ROM SAF Report 23
Ref: SAF/ROM/METO/REP/RSR/023
Web: www.romsaf.org
Date: 23 January 2016

The EUMETSAT Network of Satellite Application Facilities

ROM SAF
Radio Occultation Meteorology

ROM SAF Report 23

Assessment of a potential reflection flag product

Estel Cardellach and Santiago Oliveras

IEEC, Barcelona, Spain

DOCUMENT AUTHOR TABLE

	Name	*Function*	*Date*	*Comments*
Prepared by:	E. Cardellach	ROM SAF Local Manager	23/02/2015	
Prepared by:	S. Oliveras	ROM SAF/IEEC Engineer	23/02/2015	
Reviewed by:				
Approved by:	K. B. Lauritsen	ROM SAF Project Manager	23/01/2016	

DOCUMENT CHANGE RECORD

Issue/Revision	*Date*	*By*	*Description*
1.0	23/01/2016	EC	First version of Report on Assessment of a potential reflection flag product, cf. CDOP-2 WP 2320 and 2330 (a draft version of the report, dated 23 February 2015, was prepared as internal ROM SAF contribution to MS ROM-042)

ROM SAF

The Radio Occultation Meteorology Satellite Application Facility (ROM SAF) is a decentralized centre under EUMETSAT, responsible for operational processing of GRAS radio occultation data from the Metop satellites as well as RO data from other missions. The ROM SAF delivers bending angle, refractivity, temperature, pressure, and humidity profiles in near-real time and offline for NWP and climate users. The offline profiles are further processed into climate products consisting of gridded monthly zonal means of bending angle, refractivity, temperature, humidity, and geopotential heights together with error descriptions.

The ROM SAF also maintains the Radio Occultation Processing Package (ROPP) which contains software modules that aid users wishing to process, quality-control and assimilate radio occultation data from any radio occultation mission into NWP and other models.

The ROM SAF Leading Entity is the Danish Meteorological Institute (DMI), with Cooperating Entities: i) European Centre for Medium-Range Weather Forecasts (ECMWF) in Reading, United Kingdom, ii) Institut D'Estudis Espacials de Catalunya (IEEC) in Barcelona, Spain, and iii) Met Office in Exeter, United Kingdom. To get access to our products or to read more about the project please go to http://www.romsaf.org.

Intellectual Property Rights

List of Contents

Abstract 4
1. Introduction 5
2. Support Vector Machine methodology for reflection detection 7
3. Validation of the methodology 11
4. Computing cost, time and requirements 14
4.1 FLAG: 1-BIT OR MULTIPLE BITS? 14
5. Understanding the nature of a RO-reflection 15
6. Summary 27
7. References 29

Abstract

This document presents the implementation of a technique intended to identify the presence of reflected signals in GNSS radio-occultation (RO) data. Typically, the reflection occurs at the Earth surface level, or off atmospheric layers very close to the surface. The technique is based on supervised learning methods, in particular on Support Vector Machines (SVM). The report briefly describes the SVM approach, and how it is implemented for detection of reflected signals in RO data. Other aspects described in this document are the validation of the technique and computing requirements to run it. The document includes a preliminary discussion about the nature of these reflection events, based on the statistics of occurrence over land and ocean, seasonal features, analysis of correlation with certain atmospheric or RO conditions, and comparative performance of the ECMWF model when reflections are captured or not.

1. Introduction

This document presents an application of the recognition of patterns in the time-frequency domain, in the field of the satellite radio-occultation (RO) observations. The radio-occultation technique uses a radio-link setting below (or rising above) the horizon to inspect the properties of the transmission medium along its vertical profile. In particular, we focus on limb observations of the Global Navigation Satellite System (GNSS) signals, such as the Global Positioning System (GPS), received by Low Earth Orbiters, for atmospheric retrievals. Figure 1 sketches the geometry of the GNSS-RO concept.

The GNSS-RO missions are growing in number of satellites and scientific and operational users. Nowadays, up to 9 satellites have an active GNSS limb-sounder, producing of the order of 3000 occultations per day distributed all over the Globe. Among these occultation events, 40-60% currently present reflected signal mixed up with the direct radio-link data. This percentage is expected to rise with the new receivers being deployed, with capability to gather data down to very low altitudes and higher antenna gains.

The reflection can be seen as a secondary ray, crossing the atmosphere in a different geometry than the direct ray path. It might thus contain different geophysical content, complementing the direct link observation. This could include information about the atmosphere (lowest part), as well as a source of information about the reflecting surface. This new field of research *requires a quick and reliable identification procedure of the reflection*, potentially able to work in near real-time operational mode.

This document first presents the Support Vector Machine methodology for reflection detection. Then a section with the validation results of this methodology. Another section explains the computing costs, time and software requirements, and a discussion about the use of just one bit or multiple bits for the reflection flag. Finally, there is a description of the nature of the reflection flag.

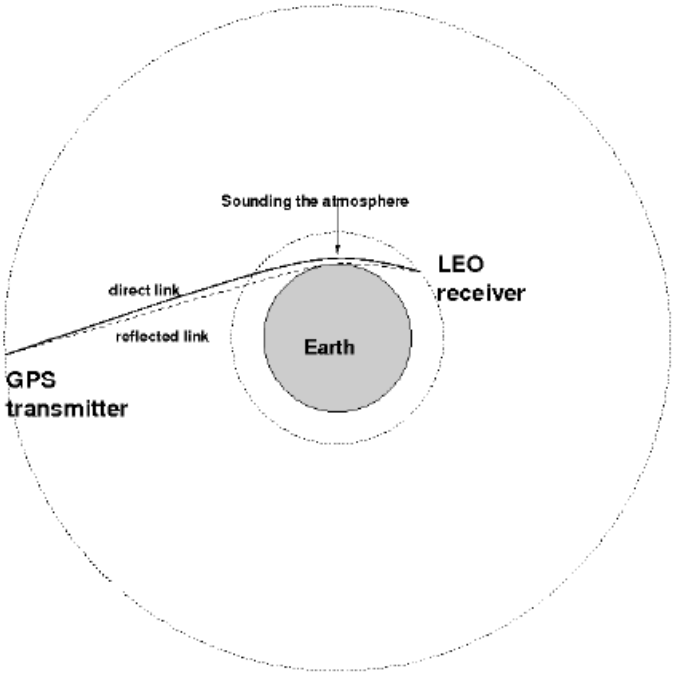

Figure 1: Sketch of a RO observation (not scaled). The reflected radio-link is also pictured (dashed line).

2. Support Vector Machine methodology for reflection detection

As identified in [Beyerle et al. (2002)], reflected signals captured in GNSS occultation data present a characteristic feature in their frequency spectra, of different Doppler frequency than the main direct radio-link one. Along time, this spectrum takes the shape of a horn smoothly approaching to the direct central frequency (in setting occultations, departing from it in rising occultations). This pattern is usually clear and easy to identify by human visual inspection (see an example in Figure 3-left). In order to analyze a statistically significant amount of occultation events, a software algorithm to replace the human visual inspection becomes imperious. During the GRAS SAF CDOP phase, several attempts were made. We could think of techniques based on physical models (e.g. to model the expected features of the reflected signals in the spectral domain), combined or not with statistical discriminators (e.g. mean power along candidate time-frequency lines). These techniques are generally good to locate the reflected signature across the radio-hologram, but present a lot of difficulties to discern (to take a decision) about the presence or not of a reflection event (large amount of missed reflections, and/or false positive ones). For these reasons, alternative approaches were tested during the GRAS SAF CDOP.

Some kernel-based learning methods have been proven to succeed in tasks of pattern recognition, that is: to take a decision about the presence or absence of a given pattern. This is exactly the problem we need to solve. We have chosen the so called Support Vector Machines (SVM, e.g. [Vapnik (1998)], or [Cristianini and Shawe-Taylor (2000)] as introductory reference).

