results are validated using external rain rates data from microwave radiometers. The impact of rain on the data potentially affected is estimated using the 1 Hz root mean square of the geophysical parameters.

The first section presents the Matching Pursuit detection algorithm as well as the new operational flag. The collocated rain data necessary to the flag validation are presented in section 2 and the effect of rain on AltiKa echo waveforms and geophysical parameters estimates in section 3. An example of the impact of rain on AltiKa waveform and geophysical parameters' estimates is presented in section 4. The results of the flagging process and their validation against independent rain data are analyzed in section 5 as well as the mean impact of rain on the geophysical parameters.

2 The Rain/cloud flag

2.1 The Matching Pursuit flag

The rain/cloud flag was defined pre-launch by Tournadre et al. (2009b) and it is only summarized in this section. A more detailed description is given in the Appendix. The detection of altimeter samples affected by rain or cloud is based on the analysis of the along-track variations of the off-nadir angle. The goal is here to detect the intervals where the off-nadir angle presents short scale coherent variations by analysis on the along-track off-nadir angle estimate, ζ^2 . The along-track $\zeta^2(t)$ is decomposed using Matching Pursuit and a dictionary of atoms defined by the wave-packet decomposition of the signal defined by Daubechies 8 (D8) mother wavelet. This wavelet was chosen because it is quite similar to the typical ζ^2 variations. The $\zeta^2(t)$ series is thus decomposed into a small number of atoms using the Mallat and Zhang MP algorithm (Mallat and Zhang, 1993), i.e.

$$\zeta^2(t) \simeq \sum_{i=1}^n a_i g_i(t) = \sum_{i=1}^n a_i \psi_{\lambda_i, t_i}(t) \quad (1)$$

where n is the number of selected atoms, a_i is the energy of atom i and g_i are the atoms chosen from the dictionary \mathcal{D} of the D8 wavelets $\psi_{\lambda,t}$.

In practice, the method can be applied to a signal of any length n . As the length is arbitrary, the signal is first padded to dyadic length (i.e. to 2^{m_1} where $m_1 = \lceil \log_2(m) \rceil$) by folding to allow discrete wavelet transform. The signal is then normalized by the ζ^2 noise computed in absence of cloud and rain and it is decomposed over the wavelet packet defined by the D8 mother wavelet. Only the wavelets of scale λ less than 9, i.e. less ~ 512 sample (about 80 km) length, are considered as longer scales can not be associated to rain or cloud. The MP algorithm is then applied to select the pertinent atoms. The atoms whose energy is larger than three times the noise level are selected.

The approximation of the signal by the selected atoms is a filtered version ζ^2 of the signal that contains only the most energetic short scale variations. The MP ζ^2 is then used to flag the samples by selecting the values whose absolute value is larger than a given threshold (here a tenth of the noise level). This selection is used to eliminate the small ripples associated located at the edges of some atoms.

As mentioned by Tournadre et al. (2009b) in their study, the MP algorithm detects the waveforms that are distorted. Surface slicks can also strongly distort waveforms and create short scale variations of ζ^2 . In such cases the signal is not attenuated but strongly enhanced (so that the phenomenon has been called “sigma0 bloom” in the literature) because of very high surface backscatter. However, the geophysical parameters, as in the case of rain, can be significantly modified. It should also be noted that sea ice and land if they are not flagged can also create short scale variations of ζ^2 .

2.2 The operational rain/cloud flag

Instrument Signal to Noise Ratio (SNR) considerations have been a main concern during all AltiKa development and the sensor’s link budget has been established by considering quite conservative hypotheses: a low case σ_0 (6.5 dB), a 3 -dB atmospheric attenuation (2 ways), and classical other system margins (3.5 dB for mispointing, ageing and overall system margin). With these hypotheses and taking into account all the sub-systems measured characteristics, the expected pre-flight SNR value was 11 dB. This can be seen as a worst case scenario. However, the mean in-flight observed SNR for the first cycles is much higher than 14 dB; the mean value being about 20.5 dB. There is actually no discrepancy between both values. Indeed, the design margins are available in addition to the 3 dB for rain attenuation, which allows withstanding very high rain rates without losing the signal.

The pre-launch analysis of the data losses due to rain and cloud was conducted using the low case scenario, i.e. a maximum 3 dB attenuation by rain (Tournadre et al., 2009a) and the global potential loss was estimated at 2 % for the 40 Hz data and 2.5 % for the 1 Hz GDR data. Thanks to the high in-flight SNR and to the robustness of the tracking algorithm, the availability of AltiKa over the ocean is much higher and data lost due to rain over ocean is lower than 0.1%. The data availability is thus almost perfect and samples that were supposed to be lost due to high rain attenuation are actually acquired and processed. As these data might potentially be strongly affected by rain, an accurate rain flagging is even more necessary than planned.

Because of limitations due to the configuration of the operational processing chain, the MP algorithm did not perform well. The parameterization was set to process ocean data and ocean data only, i.e. without sea ice or land data. In the ground segment, the sea ice flag is computed from 1 Hz data mainly because of the need of radiometer measurements. As the processing is performed on a track by track basis and that only the land editing criterion is applied, the pre-launch version of MP was largely disturbed by the presence of sea ice measurements. Indeed, sea ice waveforms have strong short-scale variations of the trailing edge that can saturate the atoms needed for the rain detection. In order to reduce the impact

of these perturbations on the algorithm, two modifications are made: the maximal number of atoms is increased and the signed square root of the off-nadir angle ($\frac{\xi^2}{|\xi^2|} \sqrt{|\xi^2|}$) is considered in order to mitigate the strong variations observed in sea ice data. Although the resulting configuration might not be optimal, the changes made allow the MP to account for sea ice data and to provide a much better detection of rain over the open ocean. The new and pre-launch settings are summarized in Table 1. The new noise level was recomputed from the square root of the off-nadir angle for the first 5 cycles. The new value of the noise given in Table 1 is half than the pre-launch one, which also partially explains the poor pre-launch flag performances.

A last point must be pointed out: due to the strong correlation between the slope of the trailing edge and the backscatter coefficient in the MLE4 estimator, as mentioned in (Quarty, 2009; Thibaut et al., 2010), a specific off-nadir angle is recomputed to provide uncorrelated input data to the MP algorithm. The slope of the logarithm of the trailing edge is estimated by a linear regression on the 40 Hz waveform samples between the 81st and the 116th waveform samples before use in the MP algorithm. Operationally, the 40 Hz data are flagged using the MP algorithm while the 1 Hz GDR data are flagged if 50% of the 40 Hz data are flagged.

3 Collocation with SSMI-S/Windsat rain rate measurements

In order to set up a qualitative and quantitative reference for the study of the impact of rain on altimeter measurements, a “combined” rain rate product is built from observations provided by instruments dedicated to the atmosphere monitoring. The goal is to create a reference allowing to distinguish as well as possible between rain and no rain event and to potentially fix an upper limit to the rain rate above which the impact on altimeter measurements is critical. Three different data sources have been used: SSMI-S (Special Sensor Microwave Imager-Sounder) on-board the F16 and F17 satellites under the Defense Meteorological Satellite Program (DMSP) and Windsat radiometer on-board Coriolis. The rain rates products are generated by Remote Sensing System using the Intercalibrated Passive Microwave Radiometer Rain Products from the Unified Microwave Ocean Retrieval Algorithm (UMORA) (Hilburn and Wentz, 2008). The daily ground track products gridded on 0.25°x0.25° maps are used. The AltiKa altimeter data have been systematically collocated with the rain data from the three missions (F16, F17 and Windsat) and the closest observation in time is kept, together with the time lag (Δt_{RR}) between the altimeter and the rain observation. This collocation has been applied to the first 9 cycles of AltiKa and the “combined” rain rate, R_c , is available for about 80% of AltiKa observations over the ocean. About 7% of R_c are positive indicating rain events. This proportion is consistent with the analysis of SSMI-S F16 rain rate data larger than 0.2 mm/hr (Wentz and Spencer, 1998). For equatorial regions where most of the rain events occur, the mean Δt_{RR} is close to 10 minutes.

