Applications of the Reflected Signals Found in GNSS Radio Occultation Events

E. Cardellach, S. Oliveras and A. Rius

ICE-CSIC/IEEC
Campus UAB, Bellaterra, Spain
estel@ieec.uab.es

ABSTRACT

Simulation and statistical empirical work has been done to show that the presence of reflected signals in GNSS occultations is an indicator of good data quality, resulting in atmospheric profiles that better match the post-analyzed ECMWF background. In order to conduct this work, an automatic tool to detect reflected signals in Level-1 data has firstly been implemented, and 102118 occultations classified according to its reflection content. Geographic and seasonal effects are also noticeable.

1 Motivation

The GNSS Radio-Occultation 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% present reflected signal mixed up with the direct radio-link data. And as more instruments are able to operate in open-loop, the amount of reflected features increases (sounding down to lower altitudes).

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 the reflecting surface. Being this work conducted under the GRAS SAF Continuous Development and Operational Phase, we will initially focus on any application of the reflected signals that might enhance the operational use of the GRAS-like observations for weather (data assimilation) and climate monitoring.

As shown in Section 2, the work under progress seems to indicate that the reflected signals do not disturb the standard retrievals of atmospheric refractivity. On the contrary, the work we present under Section 4 might suggest that the detection of reflections are a sign of clean observations, in which neither instrumental, nor complex atmospheric phenomena are affecting the measurements. If proven, this could become a useful tool for data assimilation purposes, in the form of a quality flag for rejection of data, or even as an indicator of the weight to be given to each occultation.

With these goals in view, the steps taken so far to quantitatively investigate the correlation between reflection events and quality of the retrieved profiles are the following: (a) to develop, implement, and validate a tool capable to automatically detect the presence of reflected features in the level-1 data; (b) to identify the presence/absence of reflected features in a statistically significant number of events; and (c) to analyze the quality of the occultation-derived refractivity profiles and as function of the reflection flag. These steps are presented in Sections 3 and 4.

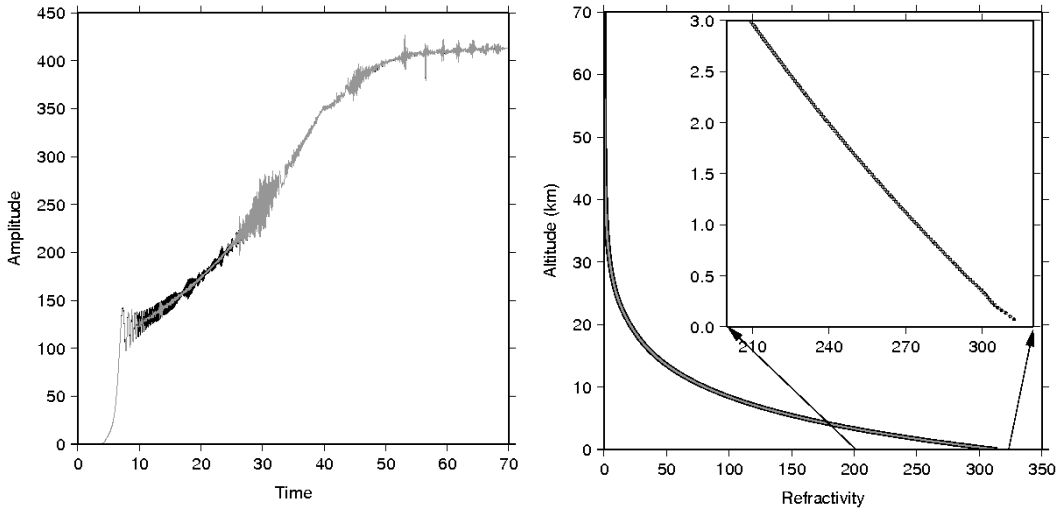

Figure 1: (left) Amplitude of the original data (black) and the clean data (gray) with no interferences due to the reflected signal (high frequency low amplitude fringes disappear in the clean data set). The cleaning process affects similarly the phase delay. (right) Refractivity profile retrieved from the original and clean data sets (black and gray respectively). Zoom-in at the lowest troposphere. The differences are negligible.

2 Do reflections contaminate the atmospheric retrievals?

During the last seconds of a setting occultation (or beginning of rising event), in which reflected signals are also present, they coherently sum with the direct-link electromagnetic field presenting amplitude and phase interference patterns. The patterns might be clear enough to extract the phase-delay information between the reflected and the direct links, and use it for surface altimetric purposes ([Cardellach et al. (2004)]). But, how these interferences affect the retrieval of the atmospheric profiles?