SVMs are supervised learning methods based on optimization theory acting on a hypothesis space of linear functions. The resulting optimal hypothesis defines/separates classes in a high dimensional feature space. Since in many real-world problems the samples to be classified cannot be linearly separated in their original space, a Kernel function is used to map them into another high dimensional feature space, in which they do (see Figure 2). In our case of study, the original sample space is the P-dimensional space of the radio-holographic images to be classified (being P the total number of pixels of the radio-hologram, x, $x \in R^P$), while the feature space is the space to where the images are mapped with an exponential kernel Φ (we used the called *radial bases function* [Joachims (2001)]). In this feature space, the hypothesis space comprises the linear functions f ($\Phi(x)$), f: $X \in R^P \rightarrow R$, such that define sub-spaces of X of positive and negative values of f (binary-setting: reflection/no reflection in the radio-hologram x for positive/negative values of f ($\Phi(x)$)). The

algorithms implemented were extracted from [Joachims (2001)], originally designed for text classification. It is a quick algorithm that can classify 16 occultations per second in a Pentium IV processor.

The training set is defined as the group of radio-holographic images with their corresponding human inspected classification tag (1/−1 for reflection/no reflection respectively), used to run the learning algorithms (optimization of the hypothesis space of the linear functions) and therefore obtain the hyper-plane that generally separates the events with reflected signals from those without reflection. Once the hyper-plane is found it defines a model f* such that can be applied to any other arbitrary radio-holographic image x to determine whether f* (Φ(x)) > 0 (reflection) or f* (Φ(x)) < 0 (no-reflection). The division itself is not perfectly defined (gaps between training samples, zones of mixed classifications...), but some margin must be established (thickness of the hyper-plane with uncertain classification). The SVM algorithms also estimate the margin associated to the model, and renormalize the system so the resulting output gives values between ±1 when the samples under analysis lie within the margin area.

We have used RO events obtained by the COSMIC mission, including their open-loop data to reach the lowest possible altitudes. The training has been made using 6468 events, 57% of them with reflected signals (visually inspected). The dimension of the images is P = 201 × 36 = 7236. The events were selected so their radio-holograms had either a clean reflection, or a clear lack of reflection. Note that some occultations show unclear reflection-like features (either too noisy, or mixed with other phenomena–probably atmospheric conditions and/or instrumental issues). Examples of the three sorts of events are displayed in Figure 3. The training avoided the unclear cases and we assumed that the resulting model f* of the SVM would determine by itself whether each of them looks closer to a reflection or closer to a non-reflection case within the margin zone. It is also worth mentioning that the algorithm identifies reflection-features, but cannot distinguish whether the reflecting surface is the Earth body (ice, ocean, land) or an atmospheric interface close to the limb of the Earth.

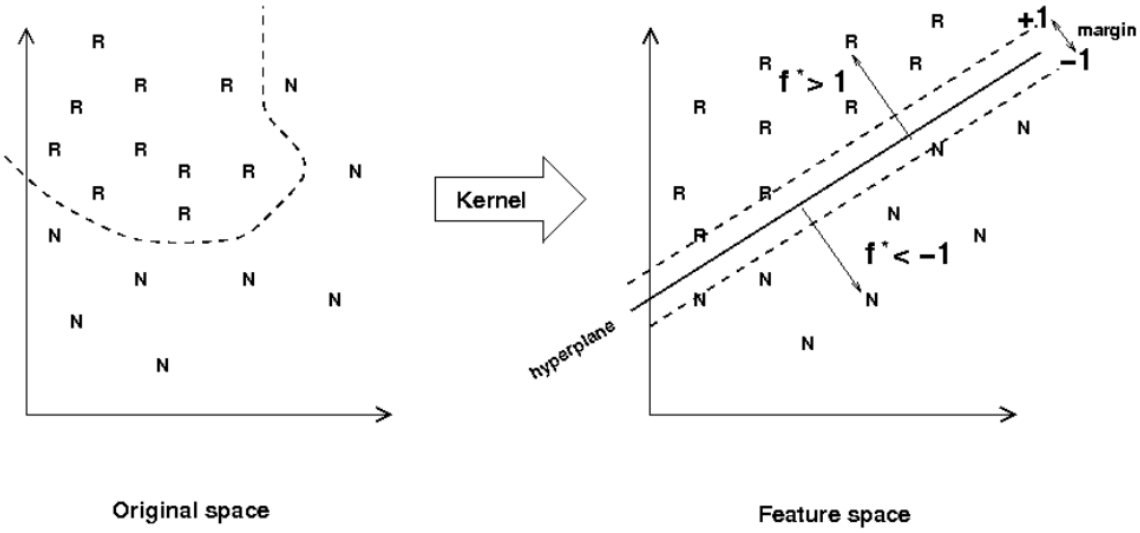

Figure 2: Sketch of the methodology behind SVM, reduced in number of dimensions for better comprehension. The original space comprises the samples to be classified. The separation between classifications might not be linear in the original space. A Kernel operation maps it into another space, the feature space, in which the classification is defined by linear functions. The margin defines the area surrounding the hyper-plane in which the classification is uncertain–data gaps in the training set or mixed cases, so the final SVM algorithms renormalize the system to the measure of the margin thickness. Then, any SVM evaluation yielding values between ±1 lie in the uncertainty zone.

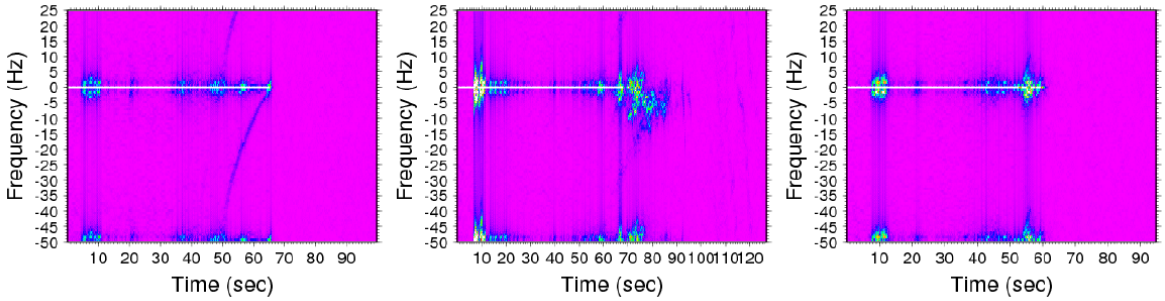

Figure 3: Radio-holographic images of three examples of occultation events visually classified as clear reflection, unclear reflection, and clear no-reflection (left to right respectively). The training set is compound by clear cases (left, right) solely. Unclear events (center) tend to be classified within the margin zone by the SVM (flag between ±1). These examples are taken from COSMIC data–including open-loop–for day 213 in 2006. Frequencies below -25 Hz are simply a copy of the range 0-25 Hz, to unfold possible aliases–as observed along the reflection horn.

3. Validation of the methodology

The model generated by the training has been evaluated in a three step sequence:

• **Evaluation of the training set:** this is a trivial step, but necessary to check that the fundaments of the algorithm work correctly. The result is 100% correct, and the SVM algorithm gives values between -3 and -1 for non-reflection events, and 1 to 3.1 for reflection events. In 99.97% of the cases the flag lies out or at the edge of the margin.

• **Evaluation of the set of events we considered unclear after visual inspection:** The SVM was then used to classify 5690 occultations that had been identified as unclear during the visual inspection (and not included in the training set). The flag values mostly lie within the margin zone: in 90% of the cases. This is consistent with the visual inspection.

Moreover, part of the set of unclear events (4750 occultations) were visually re-inspected to refine the classification (closer to reflection? or to non-reflection?), and also evaluated with the SVM. The overall success is at 65%-level, a satisfactory result since there is large subjectivity in the visual re-inspection (it is not always easy to decide whether a reflection is present or not), and most of the cases lie in the margin zone (90%). The percentage of success rise to 75% in those cases of flag beyond ±0.5.

• **Evaluation of an independent set of events:** A set of occultations not used during the training process is evaluated with the SVM, and also visually inspected to flag them as clear-reflection, clear-no-reflection, unclear. The set of clear events (1666 cases, 64% with reflection) are correctly classified by the SVM at 94%-level. Moreover, the distribution of success clearly shows that beyond the margin zone (SVM value < -1 or > 1) the detection algorithm is 99.8% confident (see Figure 4), and there is 98.5% success when $|SVM|>0.5$.