4 Impact of the atmospheric liquid water on altimeter waveforms

An example that shows the rain impact on AltiKa data as well as the problem of collocating microwave radiometer rain data and altimeter is presented in Figure 1-a with AltiKa waveforms (accounting for the altimeter Automatic Gain Control, AGC) over a portion of track number 121, cycle 5. The co-localized rain rate (R_c) at a time lag lower than 5 minutes is also shown in the figure. SSMI-S/Windsat rain data indicate a quite strong rain event with a rain rate varying between 4.8 and 7 mm/hr between 28.6° and 29.25° of latitude, however only two zones of perturbations are visible in the AltiKa waveforms (a), a first one between 28.6° and 28.7° latitude and a second one between 28.95° and 29.15°. The effect of

rain on waveforms for this two events corresponds very well to the simulations presented by (Tournadre et al., 2009a). Firstly, as the satellite approaches (or departs from) the rain cell edges, attenuation by rain impacts the waveform trailing edge modifying its slope. Then, when the rain cell is located at the satellite nadir the returned power is strongly attenuated (and may be null resulting in a tracker loss if the rain rate is high enough) and the entire waveform is affected. The distortion of the waveforms within the rain cells can be clearly seen in Figure 1-c presenting the mean waveforms, normalized by their maximums, computed outside and inside (delineated by the two dotted red rectangles in the figure) the rain cells. Within the rain cells, the waveform shape is significantly modified. The slopes of the leading and trailing edges are strongly altered. This example deals with a case of rainfall but the same observations could also be made in presence of high cloud liquid water content event.

The retracking algorithm implemented in the operational ground segment is the Maximum Likelihood Estimator MLE4 that tries to fit a theoretical waveform with a predetermined shape to the measured one (Amarouche et al., 2004). Any modification of the shape of the waveform can strongly modify the geophysical estimates. The algorithm fits a Brown model (Brown, 1977) to the altimeter waveform using an iterative unweighted least square estimator algorithm. The four geophysical parameters estimated are the altimeter range and significant wave height (SWH) determined respectively by the position of the half power of the leading edge and its width, the backscatter coefficient, σ_0 or sigma0, inferred from the waveform amplitude (in addition to the classical radar equation) and the square of the off-nadir angle (ζ^2) derived from the slope of the logarithm of the altimeter waveform trailing edge. MLE4 was initially designed to account for real mis-pointing issues of the Jason-1 mission. A drawback of this algorithm, already pointed out in Quartly (and 2009,) and Thibaut et al. (2010), is to introduce a strong correlation between the slope of the trailing edge and σ_0 . Any variation of the slope of the trailing edge impacts the σ_0 estimate.

The geophysical parameters computed from the waveforms shown in Figure 1-a are plotted on Figure 2-b, -c, -d and -e. For the rain event near 29° with rain rate reaching 7 mm/hr, the MLE4 estimates present strong erroneous values. Rapid and strong variations are observed on each parameter reaching biases of about 4 dB for σ_0 , - 7 m for the Sea Level Anomaly (SLA), more than 8 m for the SWH and 0.06 deg² for ζ^2 . Concerning the rain event near 28.6°, with rain rate about 6 mm/hr, the rain impact is limited to σ_0 and off-nadir angle. The waveforms are only weakly distorted and SLA and SWH are only marginally modified. In this example, the new operational MP flag detects all the rain affected data. By design the MP algorithm flags segments of data corresponding to atoms. This implies that some good data might be flagged especially at the edges of rain events (for example near 28.9°). This example shows the strong impact of rain on MLE4 parameters but it is also a good illustration of the difficulty of collocating microwave radiometer rain data and altimeter. The largest rain rate are clearly not associated with any attenuation or modification of the waveforms near 28.8°. This certainly results from partial beam filling effect in the radiometer and the resolution difference between the two sensors.

5 Matching Pursuit flagging

5.1 Flagged data

The new operational detection algorithm has been applied to the first 9 cycles of AltiKa. The mean percentage of flagged samples for the first cycles 2 to 9 for the 40 Hz and 1 Hz data are presented in Figure 3 . The first cycle was not considered because of the large number of orbit maneuvers. The problem of sea ice mentioned earlier is clearly visible in these fields. Over sea ice 100% of the data are flagged by MP. This confirms the necessity to increase the number of atoms for a good detection of rain.

Some orbit tracks also clearly appears in the fields, south of South Africa and South America, and in the Indian Ocean. They correspond to tracks that are entirely flagged due to orbit maneuvers that induced mis-pointing of the sensor detected by the algorithm. The percentage of flagged 1 Hz data is slightly lower than that at 40 Hz because of the way the 1 Hz flag is estimated.

In the open ocean, the geographical distributions of flagged samples reflects the global distributions of precipitations (see for example Adler et al. (2003)) with maximums in the Intertropical Convergence Zone (ITCZ), the South Pacific Convergence zone (SPCZ) and at mid-latitude within the storm tracks of the northern Hemisphere oceans. In the Western Pacific Warm Pool, about 20-25% of the 40 Hz and 17-20% of the 1 Hz data are flagged. In the ITCZ and SPCZ about 15 % of the data are flagged at 40 and 1 Hz while at mid-latitude the percentage varies from 5 to 10 %. Using the 1 Hz flag to eliminate data over sea ice, the mean percentage of flagged samples is about 9%. This is larger than the proportion of rainy data given by the analysis of SSMI-S F16 rain data (Wentz and Spencer, 1998). As already mentioned in section 2.1, MP algorithm detects not only the data affected by rain but all the samples that present a deformation of the waveform shape including the ones caused by strong short scale variations of the ocean surface backscatter or sigma0 bloom. Mitchum et al. (2004) using Topex-Poseidon Ku-band altimeter data and a threshold criterion on backscatter showed that up to 5% of the data can be affected by such high radar return cross section events. Using the same kind of analysis, Thibaut et al. (2007) estimated at 3-4% the percentage of sigma0 bloom in Envisat radar altimeter data. Tournadre et al. (2006) showed that the sigma0 blooms result from streaks of high backscatter within the altimeter footprint under light wind conditions. The AltiKa radiometer Integrated liquid water content (ILWC) is thus used to assess the presence of liquid water within the altimeter footprint and to discriminate between rain and no rain. The percentage of flagged samples for which the ILWC is larger or smaller than 0.1 kg/m² are presented in figure 4 . About 5.5% of data are MP flagged and have an ILWC larger than 0.1 kg/m². Those can be considered as rainy or cloudy. The global proportion is in good agreement with the global percentage of rain. About 3.5% of the samples are MP flagged but do not contains significant liquid water. They corresponds to mis-pointing due to orbit maneuvers or to short scale variations of surface backscatter. Their geographical distribution with maximums in the tropics and at mid-latitude south of the storm tracks is in good agreement with the global distribution of sigma0 bloom presented by Mitchum et al. (2004) for Topex and Thibaut et al. (2007) for Envisat. It should be noted that although AltiKa has a significantly smaller footprint than Envisat or Topex and responds to different scales of sea surface roughness to Ku-band altimeters, the proportion and global distributions of sigma0 blooms is remarkably similar to Envisat and Topex ones.