End-to-end simulation work has been done to solve this question: a multi-phase screen model (MPS) with a given and well-known refractivity profile has been run to produce the level-1 data set (time-amplitude/phase) received at a LEO orbit. The MPS model used a reflecting surface, so reflected features are shown in the level-1 synthetic data (we call this data set `original`). Another set is generated by *cleaning* the `original` data, we call it `clean`. The *cleaning* is performed by: (1) Fourier analysis in running sliding windows to generate the radio-holograms; (2) reduce the reflected features to noise level; and (3) inverse Fourier transform back to the time-amplitude/phase domain. The Canonical Transform (CT) algorithms are then applied to the two level-1 synthetic data sets (`original` and `clean`) to obtain the estimated atmospheric profiles.

As seen in Figure 1, the refractivity profiles obtained out of the two data sets are almost identical, with negligible differences ($<0.2\%$) from the application point of view. Similarly happens with dry temperature and bending angle retrievals. This confirms that reflected features, interference fringes of high frequency fluctuations and small amplitude, do not disturb or modify the results of the atmospheric retrievals when using CT-like techniques.

3 Automatic detection of reflected signals

As identified in [Beyerle et al. (2002)], a characteristic feature of the reflected signals present in GNSS occultation data is their frequency spectra, of different Doppler than the main direct link. Along time, this spectra 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 visually identify by human 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. Some kernel-based learning methods have been proven to succeed in tasks of pattern recognition. We have chosen the so called Support Vector Machines (SVM, e.g. [Vapnik (1998)], or [Cristianini and Shawe-Taylor (2000)] as introductory reference), out of the scope of this paper but briefly explained to help understanding the results.

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 (Figure 2). In our case of study, the original samples 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, $\vec{x}$, $x \subseteq \mathscr{R}^P$), 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 $f(\Phi(\vec{x})), f : X \subseteq \mathscr{R}^P \rightarrow \mathscr{R}$, such that define sub-spaces of X of positive and negative values of f (binary-setting: reflection/no reflection in the radio-hologram $\vec{x}$ for positive/negative values of $f(\Phi(\vec{x}))$). 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 $\vec{x}'$ to determine whether $f^*(\Phi(\vec{x}')) > 0$ (reflection) or $f^*(\Phi(\vec{x}')) < 0$ (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.

At this stage of the study setting-occultations have been used solely, all from the COSMIC mission, and including the open-loop data to reach the lowest possible altitudes. The training has been made using 6468 events, 57% of them with reflected signals. The dimension of the images is $P = 201 \times 36 = 7236$. 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 and/or instrumental). Examples of the three sort of events are displayed in Figure 3. The training avoided these cases and assumes the resulting model f^* of the SVM will determine by itself whether each of them looks closer to a reflection or closer to a non-reflection. 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.

3.1 Validation

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

- **to evaluate 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.
- **to evaluate the set of events we considered unclear after visual inspection:** The SVM classified 5690 occultations that had been identified as unclear by visual inspection. The flag values mostly lie within the *margin* zone: in 90% of the cases. This is consistent with the visual inspection.