The exercises described above, as well as the training process, were performed with setting occultations only. After having solved a minor bug with rising occultations, another validation was performed, this time including rising and setting occultations. We used the same model for both types (rising/setting), extended with more training events. The new validation test was performed on 3350 RO events of February 2007 and 2257 RO events of November 2008 visually inspected and manually flagged. The validation determined that the percentage error of the SVM flag is acceptable for

values greater than 0.25 (97.81% success February 2007 sub-set, 99.47% for November 2008). These values represent an improvement of the first validation exercise explained before. Some other validation results are summarized in Table 1.

The conclusion of the validation exercises is that the SVM here implemented can be trusted when the resulting values are near or beyond the margin zone. The final threshold to be used will depend on the application. In order to extract statistical information (percentage of reflections, seasonal effects), a threshold of around ±0.5 or even ±0.25 can be used, which ensures at least 98.5% and 97.81% success respectively. For applications in which the reflection must be processed (e.g. altimetry retrievals such as in [Cardellach et al. (2004)], or atmospheric retrievals such as in [Boniface et al., 2011]), the reflected signal needs to be clean and more restrictive thresholds should be used.

Threshold	Percentage success	
	02/2007	11/2008
1	99.76	100
0.75	99.52	100
0.5	98.94	99.90
0.25	97.81	99.47
0.1	96.00	98.98
0	94.86	98.11

Table 1: Percentage success in the validation of the SVM detection algorithm, using two different sets of RO data visually inspected (3350 RO events in February 2007, 2257 in November 2008). The validation success depends on the threshold used to define a reflection event (reflection when SVM flag >threshold).

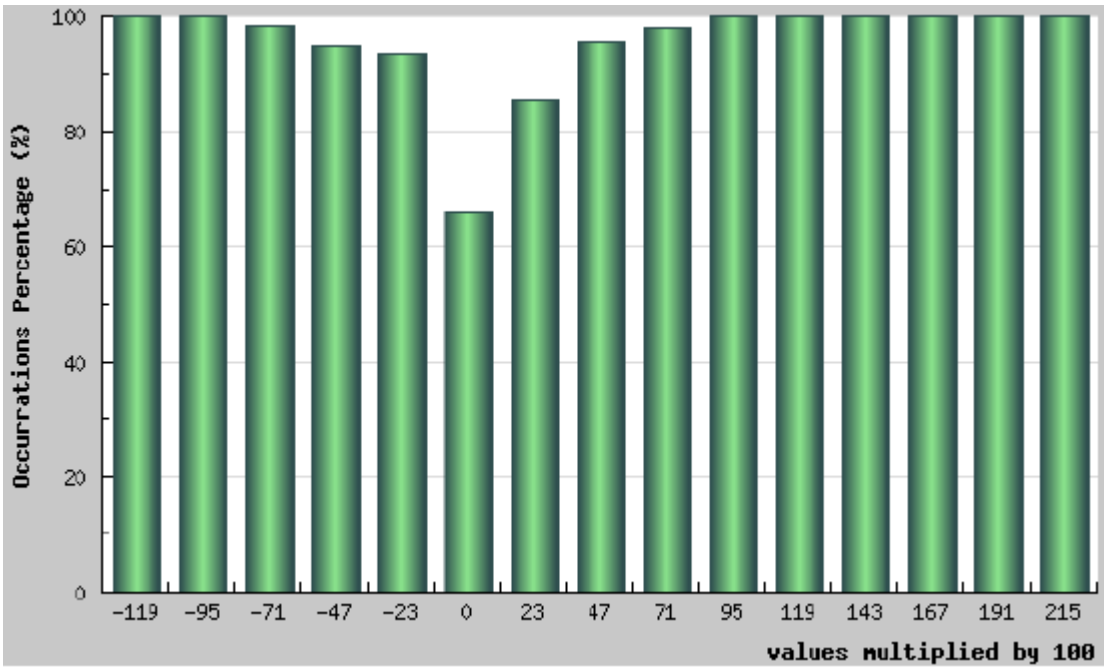

Figure 4: Evaluation of 1666 events of clear reflection/non-reflection features that were not used in the training process. On the x-axis, the SVM output value multiplied by 100 for practical reasons: positive values when the SVM considers a reflection event; negative values when the SVM do not detect the reflection; values within ±100 *lie in the margin. The y-axis presents the percentage success in that particular SVM-flag range, as compared to our visually inspected classification. It is clear that SVM only mistakes when the events lie in the margin zone, mostly within* ~ ±0.25*.*

4. Computing cost, time and requirements

Computing costs are 2/3 seconds per occultation to generate the image P to be classified by SVM. Of course this process could be parallelized with a computer with many cores. Once the images to be tested are generated, then the SVM model f* can be applied. It requires a few minutes to load the SVM model and once the model is charged in memory, the classification of the occultations is processed at an average speed of 16/occultations per second.

Therefore, if we consider the processing of 600 occultations, which could be the average occultations per day of a single GRAS receiver, they could be processed with a single processor in about 10 minutes.

The software requirements are listed below:

- Some custom scripts/programs are written in PHP and Fortran (they read the L1a occultations and generate the testing files with the vectors to be classified), and
- [SVMlight] is the software which reads the model and process the classification. This piece of software is a C compiled program (external software, free for scientific usage).

4.1 Flag: 1-bit or multiple bits?

The BUFR files have an available empty bit in which the reflection flag could be stored. Because it is one bit only, a SVM threshold should be selected (reflection when SVM>0.25? SVM>0.50? Other threshold?). As mentioned before, this threshold might change as a function of the final application. The possibility of allocating 2 or more bits would facilitate flagging reflected signals for several applications. If this is not possible, the threshold must be link to near-real time applications of NWP, where BUFR files are the standard interchange. For other off-line applications, the flag could be stored in separate files with larger precision, and distributed through the ROM SAF server.

5. Understanding the nature of a RO-reflection

A first step towards understanding the nature of reflected signals found in RO data is to analyze statistically when and where these events are captured. It is also an exercise to test the potential capabilities of the reflection-flag to map or sense certain phenomena.

A set of several months of COSMIC data has been used to look at the statistics of the reflection events, as compiled in Table 2. This represents nearly 360000 COSMIC radio-occultations globally distributed. Defining a RO-reflection event as one with SVM flag >0.25, the percentage of RO with reflections sums up to 36%. 45% of the setting occultation do present reflected signals, against 24% of the rising RO. This is an indicator that instrumental problems might hinder the reception of reflected signals. If we focus on RO's over the Oceans, 44% of them present reflected signatures. The statistics present a clear dependence on the latitude: the percentage of reflections increase with the latitude as shown in Table 3.

Month	# of RO
2006-07	11565
2006-08	22226
2007-01	29237
2007-02	48098
2007-08	11218
2007-11	41868
2008-11	57084
2008-12	71776
2009-11	64920
2009-12	1615

Table 2: COSMIC data used in this study.

Latitudes (deg)	Percentage Reflections		
	OCEAN+LAND	OCEAN	LAND
$\|lat\| \leq 20°$	22	32	8
$20° < \|lat\| \leq 50°$	41	51	12
$50° < \|lat\| \leq 70°$	53	75	14
$70° < \|lat\| \leq 90°$	72	80	64

Table 3: Percentage of reflected signals within different latitudinal belts, and reflecting surfaces. Here reflection is defined as SVM flag > 0.25.

According to these results, reflections might occur at any sort of surface (ocean, ice, land). However, land-reflections are spare and concentrate in smooth areas, free of thick vegetation, or continental ice (Figure 5). Given the electromagnetic band and the slant geometry of these observations, general lack of reflections over rugged terrain and thick vegetation canopy structures is expectable. On the other hand, the Ocean surface at slant geometries should generally behave as a smooth reflecting

surface. However, there is still a significant portion of the Ocean RO events that do not present reflected signals. Instrumental problems might trigger this. Indeed, the reflected signatures are captured at the deepest part of the occultation observation, when the direct radio-link scans the lowest part of the atmosphere, and where the tracking algorithms might have difficulties to work properly. This would explain a percentage of missing reflections in Oceanic RO data, as well as the difference in percentage reflections between rising and setting occultations. However, as displayed in Figure 6, the percentage of captured Ocean reflected signals strongly depends on the Latitude, and the distribution figures also present seasonal features in mid-latitudinal belts. These geo-physical patterns do not seem to relate to instrumental problems directly, and other explanations must be addressed to understand the nature of the RO-reflection events.