The zonal average of the percentage of 1 Hz flagged samples (global, rainy/cloudy and bloom) is presented in figure 5. The meridional structures of the rain flagged samples are similar to previously published rain climatologies (Wentz and Spencer, 1998). The maximum (~10%) occurs at the equatorial latitudes associated with the ITCZ. A secondary maximum associated to the SPCZ is observed between 10°S and the Equator. As for precipitations, the extra-tropical percentage is greater (by about 2%) in the Northern Hemisphere than in the Southern Hemisphere. Within the mid-latitude storm tracks, near 40° of latitude, the percentage of rainy samples is about 6.5% in the northern hemisphere and 5% in the southern hemisphere. This north-south difference is also observed in the precipitation climatology. The distribution of blooms also presents this north-south asymmetry with about 2% more events in the northern hemisphere than in the southern one. The maximums are found in the equatorial regions and at mid-latitude equatorward of the storm tracks.

5.2 Validation of the matching pursuit rain flag with co-localized Windsat/SSMIS rain rate data

The only way to assess the performances of the MP flag is to compare the results of the flagging process with coincident rain data. A first validation consists in quantifying detection performances of the MP flag as a function of rainfall (presence and intensity). However as already shown by Tran et al. (2008) and Tournadre and Bhandari (2009), the high spatial and temporal variability of rainfall and the small lengthscale of rain events imply that the larger resolution radiometer rain data might not be representative of the smaller resolution altimeter data rain conditions. To mitigate these phenomena, positive rain rates with a maximum 5 min time lag are considered as rainy for validation only if the AltiKa radiometer ILWC is larger than 0.1 kg/m^2 . This criterion is used to ensure the presence of at least some atmospheric liquid water near or within the AltiKa altimeter footprint. The number of such collocated ocean samples is 1209605 distributed over the 9 first cycles of AltiKa. The percentage of flagged measurements computed as a function of the rain rate (for rain rate larger than 0.5 mm/hr) is presented on Figure 6. For rain rates larger than 5 mm/hr between 80 and 90% of the data are flagged. Below this rainfall rate, the fraction of detected measurements decreases almost linearly to 37% for very light rain (0.5 mm/hr). For 1 mm/hr rain rates (light rain) about 50% of the data are detected as rainy. This results are in good agreement with the pre-launch simulations of (Tournadre et al., 2009a) which showed that all data with rain rate larger than 5 mm/hr should be affected and that for rain rate smaller than 5 mm/hr depending on the size of the rain cell between 10 to 80 % of the high resolution samples can be affected. This analysis shows that the MP flagged samples and ILWC larger than 0.1 kg/m^2 are clearly associated with rain events.

To further evaluate the performances of the flag, the Tran et al. (2008) method a dichotomous discrimination is also applied. This method consists in evaluating the percentage of hits, misses, false alarms and correct negative cases as described in Table 2. As very low rain rate (0.5 mm/hr) from microwave radiometer are dubious (Wentz and Spencer, 1998), only rain rate larger than 0.5 mm/hr are considered as rainy. This analysis is thus conducted on a subset of AltiKa data for which there exists a collocated rain rate at a less than 5 min time lag. The results of the dichotomous discrimination for cycles 1 to 9 are given in Table 3 and they are also plotted on Figure 7. The proportion of samples detected as rainy (a) by MP (solid line) varies from 1.6% to 2.37% (average value 1.95%). This is larger than missed samples (b) which ranges from 1.08% to 1.68% (average value 1.39%). The fraction of correct negatives (d) evolves around 91% whereas the false alarm rate is relatively high (c) at about 6%. This last proportion is expected because of the detection of σ_0 bloom events. To discriminate the σ_0 blooms from real false detection of rainy samples, a threshold on the radiometer liquid water content is added: $\text{ILWC} > 0.1 \text{ kg/m}^2$. Adding this criterion decreases significantly the proportion of false alarm to about 3% while the proportion of correct negatives reaches 93-95%. As the ILWC criterion is already included to determine if a microwave rain data is rainy there is no impact on the proportion of hit and misses. These results are similar to the ones presented by (Tran et al., 2008) for the Envisat rain flag. Considering the potential discrepancies between rain data and altimeter due to partial beam filling, difference of resolution, time lag, difference of sensitivity to rain and liquid water, the MP flag appears to perform satisfactorily well.

It is also important to quantify the impact of rain on the geophysical parameters, the distribution of the standard deviation of each MLE4 parameter estimate (computed from the 40 Hz elementary measurements used to compute a 1 Hz value) is plotted for the four groups of Table 2: hits (solid black line), misses (solid gray line), False Alarm (FA) (dotted line) and Correct Negative (CN) (dashed line) in Figure 8. The CN population that represents the perturbation-free oceanic conditions is characterized by a low standard deviation for each MLE4 geophysical parameter. The maximum of the distribution is about 0.07 dB for σ_0 rms (a), 0.0025 deg^2 for ζ^2 rms (b), 0.048 m for the range rms (c) and 0.3 m for the SWH rms

(d). The distributions have narrow peaks and fall rapidly after the maximum especially for σ_0 and ζ^2 . In contrast, the distributions of the “hits” group are very broad especially for σ_0 and ζ^2 showing the large impact of rain on the geophysical parameters estimates. The “Misses” and FA group distributions are quite similar for all parameters and have much lower probability than the hits group for large rms values but significantly larger one than the CN one. The shapes of the distribution shows that the two groups (“Misses” and FA) contains both correct and rain affected samples in a proportion close to 50%. This reflects problems of collocation of rain data as well as the imperfections of the MP flag which certainly flags good data especially near the edges of rain cells.

6 Conclusion

Because of the use of a single frequency for the AltiKa altimeter, classical dual frequency rain flag such as the ones used for the Jason and Envisat satellite could not be used. As Ka band is highly sensitive to atmospheric liquid water, it is necessary to accurately detect and flagged the data potentially affected by rain. Based on simulation of echo waveforms at Ka band and results from the analysis of Ku-band altimeter in presence of rain a new rain flag based on the detection of short scale variations of the off-nadir angle estimates (i.e. the change of the slope of the plateau of the waveform) has been developed and implemented in the ground segment. The detection algorithm is based on a Matching Pursuit algorithm that decomposes the along-track signal into a limited number of atoms selected by their energy. The pre-launch parameters of the algorithm were estimated using Jason-1 data. The first GDR reprocessing of the data is used to improve the algorithm that was not performing correctly. To take into account the fact that sea ice data are not screened before the detection and saturate the number of atoms, the square root of the off-nadir angle is considered and the number of atoms is increased. The noise level of the off-nadir angle, a crucial parameter for the selection of the atom, has also been estimated from real data. The noise level is half the value estimated from Jason-1 data and used pre-launch.