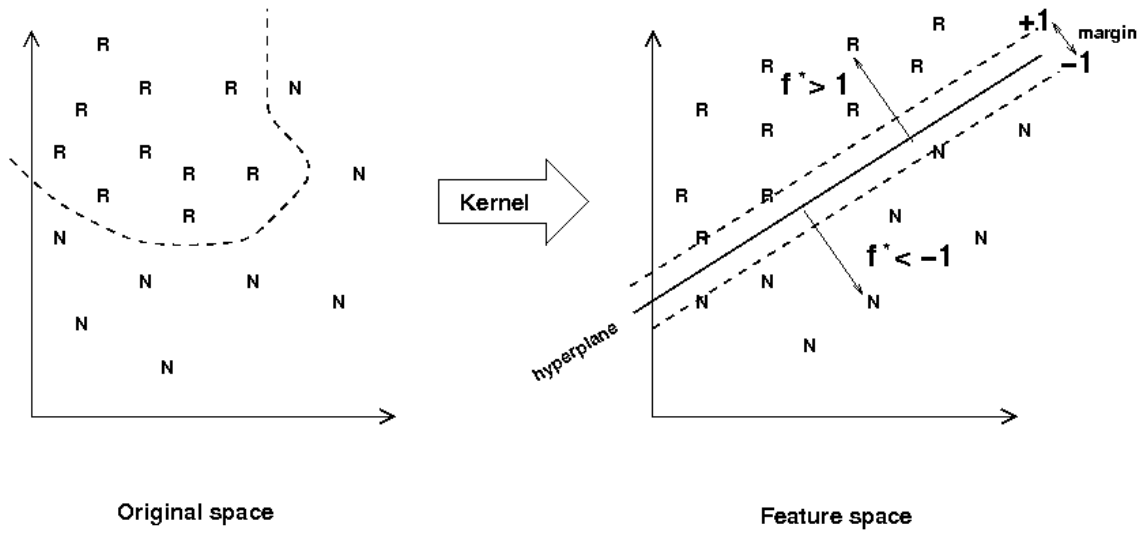

Figure 2: Sketch of the methodology behind SVM, reduced in number of dimensions for better comprehension. The original space comprises the samples to be classified. The separation between classifications might not be linear in the original space. A Kernel operation maps it into another space, the feature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

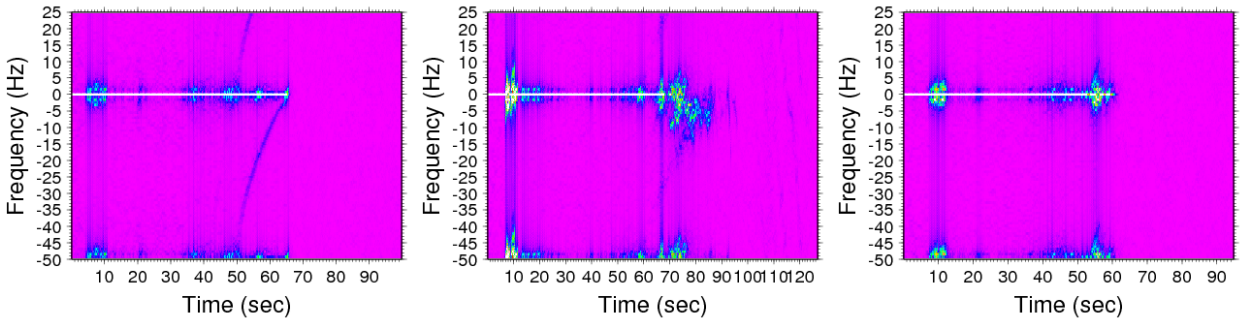

Figure 3: Radio-holographic images of three examples of occultation events visually classified as clear reflection, unclear reflection, and clear no-reflection (left to right respectively). The training set is compound by clear cases (left, right) solely. Unclear events (center) tend to be classified within the margin zone by the SVM (flag between ± 1). 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.

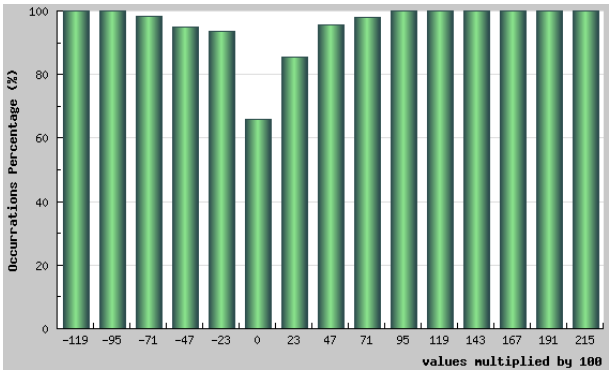

Figure 4: Evaluation of 1666 events of clear reflection/non-reflection features that were not used in the training process. On the x-axis, the SVM output value multiplied by 100 for practical reasons: positive values when the SVM considers a reflection event; negative values when the SVM do not detect the reflection; values within ±100 lie in the margin. On the y-axis 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, in particular within ∼ ±0.8.

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 (not always is 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.

- **to evaluate 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 show that beyond the *margin* zone (SMV value < -1 or > 1) the detection algorithm is 99.8% confident (see Figure 4), and there is 98.5% success out of the range $(-0.5, 0.5)$.

The conclusion of the validation 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 will be used (Section 3.2), which ensures 98.5% of success. For applications in which the reflection must be processed (e.g. altimetry retrievals such as in [Cardellach et al. (2004)]), the reflected signal needs to be clean and more restrictive threshold should be used.

3.2 Some statistical results

The SVM trained with 6468 occultations, and validated as described in the preceding Section, has been used to classify 102118 COSMIC setting events, including open loop data. 60% of the cases are out of the reduced margin (SVM flag threshold ±0.5–used in the rest of the Section).