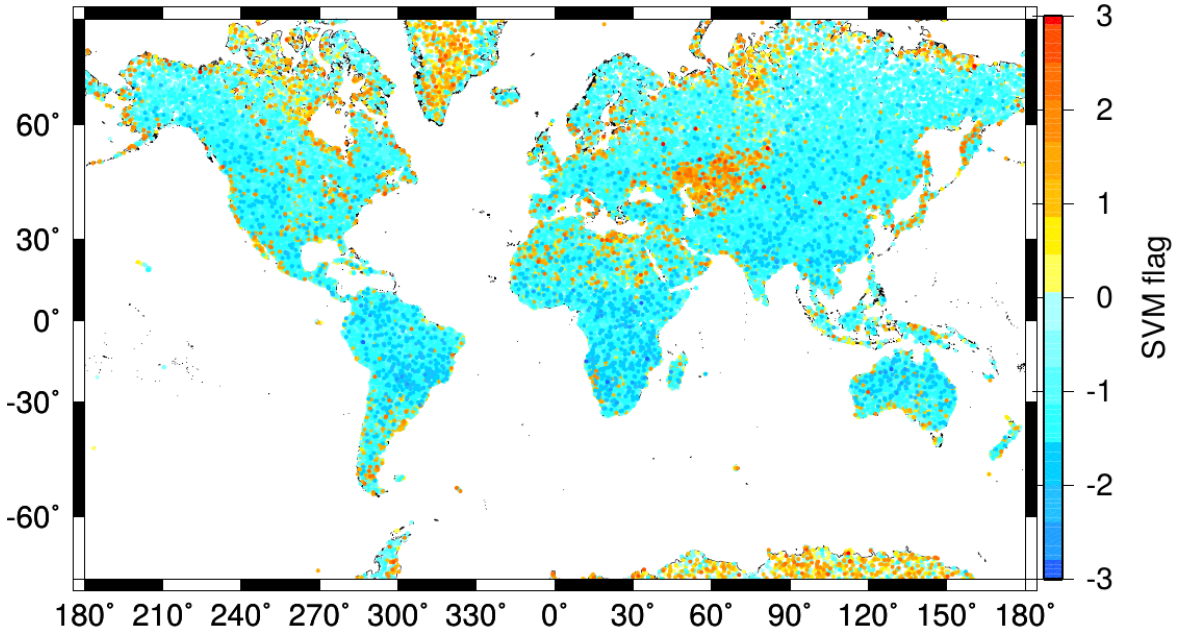

Figure 5: ~70,000 land RO events: reflection might occur over smooth terrain areas, free of thick vegetation (deserts, tundra, grasslands), or over continental ice.

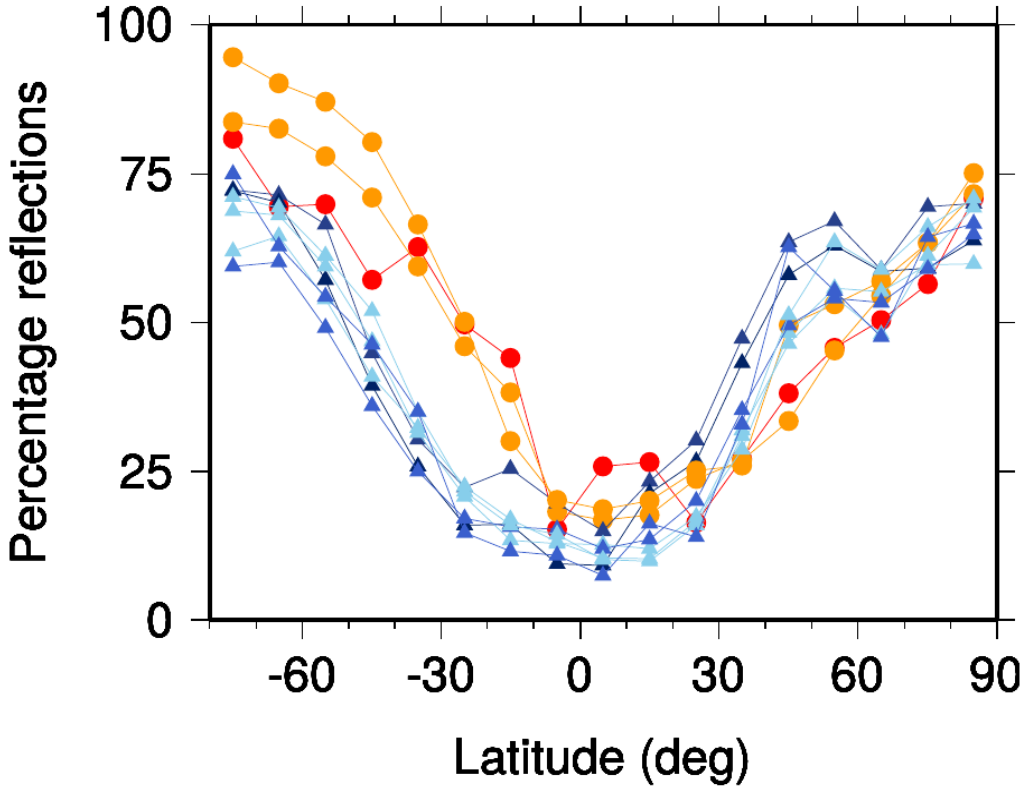

Figure 6: Statistics of reflection occurrence in RO events over the Oceans, obtained with the COSMIC constellation during several time-intervals (2006-07; 2006-08; 2007-01; 2007-02; 2007-08; 2007-11; 2008-11; 2008-12; 2009-11; 2009-12). July and August statistics are plotted with circles and warm colors, whereas November to February in triangles and cold colors. A SVM threshold of 0.5 has been considered.

The depth of the RO profile might also be related to the capacity to capture a reflection. We have checked the hypothesis that when a RO event does not reach the bot-

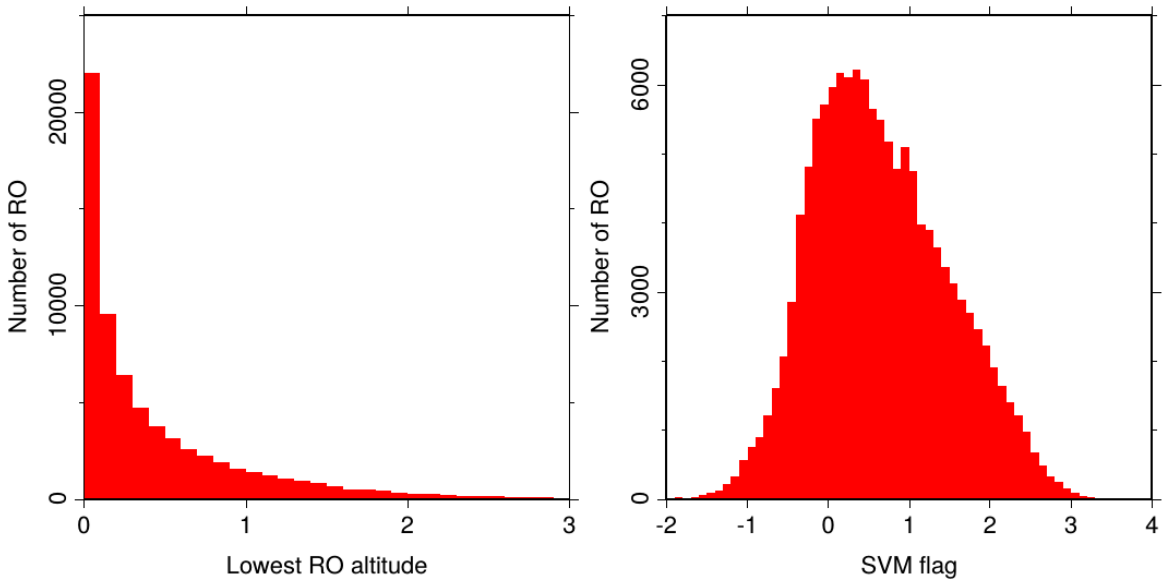

Figure 7: (left) Histogram of lowest altitude of the profile in Ocean RO events flagged with SVM > 0.5 (likely reflection). (right) Histogram of SVM values for Ocean RO events that reach the bottom kilometer of the troposphere (Hmin <1 km).