The upgraded rain flag has been applied to the first 9 of AltiKa. Globally about 9% of the data are flagged with maximum between 10 and 25% in the ITCZ, the SPCZ and the mid-latitude storm tracks. As expected MP flags not only rain affected data but also the ones for which the waveforms are distorted by short scale variations of surface backscatter (sigma0 bloom). ILWC from the AltiKa radiometer is used to assess the presence of atmospheric liquid water and thus to discriminate the rain affected samples The global distribution of rain flagged samples reflects very well the distributions of oceanic precipitations. The global mean percentage is about 5-6% in agreement with the percentage of rainy data in SSMI-S F16. About 3-4% of the data MP flagged but not rainy according to the ILWC threshold correspond to sigma0 bloom events. This proportion is very similar to the ones published for Topex and Envisat Ku-band altimeter data.

The flag is validated using collocated rain data from the SSMI-S and Windsat radiometers. Only data with time lags less than 5 min and for which there is liquid water (AltiKa ILWC > 0.1 kg.m⁻²) are considered to limit the effect of the rain sporadic nature. The flag detects very well rain rate larger than 2 mm/hr (more 60% detection) and almost all the rain rate larger than 5 mm/hr. Using the (Tran et al., 2008) dichotomous discrimination the percentage of hits, misses, false alarms and correct negative is estimated. The results are similar to the ones published for the Envisat altimeter rain flag (Tran et al., 2008). The correct negative represents 93.5 % of the data, the hits about 2% the false alarms about 3% and the misses about 1.5%.

The analysis of the root mean square of each MLE4 parameter estimate (computed from the 40 Hz elementary measurements used to compute a 1 Hz value) and for the four groups shows that MP clearly

detects samples that have a high rms for all parameters. On the other hand, the correct negative group presents very narrow distributions with low maximums of backscatter and off-nadir angle rms typical of rain free conditions. The false alarm and misses groups both contains correct and erroneous data but present a significantly lower probability of high rms values of geophysical parameters.

The upgraded MP flag performances are thus very satisfactory and very similar to the ones obtained for dual-frequency rain flag such as the Envisat one.

Acknowledgements

Windsat data are produced by Remote Sensing Systems and sponsored by the NASA Earth Science MEaSUREs DISCOVER Project and the NASA Earth Science Physical Oceanography Program. RSS Windsat data are available at www.remss.com. SSMI and SSMI-S data are produced by Remote Sensing Systems and sponsored by the NASA Earth Science MEaSUREs Program and are available at www.remss.com.

Appendix: Matching Pursuit

Matching pursuit was originally formulated by Mallat and Zhang (1993) as a technique for identifying the time/frequency content of a time series whose spectral properties evolve over time. The basic idea was to construct a large “dictionary” of explanatory functions that are localized both in time and in frequency and then to analyze a time series by projecting it against the functions in the dictionary. Matching pursuit can be adapted to explore other properties of a time series besides its time/frequency content.

Below we briefly recall the basic ideas: we seek a linear expansion approximating the analyzed signal $s(t)$:

$$s(t) \approx \sum_{i=1}^M a_i g_i(t) \quad (2)$$

in terms of functions g_i chosen from a large and redundant set of basic functions (dictionary \mathcal{D}). The problem of choosing M functions, which explain the largest part of variance of a given signal is a *N-P hard problem* Davis *et al.* (1997), i.e. computationally intractable. MP offers a sub-optimal solution, obtained by means of an iterative algorithm. In the first step, the function g_0 which gives the largest product with the signal s is chosen from the dictionary \mathcal{D} , composed of normalized functions ($\|g_n\| = 1$). In each of the consecutive steps, the function g_n is matched to the signal R_s^n which is the residual left after subtracting results of previous iterations:

$$\begin{cases} R_s^0 = s \\ R_s^{n+1} = R_s^n - \langle R_s^n, g_n \rangle g_n \\ g_n = \operatorname{argmax}_{g_i \in \mathcal{D}} (\langle R_s^n, g_i \rangle) \end{cases} \quad (3)$$

For a complete dictionary the procedure converges to s with $M \rightarrow \infty$ Mallat and Zhang (1993). In practice we use finite expansions

$$s = \sum_{n=1}^M \langle R_s^n, g_n \rangle g_n \quad (4)$$

Orthogonality of R_s^{n+1} and g_n in each step implies the conservation of energy

$$\|s\|^2 = \sum_{n=1}^{M-1} |\langle R_s^n, g_n \rangle|^2 + \|R_s^M\|^2 \quad (5)$$

The complete demonstration of the convergence and energy conservation is given in Mallat and Zhang (1993). The most important feature of this decomposition is that it is a greedy algorithm, i.e. that it chooses at each iteration a function that is best adapted to approximate part of the signal.

Different dictionaries can be used for MP depending on the type of time-series to analyze. As originally proposed by Mallat and Zhang (1993) for audio signal, we use wavelets to create the time-frequency dictionary of atoms necessary for MP. The continuous wavelet transform (CWT) of time series s with respect to wavelet ψ is defined as

$$W(\lambda, t) = \int_{-\infty}^{+\infty} \psi_{\lambda,t}(u)s(u)du \quad (6)$$

where

$$\psi_{\lambda,t}(u) = \frac{1}{\sqrt{\lambda}} \psi\left(\frac{u-t}{\lambda}\right) \quad (7)$$

The variable λ is the scale and t is the point where the wavelet is centered.

For discrete time-series, CWT cannot be computed exactly, and discrete wavelet transform (DWT) has to be used. The DWT can be regarded as an approximation of the CWT over a so-called dyadic grid of scales. Each row is usually set to the largest integer that is less than or equal to $\log_2(N)$ where N represents the sample size (Coifman et al., 1992).

References

- Adler, R., G. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydromet.* 4(6), 1147–1167.
- Amarouche, L., P. Thibaut, O. Zanife, J.-P. Dumont, P. Vincent, and N. Steunou (2004). Improving the Jason-1 ground retracking to better account for attitude effects. *Marine Geod.* 27, 171–197.
- Brown, G. S. (1977). The average impulse response of a rough surface and its applications. *IEEE Trans. Antennas Propag.* AP-25, 67–74.
- Coifman, R. R., Y. Meyer, and M. V. Wickerhauser (1992). *Wavelets and Their Applications*, Chapter Wavelet analysis and signal processing, pp. 153–178. Jones and Bartlett, Boston.
- Davis, G., S. Mallat, and M. Avelano (1997). Greedy adaptative approximations. *J. Constr. Approx.* 13, 57–98.
- Hilburn, K. and F. Wentz (2008). Intercalibrated passive microwave rain products from the unified microwave ocean retrieval algorithm. *J. Appl. Met. Clim.* 47, 778–795.
- Mallat, S. G. and Z. F. Zhang (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* 41, 3397,3415.
- Mitchum, G., D. Hancock, G. Hayne, and D. Vandemark (2004). Blooms of sigma0 in the TOPEX radar altimeter data. *J. Atmos. Oceanic Tech.* 21, 1232–1245.
- Quarty, G. (2009). Optimizing σ_0 information from the Jason-2 altimeter. *IEEE Geosc. Remote Sens. Let.* 6 (3), 398–402.

