Efficient Collocation of GNSS Radio Occultation Soundings with Passive Nadir Microwave Soundings

Alex Meredith¹, Stephen Leroy², and Kerri Cahoy¹

¹Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
²Atmospheric and Environmental Research, Lexington, MA 02421, USA

Correspondence: Alex Meredith (ameredit@mit.edu)

Abstract. Radio Occultation (RO) using the Global Navigation Satellite Systems (GNSS) can be used to infer atmospheric profiles of microwave refractivity in the Earth’s atmosphere. GNSS RO data are now assimilated into numerical weather prediction models and used for climate monitoring. New remote sensing applications are being considered that fuse GNSS RO soundings and passive nadir-scanned radiance soundings. Collocating RO soundings and nadir-scanned radiance soundings, however, is computationally expensive, especially as new commercial GNSS RO constellations greatly increase the number of global daily RO soundings. This paper develops a new and efficient technique, called the “rotation-collocation-method”, for collocating RO and nadir-scanned radiance soundings in which all soundings are rotated into the time-dependent reference frame in which the nadir sounder’s scan pattern is stationary. Collocations with RO soundings are then found when the track of an RO sounding crosses the line corresponding to the nadir sounder’s scan pattern. When applied to finding collocations between RO soundings from COSMIC-2, Metop-B-GRAS, and Metop-C-GRAS and the passive microwave soundings of ATMS on NOAA-20, Suomi-NPP, and AMSU-A on Metop-B and Metop-C for the month of January, 2021, the rotation-collocation method proves to be 99.0% accurate and is hundreds to thousands of times faster than traditional approaches to finding collocations.

1 Introduction

Measurements made using radio occultation (RO) by the Earth’s atmosphere of the transmitters of the Global Navigation Satellite Systems (GNSS) are now routine and important contributors to numerical weather prediction and atmospheric reanalysis (Cardinali and Healy, 2014; Banos et al., 2019, and references therein). GNSS RO data fills in large holes in global coverage left by the international network of radiosondes, anchors atmospheric analyses by virtue of its near-absolute accuracy (Gelaro et al., 2017; Hersbach et al., 2020), and provides cloud-free information on atmospheric water vapor in the middle to lower troposphere (Kursinski and Gebhardt, 2014; Mascio et al., 2021). GNSS RO measurements are typically inverted to yield profiles of the index of refraction, a quantity with contributions from atmospheric density, temperature and water vapor (Kursinski et al., 2000).

Collocations of GNSS RO atmospheric soundings with the soundings of cross-track scanners in low Earth orbit are useful for several reasons. First, the contributions of water vapor and nitrogen/oxygen to the index of refraction cannot be separated based
on RO measurement alone; rather, separating their contributions requires the assistance of outside constraints. The commonly
used outside constraint is the forecast of a numerical weather prediction system (e.g. Healy and Eyre, 2000), but specialized
algorithms have been proposed that implement constraints based on collocated remote soundings by different techniques. One
such algorithm considers column water vapor as inferred from microwave radiometers (Xie, 2006; Wang et al., 2017). Second,
RO data have been used as a benchmark for investigating the accuracy of other remote sensing instruments by virtue of the
near absolute accuracy of the measurements. RO soundings have been compared to collocated microwave soundings in order
to validate the microwave soundings (Schrøder et al., 2003; Ho et al., 2007; Iacovazzi et al., 2020) because microwave radiance
standards are far less accurate than the timing standards which underly the GNSS signals used. GNSS RO validation reduces
concerns about the use of satellite microwave data for climate trend studies. Inter-comparison of RO and spectral thermal
infrared sounders for the sake of validating the calibration of the infrared sounders has also been investigated (Feltz et al.,
2017). These studies find that collocation between GNSS RO soundings and soundings of passive nadir cross-track scanners is
necessary.

The computation of collocation between RO and passive nadir scanners is computationally expensive. Collocation is defined
by tolerance windows in the spatial and temporal separation between a pair of RO and passive nadir soundings. The most direct
approach to collocation is to consider a large batch of RO soundings and a large batch of passive nadir scanning data in time
periods \( \Delta T \) significantly longer than the temporal separation window \( \Delta t \) and calculating the temporal and spatial separations
between every potential pair of RO and nadir-scanner soundings to find pairs that meet the collocation criteria defined by
the tolerance windows. Because of the large numbers of passive nadir soundings involved, the computation of collocations is
extremely expensive. The expense can be reduced by decreasing the time window on passive nadir scanner soundings to be
considered to \( \Delta T = 2\Delta t \), but no further optimization is possible. We refer to collocation approaches similar to this as a brute
force method.