tom layers of the atmosphere it might be unable to collect the reflected branch of the signal. This is partially true according to statistics in Figure 7(left), which shows the histogram of the lowest profiled altitude in Ocean RO events with likely reflection (SVM>0.5). Most of these Ocean reflection events do indeed occur when the lowest altitude is below 2 km. However, as displayed in Figure 7(right), the histogram of SVM flag values corresponding to Ocean RO events that reach the bottom kilometer of the troposphere (Hmin <1km) do present a significant portion of negative values. These negative values correspond to non-reflection RO, in spite of reaching the bottom kilometer of the atmosphere. Therefore, reaching the bottom layers of the troposphere seems to be a condition to capture Ocean reflections, but it is not sufficient to guarantee a reflection event.

We might think that, as for land RO reflections, the surface roughness might play some role. However, at grazing angles of elevation, the effective roughness of the Ocean is always negligible, the vertical distance between the peaks and the valleys of the waves, H, effectively becoming H_eff = H sin(e), where the elevation angle e is of the order and less than 1 degree. Indeed, the Ocean surface roughness conditions do not drive the presence or lack of reflected signals, as empirically analyzed by cor-

relating the QuikSCAT scatterometer surface wind at the location/time of the RO event (daily means at 0.5° spatial resolution) and the SVM reflection flag (Figure 8(left)). The resulting correlation coefficient is r = −0.006. On the other hand, the sea surface temperature (daily observations as provided by ODYSSEA-analysis at 0.1° spatial resolution [Piolle and Autret, 2007]) does slightly anti-correlate with the SVM-flagging: high sea surface temperatures relate to more negative SVM values (lack of reflections). Although the correlation coefficient is relatively weak (r = −0.4), the anti-correlation can be appreciated in Figure 8(right).

The relationship between lack of reflections and hot water surface temperatures is also visible in Figure 9 (top and centre), where the monthly average (December 2008) of percentage reflections and sea surface temperature within cells of 4°×4° are mapped using reversed color scales. In this Figure, the features along the West coast of North and South America, as well as West South Africa and West Australia are consistent in both higher percentage of reflections and cold water surface. Similarly, the hot water features located at the North-East of Australia, central America and West North Africa are also captured by the lack of reflections in these regions. The main inconsistency is located just north of Australia, probably due to the dense distribution of islands. The reason is that we flag Ocean/Land RO based on the location of the lowest tangent point of the RO profile. However, the horizontal extent (or footprint when related to the reflection) might be up to 100 km along the observation direction. Therefore, we might have flagged (and used in these Figures) Ocean events which might have Land contamination in areas with large concentration of islands, as well areas bordering the continental lines.

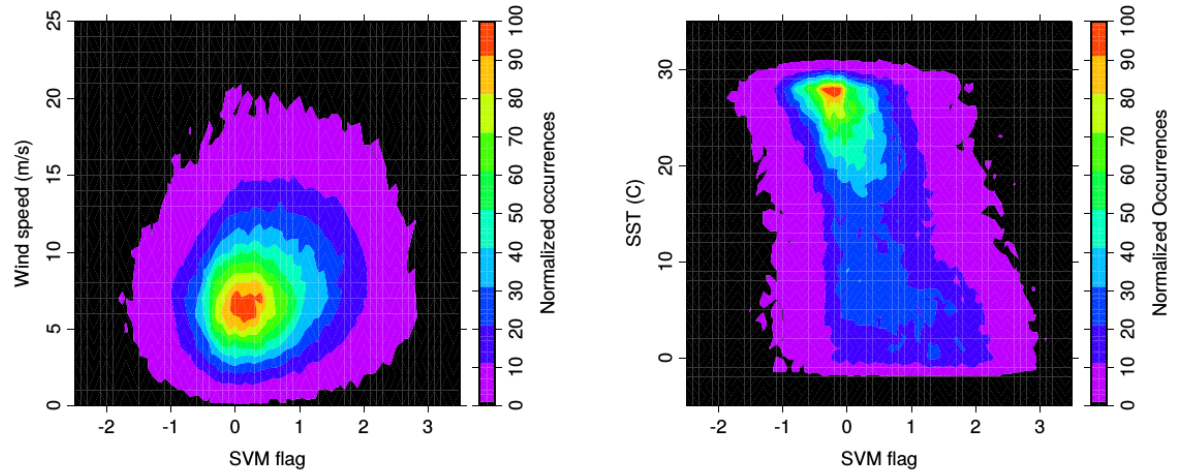

Figure 8: (left) Correlation between the SVM-flag (positive for reflection, negative for no-reflection) and QuikSCAT sea surface winds, given as normalized 2D-histogram. (right) Correlation between the SVM-flag and ODYSSEA sea surface temperature, given as normalized 2D-histogram.