- Quartly, G. D., T. H. Guymer, and M. A. Srokosz (1996). The effects of rain on Topex radar altimeter data. *J. Atmos. Oceanic. Tech.* 13, 1209–1229.
- Thibaut, P., F. Ferreira, and P. Femenias (2007). Sigma0 bloom events in the envisat radar altimeter data. In European Space Agency (Ed.), *Proc. Envisat Symposium 2007, Montreux, Switzerland 23-27 April 2007*, Volume ESA SP-636.
- Thibaut, P., J. Poisson, E. Bronner, and N. Picot (2010). Relative performance of the MLE3 and MLE4 retracking algorithms on Jason-2 altimeter waveforms. *Marine Geodesy* 33, 317–335.
- Tournadre, J. (1998). Determination of rain cell characteristics from the analysis of Topex altimeter echo waveforms. *J. Atmos. Oceanic Technol.* 15(2), 387–406.
- Tournadre, J. (1999). Estimation of rainfall from Ka band altimeter data : computation of waveforms in presence of rain. In *Procs of IGARSS-99, IEEE Press, Piscataway, NJ, Hambourg, Allemagne, 28 jun.-2 jul.*, Vol. I, pp. 197–199. IEEE Press.
- Tournadre, J. (2004). Validation of Jason and Envisat dual frequency rain flags. *Marine Geodesy* 27, 153–170.
- Tournadre, J. and S. Bhandari (2009, JAN). Analysis of Short Space-Time-Scale Variability of Oceanic Rain Using TOPEX/Jason. *J. Atmos. Oceanic Technol.* 26(1), 74–90.
- Tournadre, J., B. Chapron, N. Reul, and D. C. Vandemark (2006, April). A satellite altimeter model for ocean slick detection. *J. Geophys. Res.* 111(C4), C04004.
- Tournadre, J., J. Lambin-Artru, and N. Steunou (2009a). Cloud and Rain Effects on AltiKa/SARAL Ka-Band Radar Altimeter-Part I: Modeling and Mean Annual Data Availability. *IEEE Trans. Geosc. Remote Sens.* 47(6), 1806–1817.
- Tournadre, J., J. Lambin-Artru, and N. Steunou (2009b). Cloud and Rain Effects on AltiKa/SARAL Ka-Band Radar Altimeter-Part II: Definition of a Rain/Cloud Flag. *IEEE Trans. Geosc. Remote Sens.* 47(6), 1818–1826.
- Tournadre, J. and J. C. Morland (1997). The effects of rain on TOPEX/Poseidon altimeter data. *IEEE Trans. Geosc. Remote Sens.* 35(5), 1117–1135.
- Tran, N., E. Obligis, and F. Ferreira (2005). Comparison of two Jason-1 altimeter precipitation detection algorithms with rain estimates from the TRMM Microwave Imager. *J. Atmos. Ocean. Technol.* 22, 782–794.
- Tran, N., J. Tournadre, and P. Femenias (2008). Validation of Envisat Rain Detection and Rain Rate Estimates by Comparing With TRMM Data. *IEEE Geosc. Rem. Sens. Let.* 5(4), 658–662.
- Vincent, P., N. Steunou, E. Caubet, L. Phalippou, L. Rey, E. Thouvenot, and J. Verron (2006). AltiKa: a Ka-band Altimetry Payload and System for Operational Altimetry during the GMES Period. *Sensors* 6, 208–234.
- Wentz, F. J. and R. Spencer (1998). SSM/I Rain Retrievals within a Unified All-Weather Ocean Algorithm. *J. Atmos. Ocean. Tech.* 55, 1613–1627.

Table 1: Main parameters of the Matching Pursuit Algorithm pre-launch and for the GDR first reprocessing

Parameter	Pre-launch	GDR reprocessing
Noise level	$8.2 \times 10^{-4} \text{ deg}^2$	0.02 deg ($4 \times 10^{-4} \text{ deg}^2$)
Maximal number of atoms	200	450

Table 2: Discrimination table used to validate the rain flag

		SSMI-S/Windsat Rain rate $> 0.5 \text{ mm/hr}$ and $\text{ILWC} > 0.1 \text{ kg/m}^2$	
		Yes	No
MP flag	Yes	Hits	False alarms (FA)
	No	Misses	Correct negatives (CN)

Table 3: Percentages of Hits, Misses, FA and CN for cycles 1 to 9 with flag MP and flag MP and $\text{ILWC} > 0.10.1 \text{ kg.m}^{-2}$.

cycle number	1	2	3	4	5	6	7	8	9
MP flag									
Hits	2.19	1.92	1.82	1.62	1.88	1.98	1.75	2.09	2.38
Misses	1.06	1.31	1.58	1.67	1.36	1.50	1.42	1.28	1.30
FA	12.63	6.84	4.64	3.94	5.97	5.99	5.88	6.74	5.56
CN	84.12	89.93	91.96	92.77	90.79	90.53	90.95	89.89	90.76
MP flag and $\text{ILWC} > 0.1 \text{ kg.m}^{-2}$									
Hits	2.17	1.91	1.81	1.61	1.88	1.97	1.74	2.08	2.36
Misses	1.08	1.32	1.59	1.67	1.36	1.5	1.43	1.29	1.31
FA	4.01	2.57	2.24	2.12	2.92	3.28	2.65	2.70	2.41
CN	92.74	94.2	94.36	94.60	93.84	93.25	94.18	93.93	93.92