Another method of collocation is motivated by the density and pattern of passive nadir soundings: they are so dense that
they leave no gaps in their coverage during nominal operations, and their coverage pattern can be predicted precisely using an
orbit propagator. If the reference frame for a collocation is one in which the scan pattern is just a stationary line of soundings,
then a collocation is found when the location of an RO sounding in the scan-pattern reference frame crosses the nadir scanner’s
scan line. The advantage of such an algorithm is that the actual geolocations and times of the passive nadir soundings need not
be considered at all; only the geolocations and times of the RO soundings do. Consequently, the determination of collocations
should be greatly accelerated over a brute force method. The algorithm for collocation involving rotation into the reference
frame of the nadir scan pattern we refer to as the rotation-collocation method.

The rotation-collocation method promises a great increase in efficiency over any brute force collocation method, but two
complications must be addressed. Each is associated with a key assumption of the rotation-collocation method, and the errors
that result must be quantified. The first assumption is that the scan of the passive nadir scanner is defined precisely as a line
in its own reference frame. In actuality, rather than a simple line, the scan pattern is a co-linear set of footprints of finite,
non-zero sizes and distorted elliptical shapes, with greater distortion at the ends of the scan. The second assumption is that
a simple orbit propagator and a range of scan angles of the passive nadir scanner is sufficient to determine the coverage of
the scan footprints. These assumptions can be validated and the associated errors quantified by direct comparison of a set of collocations determined by the rotation-collocation method to a set of collocations determined by a brute force method, the latter serving as a truth standard. Rates of false positives and false negatives can be estimated. Once these complications are addressed, then all that remains is to compute how great an acceleration in computation is gained by the rotation-collocation method over a standard brute force method.

This paper is organized as follows: Section 2 contains a description of the brute force method and a theoretical exposition of the rotation-collocation method. Both will be applied to candidate data sets in order to validate the rotation-collocation method and to determine the acceleration gained by the rotation-collocation method. Section 3 describes the data sets that will be used in the study and define the experimental setup. Section 4 contains an analysis of the experiments, including a quantification of the daily numbers of collocations of RO soundings with passive nadir microwave soundings. Section 5 is a discussion of the results with a suggestion for future satellite deployments. Section 6 presents the final conclusions.

2 Approach and theory of the collocation algorithms

This section describes the details of the brute force and rotation-collocation algorithms. Collocations are defined as RO soundings that are separated from a passive nadir sounding by at most $\Delta t$ in time and $\Delta d$ in distance. First, the details of two approaches to the brute-force algorithm are presented, and then two approaches to the rotation-collocation algorithm are presented.

2.1 The brute-force algorithm

The brute-force algorithm uses two checks: a spatial check and a time check. Because the brute-force algorithm makes no approximation, the brute-force method is a truth metric against which the accuracy of our rotation-collocation methods can be evaluated. This subsection describes two implementations of the brute-force algorithm: the first implementation considers all soundings of the nadir scanner over the course of a day when searching for collocations with RO soundings, and the second implementation improves efficiency by sorting the nadir scanner soundings in time and windowing the soundings to within $\Delta t$ in time of the RO sounding before searching for spatial collocations.

2.1.1 Brute-force #1: All nadir scan soundings

The first brute-force approach compares every RO sounding to every nadir scanner sounding over a one-day period, performing a spatial check and a time check for every RO-nadir scanner sounding pair. A generic RO sounding has latitude $\theta_{RO}$, longitude $\lambda_{RO}$, and sounding time $t_{RO}$; and a generic nadir scanner sounding has latitude $\theta_{NS}$, longitude $\lambda_{NS}$, and sounding time $t_{NS}$. The spatial and temporal checks for collocation are:

$$\sqrt{\left(\lambda_{NS} - \lambda_{RO}\right)^2 \cos^2 \theta_{RO} + \left(\theta_{NS} - \theta_{RO}\right)^2} < \frac{\Delta d}{R_E}$$

(1a)

$$|t_{NS} - t_{RO}| < \Delta t$$

(1b)
in which $R_E$ is Earth’s radius at the equator and longitudes and latitudes have units of radians. Note that equation 1a assumes small separations, $\delta d \ll R_E$. The temporal check is performed first, which permits a minor speed optimization using early termination: the logging practices of typical nadir scanner instruments generally associate a single time for a fixed number of footprints, thereby permitting a brute force method to greatly reduce the number of time checks.