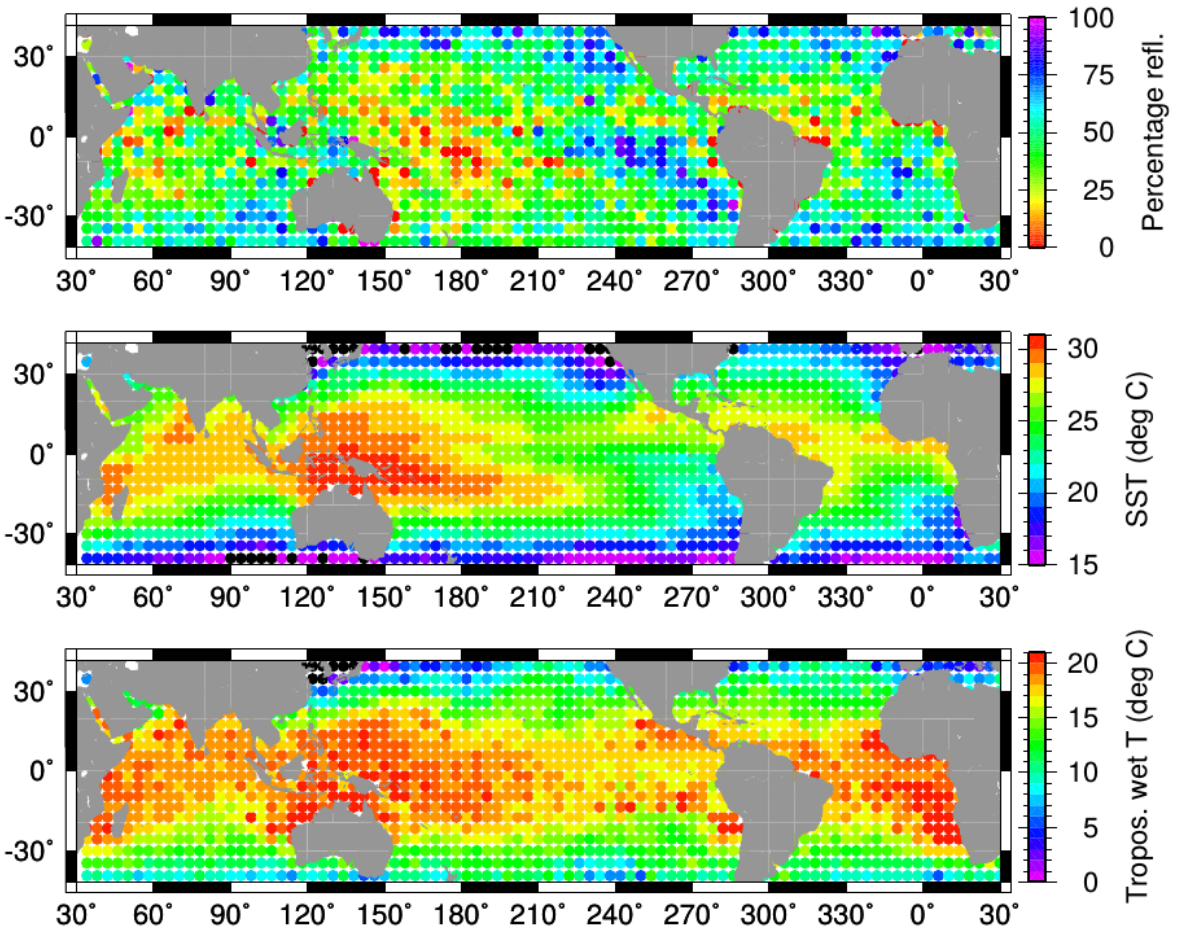

Figure 9: (top) Percentage of RO events with presence of reflected signals within cells of 4°× 4°, computed with COSMIC data from December 2008. (centre) Average sea surface temperature (ODYSSEA) within same cells and time period. (bottom) Wet temperature averaged at the bottom 2-km of the troposphere as given by COSMIC RO analysis at UCAR/CDAAC, within same cells and period.

Looking at Figures 8(right) and 9(top/centre), one might think that the temperature affects the reflectivity of the surface. This is in general true, since the Fresnel reflection coefficients depend on the surface temperature. However, at grazing angles of observation, the Fresnel coefficients for the co-polar component of circularly polarized signals (RHCP incident, RHCP reflected) do not change significantly with temperature [Ulaby, et al., 1986]: <0.1% variation between temperatures 1° to 20° C, at incidence angles greater than 80°.

A possible answer to this contradiction can be the characteristics of the air over masses of warm water. This would link the lack of reflected signals with some special tropospheric conditions. Several atmospheric parameters extracted from the same RO-profiles (as provided by UCAR/CDAAC) have been cross-correlated with the SVM flag (Figure 10):

- the bottom 5-km average of the refractivity profile,
- wet and dry temperature,
- wet and dry pressure;
- the accumulated water vapor pressure in the 5-km bottom layer;
- the minimum value of the gradient of the refractivity profile; and
- the maximum value of the direct-ray bending angle.

Although generally those are not correlated, most of them do present features that depend on the reflection flag value. The clearest correlation with the SVM flag is obtained by the averaged wet-temperature, with a moderate correlation factor $r = -0.5$. Similarly, the water vapor content at the bottom layers also presents some negative correlation ($r = -0.4$), whereas the minimum value of the gradient of the refractivity correlates with $r = 0.3$. The rest of parameters weakly correlate with the SVM, but the dispersion of all of them does seem to correlate with the reflection flag.

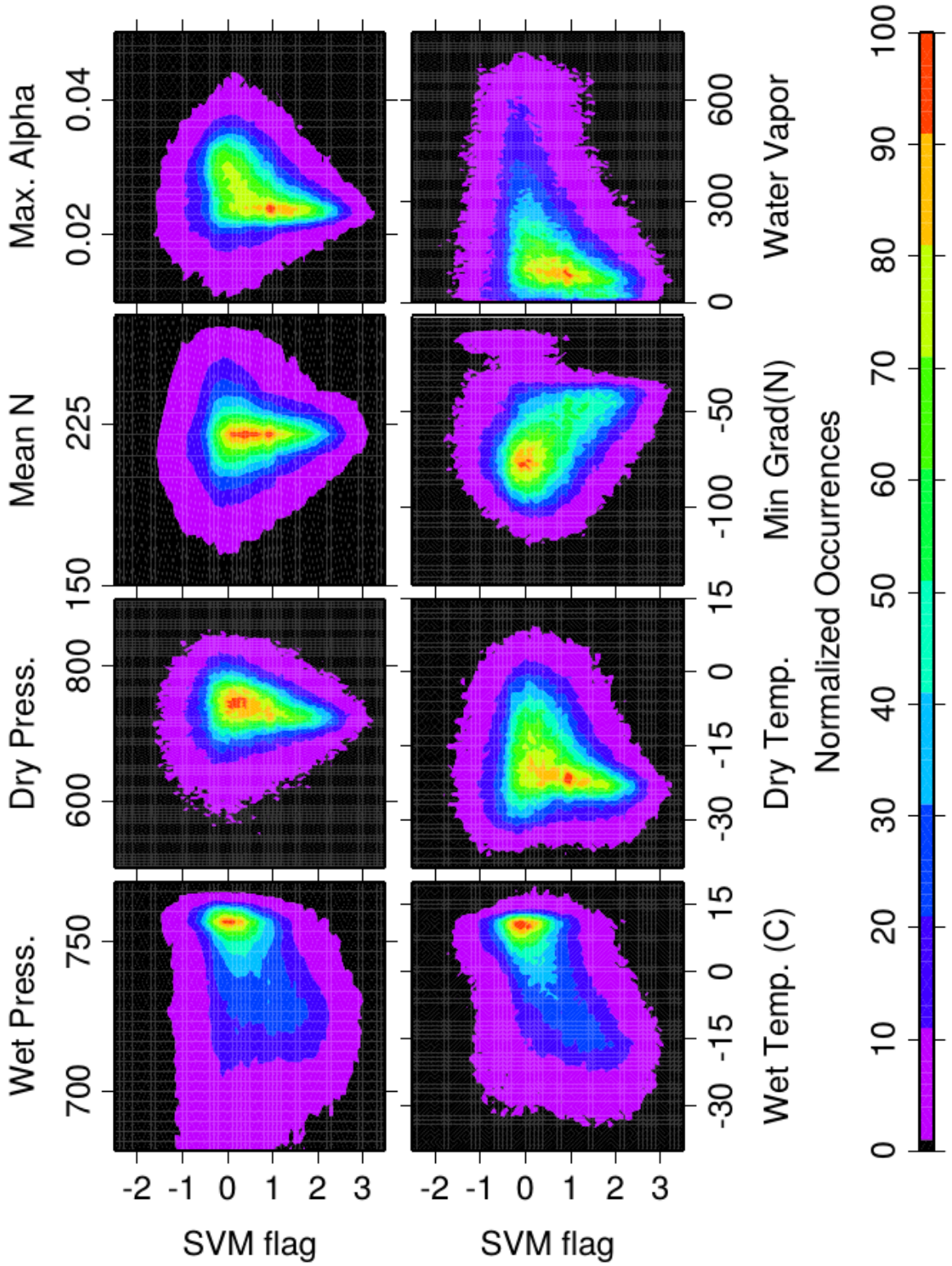

Figure 10: Correlation between the reflection SVM flag (Ocean events) and atmospheric parameters extracted from the post-processed RO profiles (averaged at the bot-tom 5-km). Refractivity given in N-units; pressures in mb; temperatures in Celsius; bending angle, α, in radians; and accumulated water vapor pressure in mb.

In the previous paragraphs we have seen that there is not a single parameter that drives the capture of the reflected signals, but a combination of tropospheric factors. [Cardellach et al. 2008] found that the comparisons between ECMWF refractivity profiles and the GNSS RO inverted ones worsen with lack of reflected signals. In other words, the combination of tropospheric factors that hinders the Ocean RO reflections seems to be the related to the atmospheric conditions that hinder ECMWF and GNSS RO profiles to converge. The work in [Cardellach et al. 2008] has been re-analyzed, increasing the number of Ocean RO events (now up to ~169,000), and the time span of the series. The results of this re-analysis confirm that both the bias and the RMS-dispersion of the error in the refractivity profile diminish significantly at the bottom 10-km of the troposphere when the Ocean RO's present reflected signals. This is shown in Figure 11, where the refractivity error has been defined as N_GPS (h) − N_ECMWF (h). Similar results are reported in [Healy (2015)].