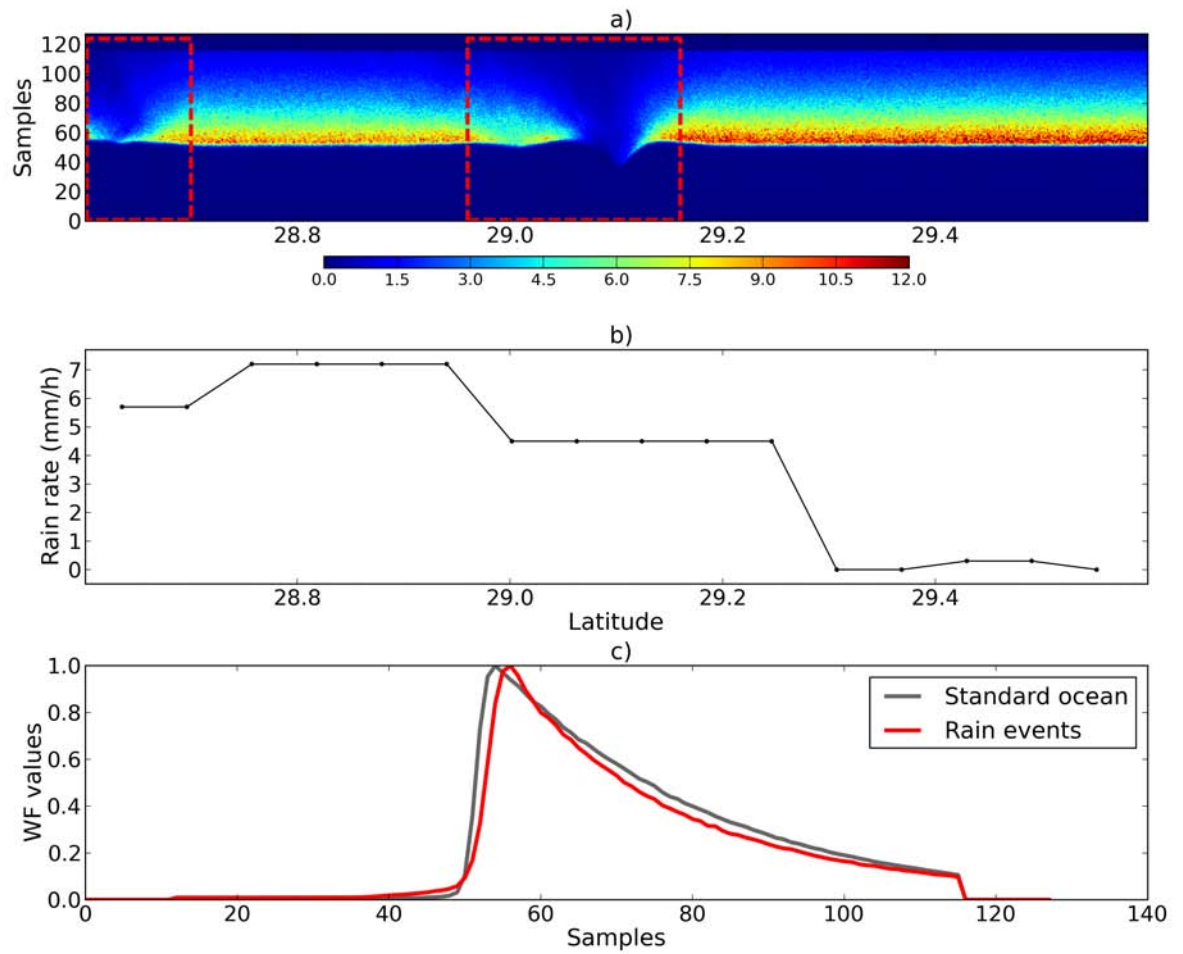


Figure 1: a) AltiKa waveforms corrected from Automatic Gain Control during rain events, Pass 121 Cycle 5, b) collocated rain rate, (c) mean normalized waveforms in nominal conditions (black line) and during rain events delineated by the red dotted rectangles in (a) (red line).

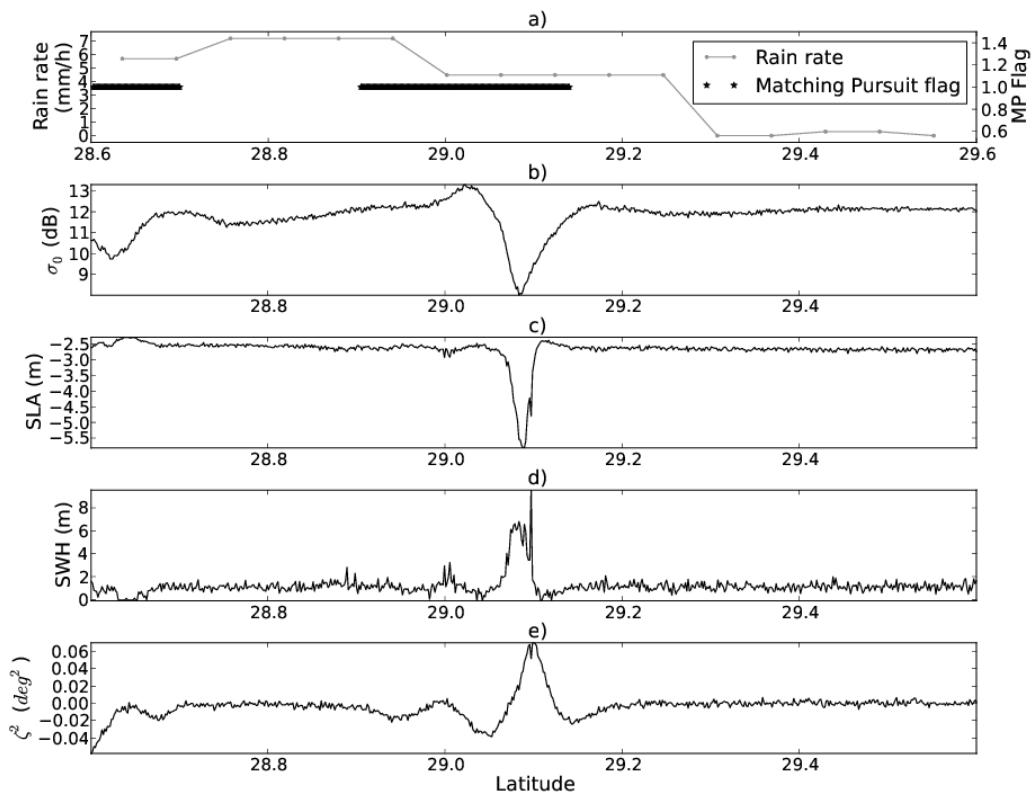


Figure 2: . Geophysical parameters estimated by MLE4 for the waveforms of figure 1. a) Colocalized rain rate measured by SSMI-S/Windsat (grey line) and the matching pursuit Flag (black dot), b) backscatter coefficient σ_0 , c) Sea Level Anomaly, d) significant wave height, e) off-nadir angle estimate.

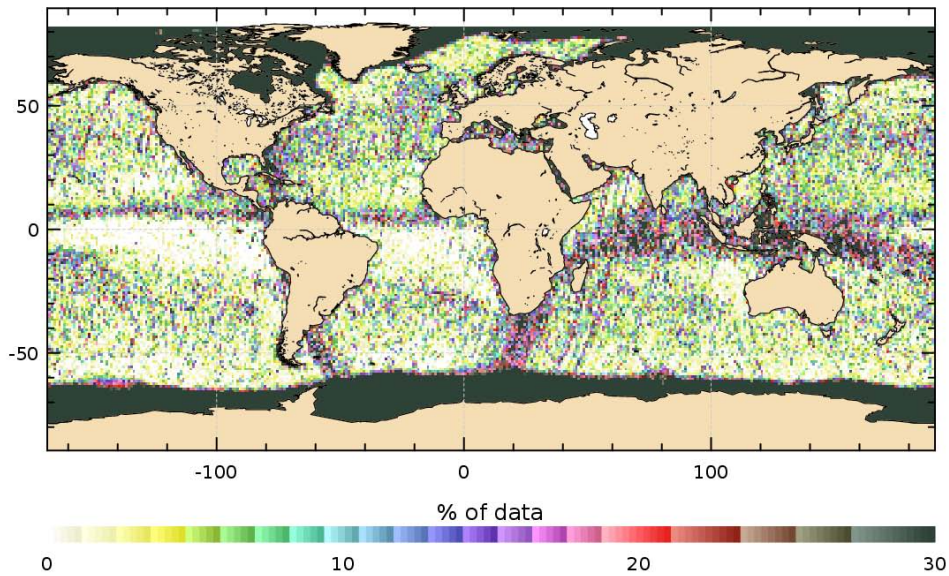


Figure 3: Mean percentage of 40 Hz (top) and 1 Hz (bottom) flagged samples using the operational MP algorithm for cycles 2 to 9 of AltiKa.