The statistics show that a large percentage of the total study set presents reflected signatures (40%). The geographic distribution of the reflections (Figure 5-left) is significant around the poles, over ice. But also present in Ocean waters, including the Tropics. Among all the occultations occurring over the Oceans, 48% present reflected signals, although the percentage is much higher in extreme latitudes (up to 81% over Southern Ocean, or around 75% in latitudes below -30°). Over land, the reflections seldom occur, 7% solely–excluding Antarctica and Northern Polar land.

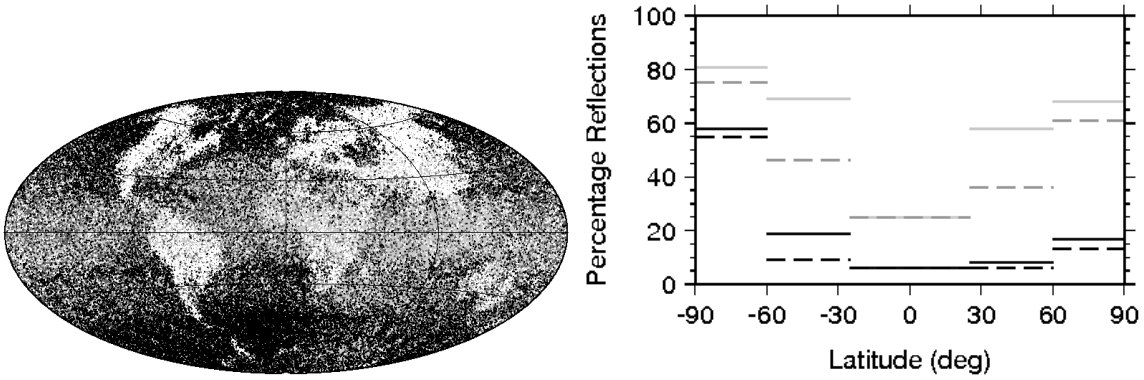

Figure 5: (left) 102118 occultations classified with the SVM: reflection (black), no-reflection (white), and unclear (gray). (right) Percentage of reflection events in separate latitude-blocks. Gray for Ocean occultations, black for land occultations. Solid lines for local winter period, dashed lines for local summer period.

The 102118 events correspond to two seasonal periods: January-February 2007 (57109 cases), and July-August 2006 (45009 cases). Over the Oceans, the percentages of reflected signals present strong seasonal signatures (Figure 5-right), especially in mid latitudes (variation of 23 percentage points). Over land, the seasonal effects are relatively weak, and probably linked to ice presence. In every region beyond Tropics, the reflections increase during winter time.

4 Reflected signals and quality of the profiles

On one hand, the atmospheric retrievals of refractivity, bending angle and dry temperature are not contaminated by the reflected signal, as seen in Section 2. On the other hand, Section 3.2 shows that reflections over the Oceans are more likely over extreme latitudes than in Tropical zones, especially during the winter. This might indicate that the reflected signals can only reach the receiver when certain surface/atmospheric/instrumental conditions apply. The surface condition is clear: land reflections seldom occur, while Ice and Oceans are good reflectors in the occultation geometry. Therefore, over the Oceans, and assuming the sea surface characteristics do not differ significantly between the Tropics and Mid Latitudes, the instrumental and atmospheric conditions must be the limiting factors to receive reflected signals. In other words, rather than *disturbing* the atmospheric profiling, the presence of reflected signals might be a testimony of the cleanness of the data, and thus of the good quality of the atmospheric sounding. This hypothesis is inspected in this Section by means of statistical comparisons of the post-processed COSMIC refractivity profiles (Level-2) with the analyzed ECMWF background, all downloaded from the COSMIC server ([Cosmic]). These required files were available for 69359 of the 102118 SVM-classified occultations.

The comparison results are presented as statistical percentage error (mean or bias μ, and standard deviation σ) on each individual altitude layer k (percentage over the total number of occultations present in the layer M_k):

$$\mu_k = \frac{1}{M_k}\sum_{i=1}^{M_k}\frac{N_i^{GPS}(h_k) - N_i^{ECMWF}(h_k)}{N_i^{GPS}(h_k)} \tag{1}$$

$$\sigma_k = \sqrt{\frac{1}{M_k}\sum_{i=1}^{M_k}\left(\frac{N_i^{GPS}(h_k) - N_i^{ECMWF}(h_k)}{N_i^{GPS}(h_k)} - \mu_k\right)^2} \tag{2}$$

The comparisons are grouped according to the SVM classification value (called *flag* hereafter), in five different categories (detailed in Table 1). Reflections are identified by positive flags, especially beyond the *margin* ($flag \geq 1$).