When the comparisons are separated in latitudinal belts, the conclusion persists, although some latitudinal regions are more sensitive than others. This is compiled in Figure 12, where we can see that the improvement of the bias with reflected signals (black and dashed-black lines) occurs everywhere except at the North Polar area (where the comparisons are all very good). This improvement is more noticeable at mid latitudes (both South and North), and at the Tropics around 2 to 5 km altitude. The Tropical negative bias below 2 km does not improve with reflected signals. The error RMS-dispersion also improves with presence of reflected signals everywhere, although at the North Pole is a weak effect, and at the Tropics affects only certain altitudes (around and higher than 3 km).

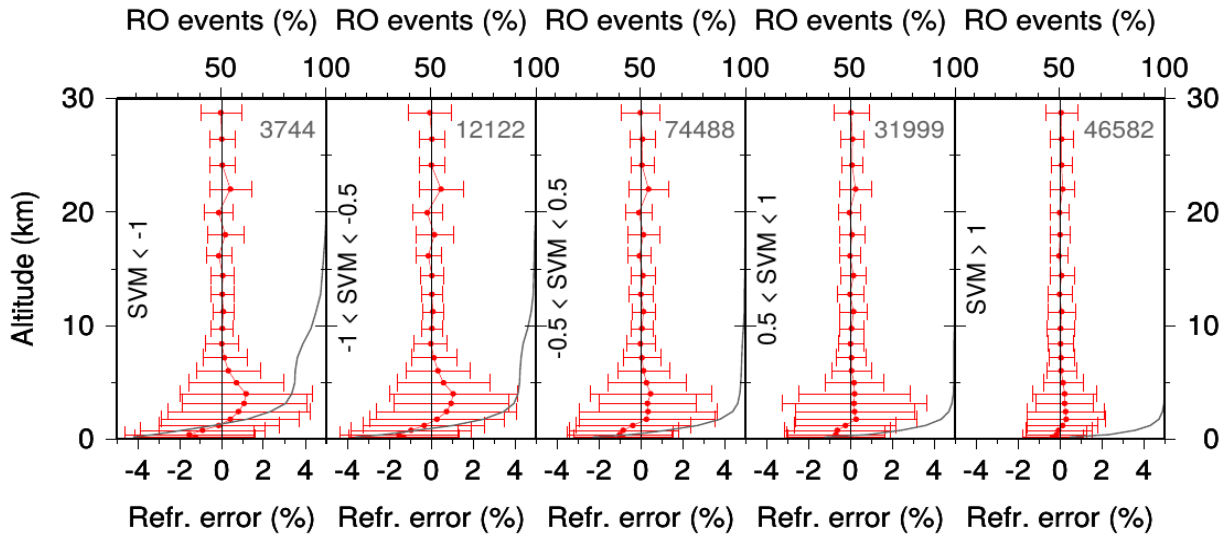

Figure 11: Statistics of the refractivity error, defined as the difference with respect to the ECMWF background (N_GP S (h) − N_ECMWF (h)), for Ocean COSMIC events (UCAR/CDAAC provided data). The statistics are separated according to the value of the SVM flag. The continuous grey line accounts for the percentage of RO used at each altitude level (right axis), and the 100% corresponds to the number printed in grey.

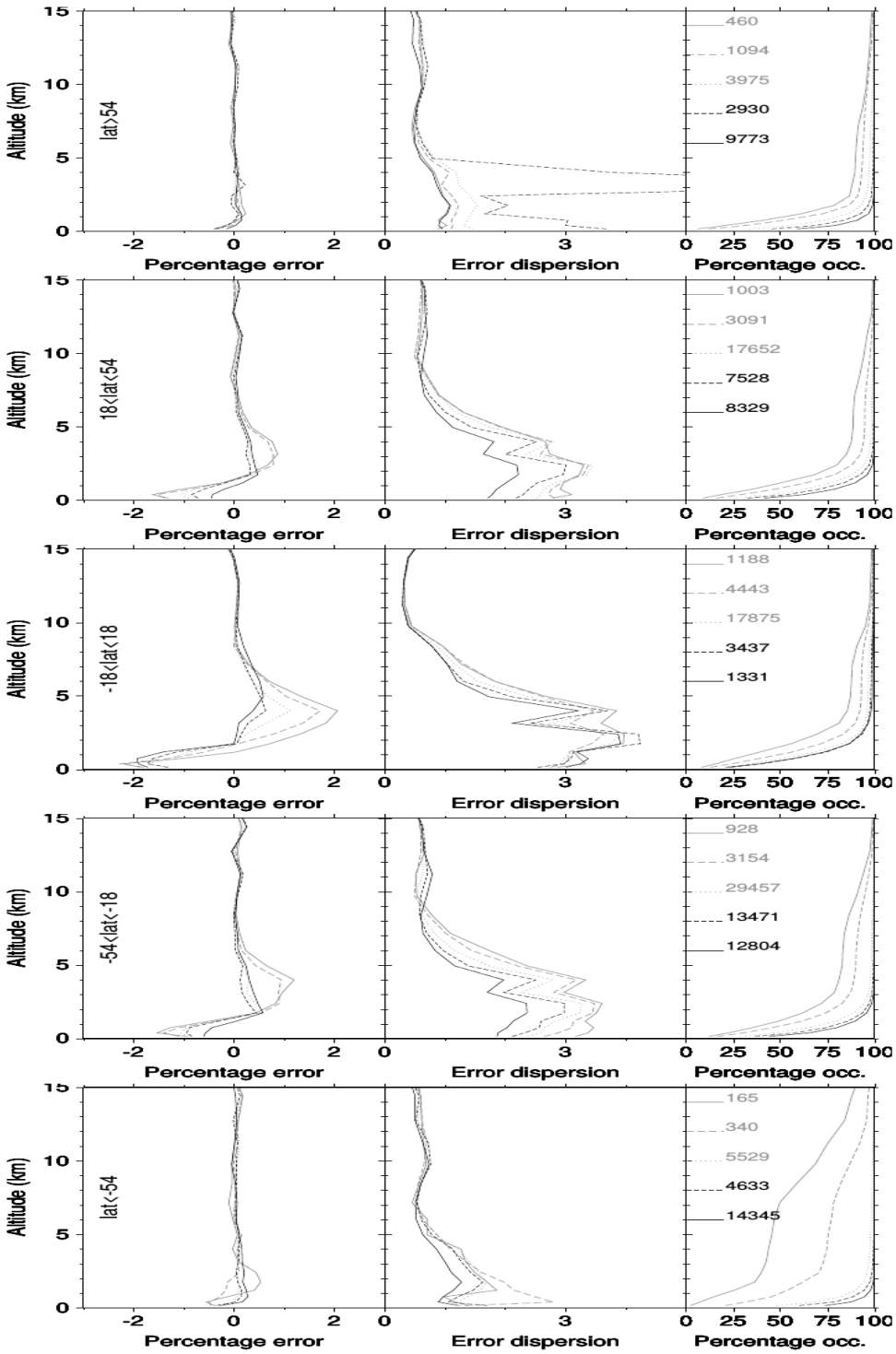

Figure 12: Statistics of (N_GPS(h) − N_ECM WF(h)) for Ocean COSMIC events (UCAR/CDAAC provided data), for different latitudinal belts (North to South from top to bottom). Each plot shows the mean, RMS-dispersion, and number of events, for 5 different ranges of SVM-values: solid-grey for SVM < -1; dashed-grey for -1 ≤SVM < -0.5; dotted for -0.5 ≤ SVM < 0.5; dashed-black for 0.5 ≤ SVM < 1; and solid-black for SVM > 1.

6. Summary

→ A supervised learning method to recognize reflection-like patterns in radio-holographic images of RO events has been selected, implemented and tested.

→ Difficulties on achieving a robust method based on physical models are the main reason behind this decision.

→ The selected method is called 'Support Vector Machine' (SVM), and it uses optimization theory acting on a hypothesis space of linear functions. A kernel is used to linearize it, and after a re-normalization, the output is a real number that indicates a clear reflection if the SVM value is larger than 1, or clear non-reflection event when SVM < -1. Values between -1 and +1 lay in the margin zone (mixed statistics or gaps in the training model).