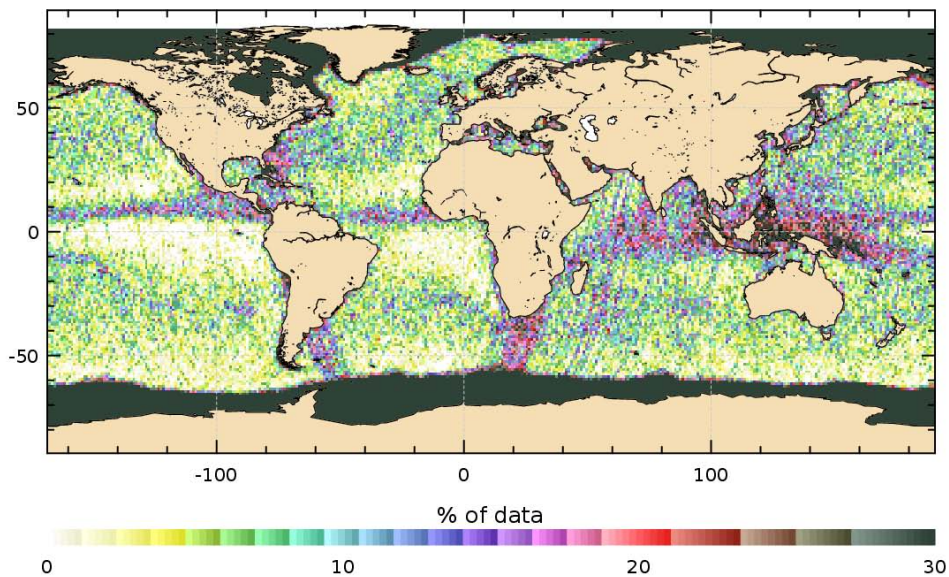
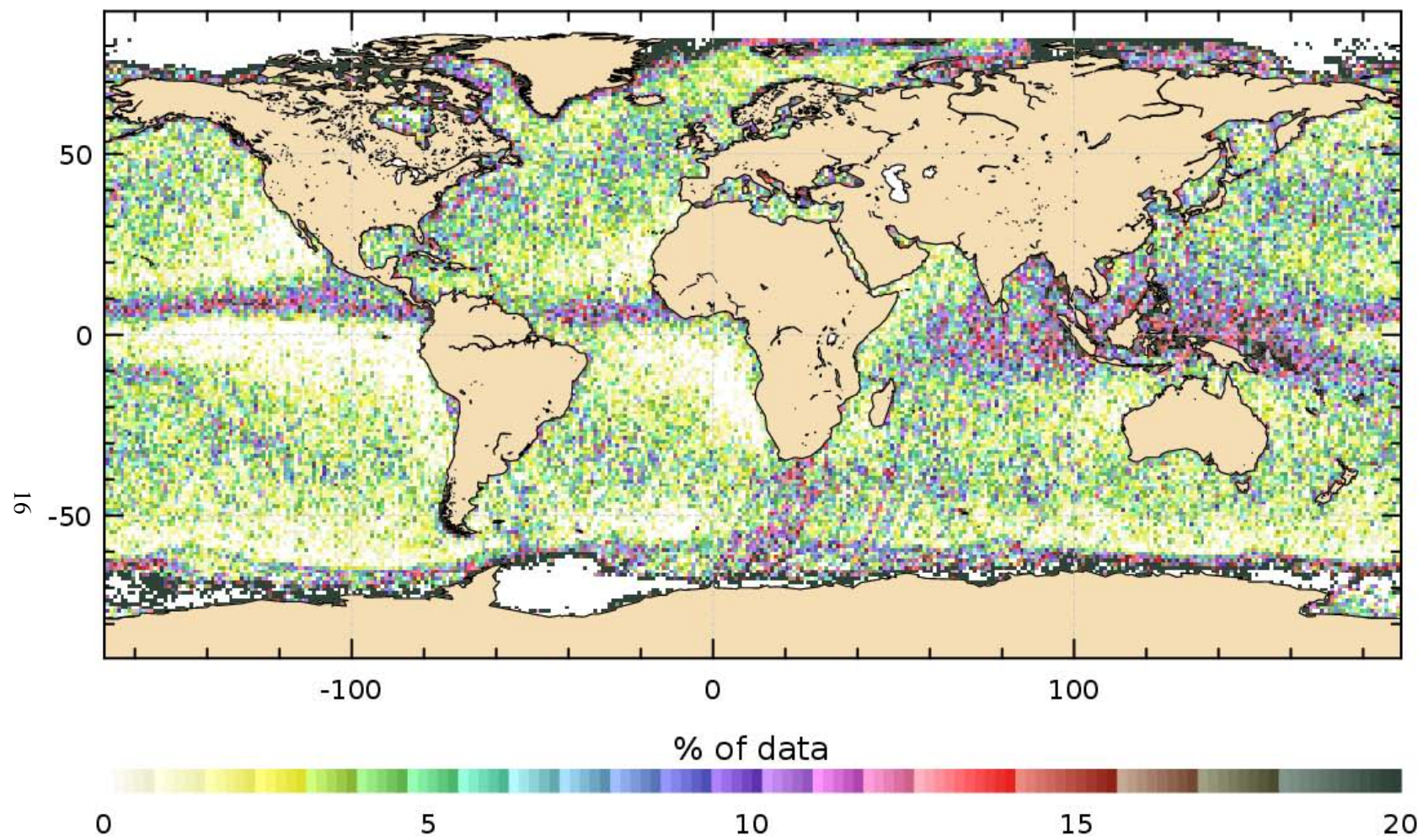
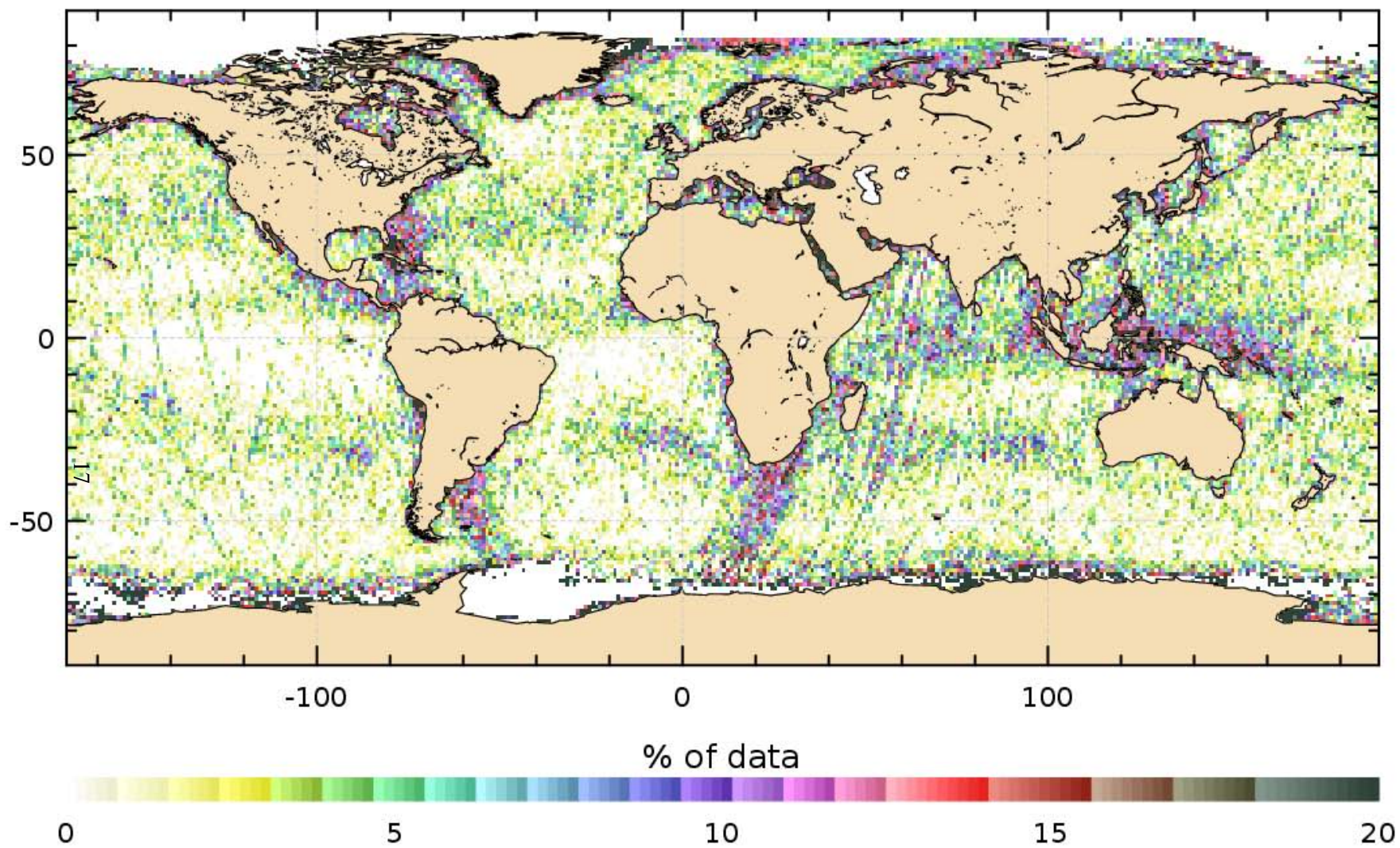