Flag group ID	Range	Description
NR	$flag \leq -1$	No-Reflection
PNR	$-1 < flag < -0.5$	Possible No-Reflection
U	$-0.5 \leq flag \leq 0.5$	Unclear
PR	$0.5 < flag < 1$	Possible Reflection
R	$flag \geq 1$	Reflection
Latitude Group ID	Range	Description
SP	$-90°$ to $-54°$	South Polar zones
MS	$-54°$ to $-18°$	Mid-South zones
E	$-18°$ to $18°$	Equatorial zones
MN	$18°$ to $54°$	Mid-North zones
NP	$54°$ to $90°$	North Polar zones

Table 1: Flag groups and geographic zones used in the study.

4.1 Oceans and Sea-Ice

Ocean occultations can be defined as those the tangent point of which are located in geographic coordinates over the Ocean or Sea-Ice. 70 % of the setting occultation events are oceanic, around 48000 events in our study set. The refractivity errors with respect to the ECMWF background clearly depend on the SVM classification value, in particular at the bottom layers of the atmosphere, where both the mean error and the standard deviation improve (smaller) when reflected signals are present (see Figure 6). The bias errors below 10 km altitude are reduced to less than 1% when the flag is greater than 0.5, while $\sim$2% biases appear at $\sim$5 and 1$\sim$ km altitude (positive and negative bias respectively) when the flag is below -0.5. The positive bias at 5 km height nearly disappears when the data contains reflected signals (positive flags). The lack of reflected signals also correlates with a loss of events below 10 kilometers of altitude (also noticeable in Figure 6). The standard deviations below 10 km also reduce significantly when clear reflection is detected.

The analysis is divided into geographic zones, generating separate statistics for oceanic occultations over the five different ranges of latitude detailed in Table 1. Biases are compiled in Figure 7, showing different behavior depending on the latitude. The positive bias around 4-5 km altitude happens in Equatorial and Mid latitudes, when no reflection is present. For reflection events it is reduced by up to 2 percentage points, never taking bias errors beyond 1%. The differences between the quality of occultations with and without reflections are much smaller over the Poles, although presence of reflected signals tend to better compare with ECMWF background. The amount of events present on each layer are also plotted, useful to note when the statistics are significant (i.e., lowest kilometers of SP-NR populated with 20 or less events solely).

4.2 Land Occultations

Occultations over land represent the 30% of the total setting occultations (around 21000 events in this study case). There are a few reflection events over land (7%), when excluding the Antarctica and northern areas (latitudes > 60). The reasons for reflected signals to be or not to be present in the event are unclear, most probably related to the surface conditions. In any case, the comparisons with respect to the ECMWF background show dependence on the reflection flag value similar to but not as strong as occultations over the Oceans (Figure 8, compare with Figure 6). The bottom negative bias is gradually reduced from -3% when no-reflected signals are present, to -0.5% when clear reflections appear. A weak but positive bias around 4 km altitude solely occur when no reflected signals are found.