### 2.1.2 Brute-force #2: Search-sort

This approach is similar to that of the brute-force method discussed in §2.1 but with narrowed windowing in time. The spatial check remains the same as the one given in Equation 1a; however, this approach avoids a time check by time-sorting the nadir scanner data. For each RO sounding, we search for the nadir scanner data indices corresponding to the window $[t_{RO} - \Delta t, t_{RO} + \Delta t]$. Then, we poll the nadir scanner data falling in this time window—which is guaranteed to pass the time check—and perform only the spatial check when searching for collocations.

With $n$ as the total number of nadir scanner soundings, and $r$ as the total number of RO soundings, brute-force method #1 has a time complexity of $O(rn)$ for the time check. Sorting the nadir scanner data has a time complexity of $O(n \log n)$, where $n$ is the total number of nadir scanner soundings, and searching the nadir scanner data has a time complexity of $O(\log n)$, so brute-force method #2 has a time complexity of $O(r \log n) + O(n \log n)$ for the time check, as this method performs one initial sort of the nadir scanner data and then one search of the nadir scanner data for each RO sounding.

In most cases, this method is faster than brute-force method #1, but when the number of nadir scanner soundings is very large (e.g. $\log n > r$, with $n$ the number of nadir scanner soundings and $r$ the number of RO soundings), the time required to sort the nadir scanner soundings can become long enough that brute-force method #2 takes longer than brute-force method #1. Furthermore, both brute-force methods avoid performing spatial checks for nadir scanner soundings outside of the time window, so the number of spatial checks performed by both methods is the same. The spatial check is much slower than the time check, so as the time window grows and the number of spatial checks required grows, the time taken by brute-force method #2 approaches the time taken by brute-force method #1.

### 2.2 The rotation-collocation algorithm

The rotation-collocation method has two steps: first, rotating the RO soundings into the nadir scanning satellite’s time-varying frame to find the apparent path of the RO sounding in the reference frame, and second, determining whether the apparent path of the RO sounding intersects the nadir scanner’s pattern. The two patterns can only intersect if both the spatial and the temporal checks are satisfied.

Figure 1 illustrates the transformations undertaken by the rotation-collocation method. Panel (a) shows an RO sounding from COSMIC-2 E1 occurring at 00:23:52 UTC on January 2, 2021, and the pattern of NOAA-20 Advanced Technology Microwave Sounder (ATMS) soundings occurring within $\Delta t = 600$ seconds of the RO sounding, from 00:13:52 to 00:33:52 UTC. Panel (b) shows the COSMIC-2 RO sounding and the ATMS soundings rotated into the time-varying frame of NOAA-20. All of the ATMS soundings collapse to a near-perfect single line that extends upward and downward by an amount related to the range of the nadir scanner’s scan angles. Also notice that the single RO sounding is represented as a series of points corresponding
Figure 1. (a) A radio occultation from COSMIC-2 E1 plotted against contemporaneous MW soundings from NOAA-20 ATMS. (b) The discretized apparent position of the same occultation rotated into the MW frame, plotted against the MW sounding pattern. (c) The discretized apparent position of the RO sounding plotted against the linearized MW sounding pattern. (d) The interpolated apparent position of the RO sounding plotted against the linearized MW sounding pattern.

to its apparent location in the rotated frame at varying times $t_i$ with the time window $[t_{RO} - \Delta t, t_{RO} + \Delta t]$. In this work, we refer to these apparent locations of a single RO sounding in the time-varying rotated frame as “sub-occultations”, as shown in panels (b) and (c) of Figure 1. The apparent path of the sub-occultations crosses the line of the ATMS scan pattern, indicating the existence of a collocation.