Figure 4: Percentage of 1Hz flagged samples using the operational MP algorithm for which the ILWC $> 0.1\text{kg/m}^2$ (top) and $< 0.1\text{kg/m}^2$ (bottom) for cycles 2 to 9.





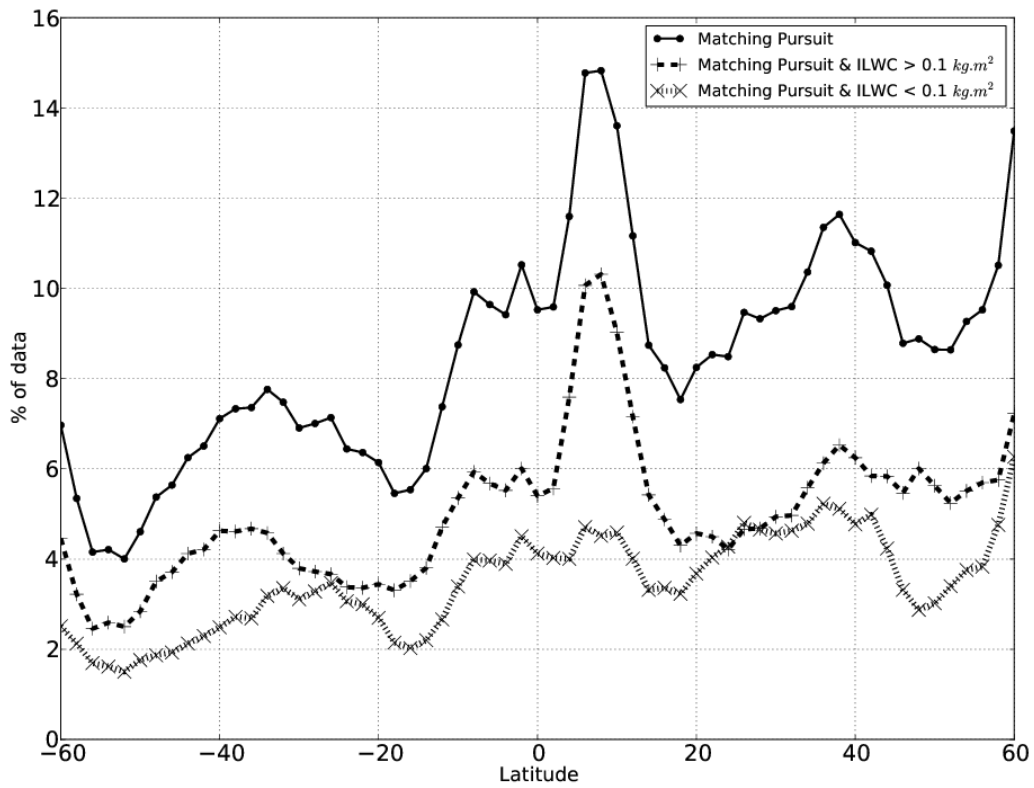


Figure 7: . Percentage of flagged samples as a function of AltiKa cycle number for (a) hits, (b) misses, (c) False alarms and (d) correct negatives.

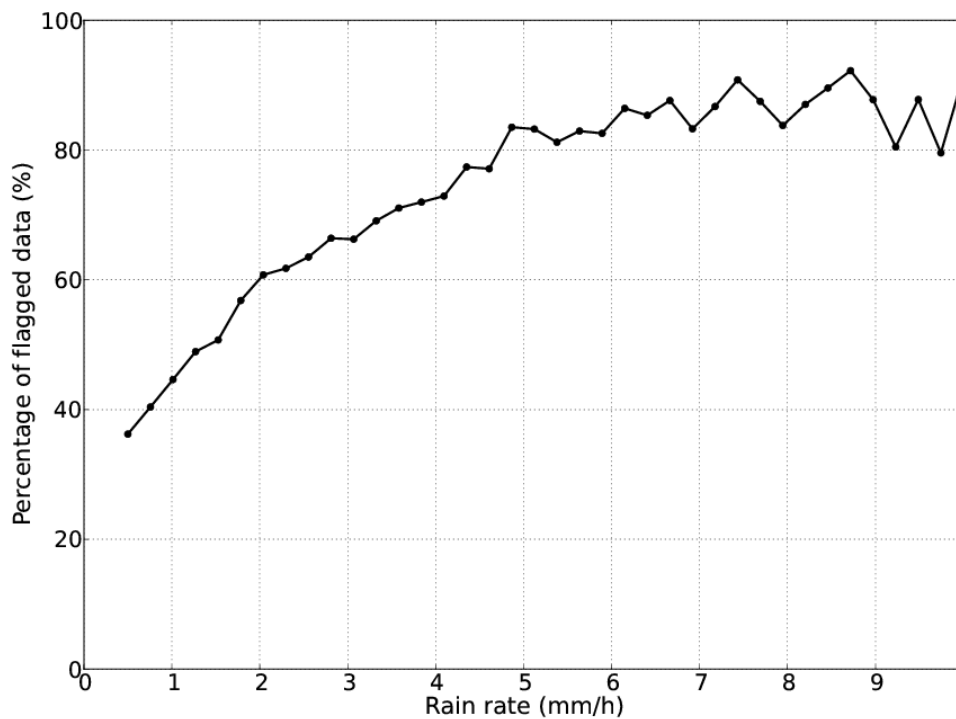


Figure 8: . Distribution of geophysical parameters root mean square (a) backscatter (b) off nadir angle, (c) range, (d) significant wave height. The four groups of Table 2 are represented by black solid lines for the hits group, the solid gray lines for the misses one, the small-dotted lines for the false alarm one and the dotted lines for the correct negative one.