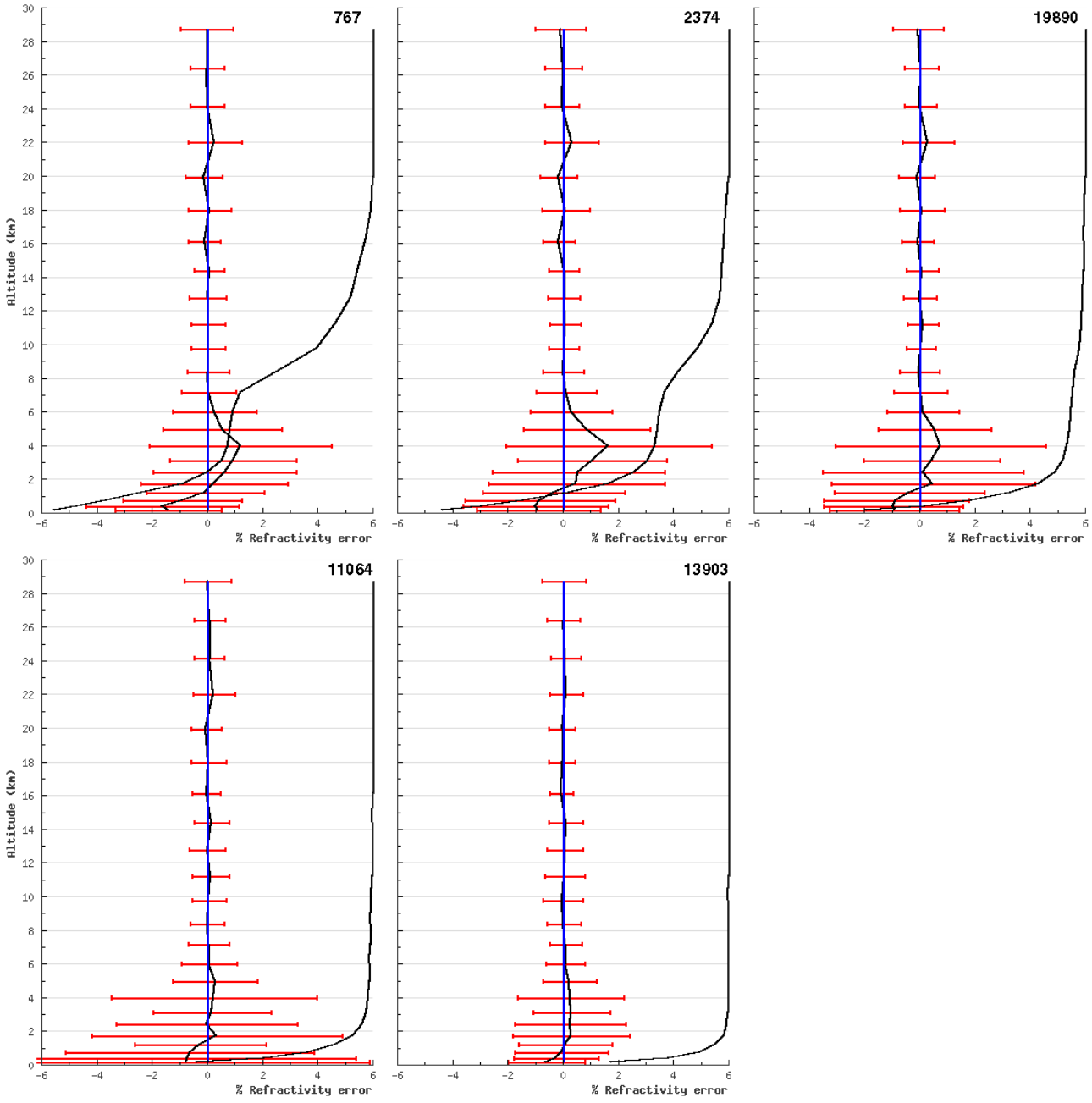

Figure 6: Mean and standard deviation of the errors in the refractivity profiles (as compared to ECMWF background) of Oceanic occultations, grouped according to the SVM flag value. Number of events per layer/case are also displayed (right end for 100% of total number on top; left end for 0%). Left to right NR, PNR, U (top) PR, and R (bottom) respectively.