Panel (c) of Figure 1 demonstrates the first approximation associated with the rotation-collocation algorithm, that a nadir scanner sounding pattern can be approximated by a perfect line at $\delta u = 0$ with $u$ the argument of latitude, or along-track coordinate of the nadir scanner, found using an orbit propagator. This approximation rests on three major assumptions: first, that the footprints of the nadir scanner, which are distorted ellipses, can be treated as single points; second, that the orbit propagator used in the rotation is perfectly accurate; and third, that the nadir scanner sounding pattern leaves no gap in coverage. This approximation has the advantage of not having to consider any of the geolocations of the nadir scanning instrument at all.

Panel (d) of Figure 1 illustrates the second approximation of the rotation-collocation algorithm, that the sub-occultations fall on a straight line in the rotated frame. Without this second approximation, the location of each sub-occultation must be computed and the line connecting consecutive sub-occultations must be checked for crossing the scan line of the nadir scanner.
With the second approximation, however, only the sub-occultations at times \( t_{RO} - \Delta t \) and \( t_{RO} + \Delta t \) are computed and the line connecting the two checked for crossing the nadir scanner scan line. The only imperfection of this approximation is that there is some minute amount of curvature associated with the path of the RO sub-occultations in the rotated frame, and that curvature becomes increasingly pronounced with longer time collocation windows \( \Delta t \).

The explicit rotation of the rotation-collocation algorithm is given by

\[
\begin{pmatrix}
x_R(t) \\
y_R(t) \\
z_R(t)
\end{pmatrix} = 
\begin{pmatrix}
cos u(t) & sin u(t) & 0 \\
-sin u(t) & cos u(t) & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 & 0 \\
0 & cos i & sin i \\
0 & -sin i & cos i
\end{pmatrix}
\begin{pmatrix}
\cos \Omega(t) & \sin \Omega(t) & 0 \\
-sin \Omega(t) & \cos \Omega(t) & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x_{ECI}(t) \\
y_{ECI}(t) \\
z_{ECI}(t)
\end{pmatrix}
\]  
(2a)

in which \( u \), \( i \), and \( \Omega \) are the argument of latitude, the inclination, and the right ascension of the ascending node of the nadir scanner satellite, and the coordinates \( x_{ECI}(t), y_{ECI}(t), z_{ECI}(t) \) are Cartesian coordinates of a location in an Earth-centered inertial coordinate system. In the collocation problem, the input coordinates are longitude \( \lambda \) and latitude \( \theta \), and so first we transform the latitude and longitude of a sounding to a position in an Earth-centered Earth-fixed (ECF) coordinate system given by \((\cos \lambda \cos \theta, \sin \lambda \cos \theta, \sin \theta)\) and then compute the ECI coordinates according to the time-dependent transformation \( t \):

\[
(x_{ECI}(t), y_{ECI}(t), z_{ECI}(t)) = L_i(x_{ECF}, y_{ECF}, z_{ECF}) = L_i(\cos \lambda \cos \phi, \sin \lambda \cos \phi, \sin \phi)
\]  
(2b)

The results of equations 2a and 2b are the rotated Cartesian coordinates \( (x_R, y_R, z_R) \). These coordinates are best interpreted as an along-track coordinate that we call “delta argument of latitude” \( (\delta u) \) and the cross-track coordinate that we call the “scan distance” \( (\delta s) \):

\[
\delta u = \arctan(y_R, x_R)
\]  
(3a)

\[
\delta s = \arcsin z_R
\]  
(3b)

in which \( \arctan(\cdots, \cdots) \) is a four-quadrant arctangent defined such that \( \tan \delta u = y_R/x_R \). Both \( \delta u \) and \( \delta s \) are distances on the Earth’s surface in units of radians. They can be converted to degrees by multiplying by \( 180^\circ/\pi \) as in figure 1 or to distance by multiplying by the radius of the Earth \( (R_E) \).

The scan pattern of the nadir-scanning satellite in the rotated frame of reference can be described as the line segment \(-\delta s_{max} < \delta s < \delta s_{max}, \delta u = 0\), where \( \delta s_{max} \) is essentially limited by \( \xi_{max} \), the maximum scan angle of the nadir-scanning instrument. The relationship between the maximum of the scan distance \( (\delta s_{max}) \) in the rotated frame and the maximum scan angle \( (\xi_{max}) \) of the scanning instrument is found using the law of sines:

\[
\delta s_{max} = \arcsin\left(\frac{a}{R_E(\theta)} \sin \xi_{max}\right) - \xi_{max}
\]  
(4)

in which \( a \) is the radius of the nadir-scanner satellite’s orbit. The computation of the scan distance allows for a minor correction associated with the oblateness of the Earth, namely that the Earth’s radius is a function of latitude \( (R_E = R_E(\theta)) \).
2.2.1 Rotation-collocation #1: Sub-occultations