→ The training has been made using 6468 events, 57% of them with reflected signals (visually inspected). It has been validated with more than 12,000 RO events, which we had previously inspected visually. The validation set included setting and rising cases, 'clear' and 'unclear' cases. 90% of those visually determined as 'unclear' obtained SVM values within the margin zone. ~98% of the events with SVM > 0.25 corresponded to cases for which visual inspection had determined that a reflection was present.

→ The processing chain to obtain the SVM takes a fraction of second to conclude 1 individual case, at a typical rate of 16 occultation per second, once the model has been charged into the memory (it takes a few minutes). Incorporating this analysis into a near-real time processing scheme is in principle possible, being the only complication when or how often would the system needed to load the model into the memory. This should be contrasted with the actual RO NRT processing chain and identify ways to keep this model in the memory for relatively long periods of time.

→ If 1-bit flag is allocated at BUFR files to account for presence or lack of reflection, the threshold must be carefully studied to optimally match the NWP applications. Other applications, that might require multiple bits, could be offered under different format.

→ The SVM tool has been used to analyze ~360000 COSMIC radio-occultations globally distributed, to understand when, how and/or why the reflections are or are not captured. The conclusions of this study are:

→ Reflections might occur at any sort of surface (ocean, ice, land), however, land-reflections are spare and concentrate in smooth areas, free of vegetation, or over continental ice.

→ Over extra-tropical oceans, >50% of the RO present reflected signals, increasing with latitude (>75% for latitudes above 50 degrees, and up to 80% above 70 degrees latitude).

→ Over the ocean, reflections present seasonal signatures, less percentage of reflected signals during local summer periods.

→ Over the ocean, reflection events do not correlate at all with roughness/wind condition, as at these slant geometries the effective roughness tends to zero (going as ~sin(elevation angle)).

→ Over the ocean, reflection events do correlate somehow with cold sea surface temperatures (more percentage of reflections over cold waters, e.g. south-east oceans). However, the co-polar reflection coefficient at slant angles is not temperature-dependent.

→ Over the ocean, reflection events do anti-correlate somehow with tropospheric variables related with wet atmospheric conditions.

→ RO profiles where reflected signals are present tend to present better comparisons between their RO-products and ECMWF ones.

7. **References**

[Beyerle et al. (2002)] Beyerle, G., K. Hocke, J. Wickert et al. (2002), GPS radio occultation with CHAMP: A radio holographic analysis of GPS signal propagation in the troposphere and surface reflections, J. Geophys. Res., 107(D24), 4802, doi:10.1029/2001JD001402.

[Boniface et al. (2011)] Boniface, K., J. M. Aparicio, Cardellach, E., Meteorological information in GPS-RO reflected signals, Atmospheric Measurement Techniques, 4, pp. 1397-1407, 2011, Jul, http://www.atmos-meas-tech.net/4/1397/2011/amt-4-1397-2011.pdf, 10.5194/amt-4-1397-2011

[Cardellach et al. (2004)] Cardellach, E., Ao, C. O., de la Torre Juárez, M., Hajj, G. A., Carrier phase delay altimetry with GPS-reflection/occultation interferometry from low Earth orbiters, *Geophysical Research Letters*, 31, pp. 10402-+, 2004, May, 10.1029/2004GL019775

[Cardellach et al. (2008)] Cardellach, E., Oliveras, S., Rius, A., Applications of the reflected signals found in GNSS radio occultation events, ECMWF Proceedings: GRAS SAF Workshop on Applications of GPSRO Measurements , 2008, ECMWF, Shinfield Park, Reading, Berks RG2 9AX, England, Jun, ECMWF

[Cardellach et al. (2009)] Cardellach, E., Oliveras, S., Rius, A., GNSS Signal Interference Classified by means of a Supervised Learning Method Applied in the Time-Frequency Domain, IEEE Proceedings of 2009 2nd International Congress on Image and Signal Processing, 2009, Oct, Institute of Electrical and Electronic Engineers, Tianjin, China, ISBN 13: 978-1-4244-4130-3, doi:10.1109/CISP.2009.5302135

[Cardellach et al. (2010)] Cardellach, E., Oliveras, S., Rius, A., Tropospheric Information Content Embedded in GNSS RO Reflected Signals, Proceedings of the International Beacon Satellite Symposium 2010, gAGE/UPC, 2010, June

[Cosmic] CDAAC COSMIC server: http://cosmic-io.cosmic.ucar.edu/cdaac/

[Cristianini and Shawe-Taylor (2000)] Cristianini N., and J. Shawe-Taylor (2000), An introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, ISBN 0 521 78019 5, 2000, 2006 re-printing

[Healy (2015)] Healy, S., The use of the GPS radio occultation reflection flag for NWP applications, ROM-SAF Report No. 22, Ref. SAF/ROM/METO/REP/RSR/022, available at http://www.romsaf.org

[Joachims (2001)] Joachims, T. (2001), Learning to classify text using support vector machines: Methods, Theory, and Algorithms, Kluwer Academic Publishers, ISBN 0 7923 7679 X.

[SVMlight] SVMlight official webpage: http://svmlight.joachims.org/

[Vapnik (1998)] Vapnik, V. (1998), Statistical Learning Theory, Wiley, Chichester, GB.

[Piolle and Autret, 2007] Piolle,J.F., and E. Autret (2007), ODYSSEA Global SST Analysis - User manual, MERSEA-WP02-IFR-STR-001-1A, CERSAT-IFREMER,

http://www.mersea.eu.org/Satellite/MERSEA-WP02-IFR-STR-001-1A.doc

[Ulaby et al., 1986] Ulaby, F.T., R.K. Moore, and A.K. Fung, Microwave Remote Sensing, Active and Passive, Vol.III, Arctech House, Inc., 1986, ISBN 0-89006-192-0, Norwood, MA.

ROM SAF (and GRAS SAF) Reports

SAF/GRAS/METO/REP/GSR/001	Mono-dimensional thinning for GPS Radio Occultation
SAF/GRAS/METO/REP/GSR/002	Geodesy calculations in ROPP
SAF/GRAS/METO/REP/GSR/003	ROPP minimiser – minROPP
SAF/GRAS/METO/REP/GSR/004	Error function calculation in ROPP
SAF/GRAS/METO/REP/GSR/005	Refractivity calculations in ROPP
SAF/GRAS/METO/REP/GSR/006	Levenberg-Marquardt minimisation in ROPP
SAF/GRAS/METO/REP/GSR/007	Abel integral calculations in ROPP
SAF/GRAS/METO/REP/GSR/008	ROPP thinner algorithm
SAF/GRAS/METO/REP/GSR/009	Refractivity coefficients used in the assimilation of GPS radio occultation measurements
SAF/GRAS/METO/REP/GSR/010	Latitudinal binning and area-weighted averaging of irregularly distributed radio occultation data
SAF/GRAS/METO/REP/GSR/011	ROPP 1D-Var validation
SAF/GRAS/METO/REP/GSR/012	Assimilation of Global Positioning System Radio Occultation data in the ECMWF ERA-Interim re-analysis
SAF/GRAS/METO/REP/GSR/013	ROPP PP validation
SAF/ROM/METO/REP/RSR/014	A review of the geodesy calculations in ROPP
SAF/ROM/METO/REP/RSR/015	Improvements to the ROPP refractivity and bending angle operators
SAF/ROM/METO/REP/RSR/016	Simplifying EGM96 undulation calculations in ROPP
SAF/ROM/METO/REP/RSR/017	Simulation of L1 and L2 bending angles with a model ionosphere
SAF/ROM/METO/REP/RSR/018	Single frequency radio occultation retrievals: impact on numerical weather prediction
SAF/ROM/METO/REP/RSR/019	Implementation of the ROPP two-dimensional bending angle observation operator in an NWP system
SAF/ROM/METO/REP/RSR/020	Interpolation artefact in ECMWF monthly standard deviation plots
SAF/ROM/METO/REP/RSR/021	5th ROM SAF User Workshop on Applications of GPS radio occultation measurements
SAF/ROM/METO/REP/RSR/022	The use of the GPS radio occultation reflection flag for NWP applications

ROM SAF Reports are accessible via the ROM SAF website http://www.romsaf.org