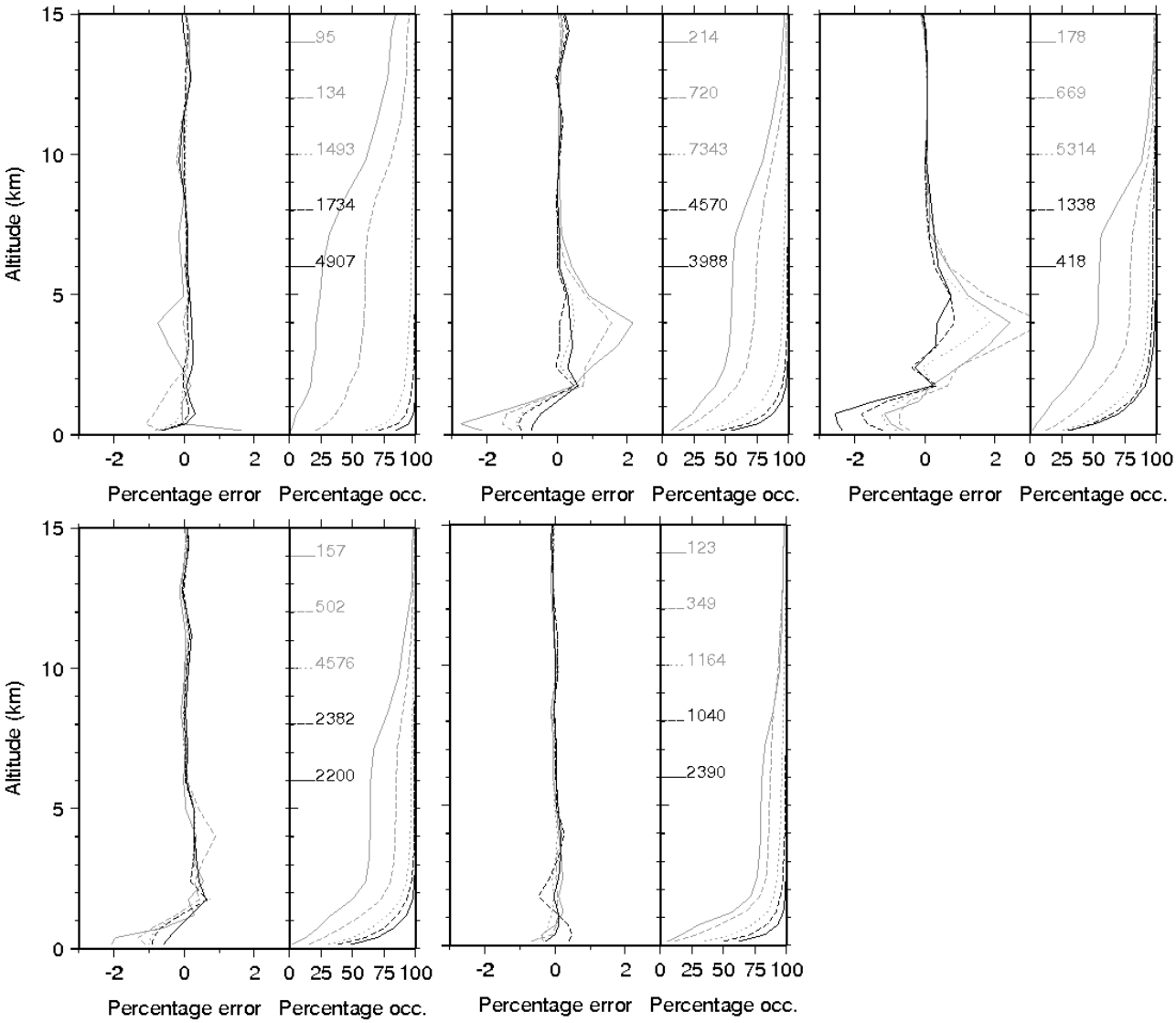

Figure 7: Biases in the refractivity profiles (as compared to ECMWF background) of oceanic occultations from the different geographic zones listed in Table 1: SP, MS, E (top) MN, NP (bottom) from left to right respectively. Each plot contains five statistics, grouped according to the SVM flag value as listed in Table 1: NR (solid-gray), PNR (dashed-gray), U (dotted), PR (dashed-black), and R (solid-black) respectively. On the left of each plot, the percentage of occultations considered in each class/layer, being 100% the number in the legend.

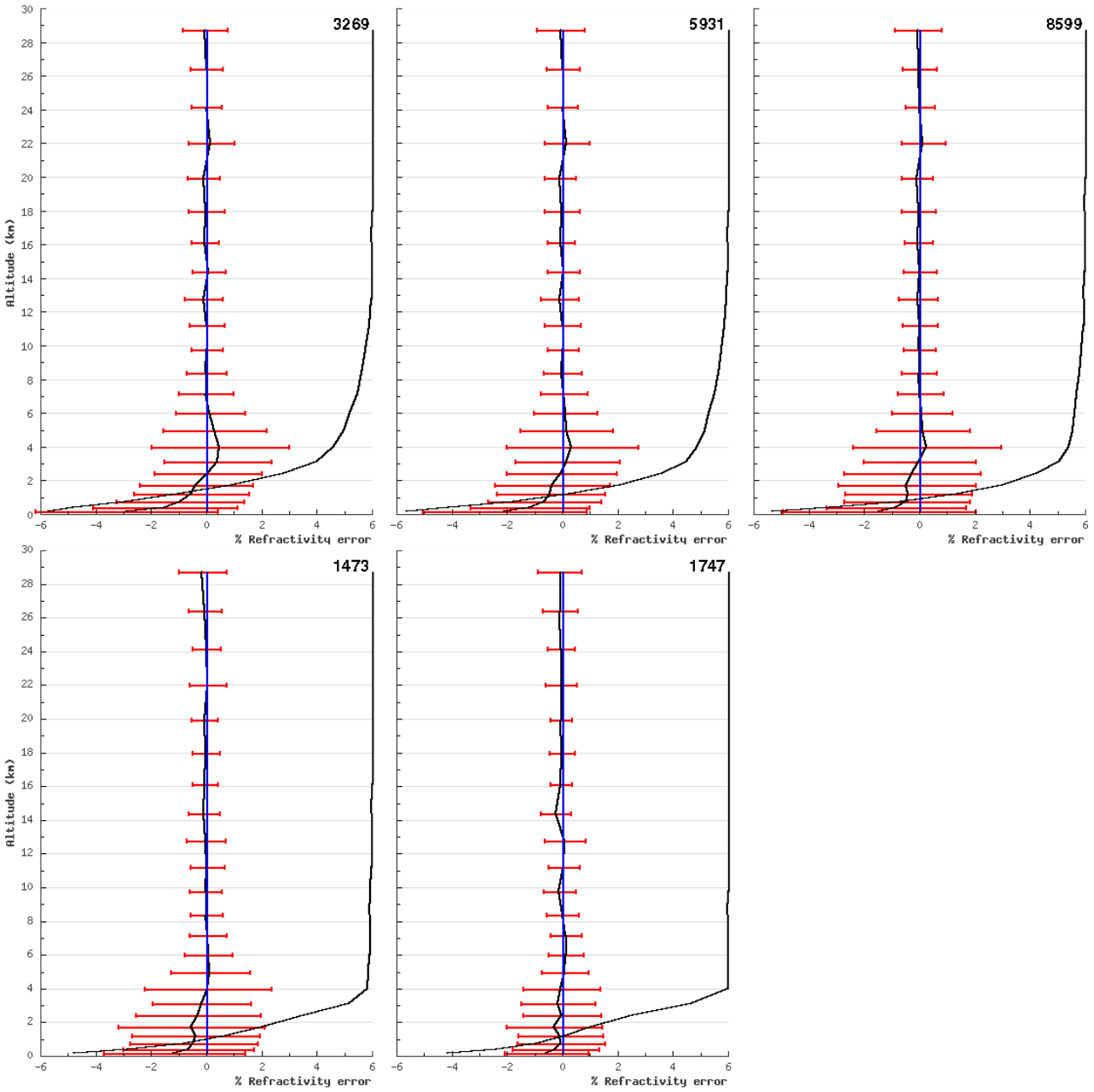

Figure 8: Same as Figure 6 but for occultations over land.

5 Conclusions

In order to quantify the effects of the reflected signals into the GNSS RO retrieved atmospheric products, simulation work has been done to detect variations in the inferred profiles provoked by the presence of reflected signals. This work shows that the Canonical Transform-like methods are insensitive to leakages of reflected signals.

A Support Vector Machine has also been implemented to detect presence of reflected signals in Level-1 GNSS RO data (phases/amplitudes). It has been validated, achieving 94% of overall success, 99.8%-success in events with classification values beyond the SVM-*margin*, and 98.5% success beyond the reduced margin (± 0.5). The SVM has been applied to 102118 setting COSMIC occultations (including open-loop data) to find geographical and seasonal patterns. Most reflections occur over Oceans, following a pattern that reminds the sea surface temperature distribution. Reflections tend to increase during local winter time, especially over Mid Latitude Oceans. Level-2 profiles (GPS-derived and background) were available for 69359 occultations out of the study set, and they have been used to generate statistics about the correlation between the reflection flag value and the quality of the retrieved profiles (quality as compared to ECMWF post-analyzed background). Over Oceans and Sea-Ice, occultations with reflected signals compare better with the background (bias and RMS) than occultations without reflected signals. A weaker but similar trend is observed in land occultations. This seems to indicate that rather than disturbing the atmospheric retrievals, the presence of reflections attests that the instrumental and atmospheric conditions are fine enough to get a clean Level-1 data set, and allow the RO techniques to generate quality profiles. We plan to include a reflection-flag in the operational GRAS products. Future work should investigate how to use this information to assist the assimilation of RO products into the NWPM.

Acknowledgments

This work is being conducted under the GRAS SAF CDOP project. E.C. has a Ramón y Cajal Award Contract of the Spanish Ministry of Science and Education. The simulations in Section 2 were performed by Michael Gorbunov (Institute for Atmosphere Physics, Russian Academy of Sciences), with participation of Kent Lauritsen (Danish Meteorological Institute, Denmark). The GRAS SAF team has nourished this work with comments and ideas.

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.

[Cardellach et al. (2004)] Cardellach, E., C.O. Ao, M. de la Torre-Juárez, and G.A. Hajj (2004), Carrier phase delay altimetry with GPS-reflection/occultation interferometry from low Earth orbiters, *Geophs. Res. Letters*, 31, L10402, doi:10.1029/2004GL019775.

[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

[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.

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