In order to determine collocation, then, it is only necessary to check whether the path of the RO sounding in the rotated frame crosses the line associated with the scan pattern of the nadir-scanning instrument. Recall that the RO sounding is a trajectory in this frame because the coordinate system rotates with the scan line of the nadir scanner, which itself is moving during the time window \[[t_{RO} - \Delta t, t_{RO} + \Delta t]\]. We define the apparent trajectory of sub-occultations for a generic RO sounding with longitude \(\lambda_{RO}\), latitude \(\theta_{RO}\), and time \(t_{RO}\) at times \(t_i\) by

\[t_i = t_{RO} + dt \left( i - (N + 1)/2 \right)\]  

(5)

in which \(dt = 2\Delta t/(N-1)\) is the time separation between consecutive sub-occultations and \(N\) is the number of sub-occultations. The position of each sub-occultation is computed in the rotated frame (recall that the transformations of equations 2a and 2b are both time-dependent). Each segment connecting consecutive sub-occultations in the rotated frame is checked for crossing the scan line \(\delta u = 0\) of the nadir-scanning instrument. If any segment crosses the scan line, the temporal check for collocation is satisfied. If the intersection occurs at a scan distance \(|\delta s| < \delta s_{\text{max}}\), then the spatial check for collocation is satisfied. When both the spatial and temporal checks are satisfied, a collocation is found.

The computational expense of this approach to rotation-collocation algorithm comes from running an orbit propagator as implied for determination of \(u(t)\) in equation 2a, which is executed \(N\) times for each RO sounding. If there are \(r\) total RO soundings and \(N\) sub-occultations per RO sounding, the time complexity of orbit propagation is \(O(rN)\), and does not depend on the number of nadir scanner soundings. As a result, the rotation-collocation method is significantly faster than either brute-force method when there are large numbers of nadir scanner soundings.

2.2.2 Rotation-collocation #2: Linearized

In the linearized approach to the rotation-collocation algorithm, the positions of only two of the RO sub-occultations are computed, and those are at \(t = t_{RO} - \Delta t\) and at \(t = t_{RO} + \Delta t\), and the line segment connecting those two positions in the rotated frame is checked for crossing the scan line. If it does cross the scan line (\(\delta u = 0\)), the temporal check is satisfied, and if it crosses the scan line at \(|\delta s| < \delta s_{\text{max}}\), then the spatial check is satisfied and a collocation is found.

The computational expense of this approach to the rotation-collocation algorithm comes from running an orbit propagator as implied for determination of \(u(t)\) in equation 2a, which is executed only 2 times for each RO sounding. As such, if there are \(r\) total RO soundings, the time complexity of orbit propagation is \(O(r)\). Recalling that the time complexity of orbit propagation is \(O(rN)\) for the rotation-collocation algorithm with sub-occultations, when \(N\) is much greater than 2, the linearized approach to collocation is much faster than the sub-occultation approach; however, it can be less accurate because the path of the RO sounding in the rotated frame is not strictly a straight line. The greater the temporal window \(\Delta t\) is, the more curved the trajectory becomes. As explored in §4.6, as \(\Delta t\) grows and the trajectory curvature increases, the number of incorrect predictions made by the linearized rotation-collocation method also increases, and using the rotation-collocation method with sub-occultations becomes necessary to preserve accuracy.
3 Experimental setup

We devise a set of experiments to test the validity of the approximations of the rotation-collocation algorithm posed in the introduction and evaluate the computational efficiency gains for each. The experiments consist of a month of geolocations of actual RO data and nadir-scanning data from January 2021. Because of the promise in using nadir microwave radiance to construct weather-independent temperature and water vapor profiles from the surface to the stratopause, we use the geolocations of highly precise, well-calibrated microwave nadir radiance data. The nadir-scanner geolocations are for the Advanced Microwave Sounding Units (AMSU-A) instruments on the Metop satellites (Metop-B and Metop-C) and for the Advanced Technology Microwave Sounders (ATMS) on the Suomi-NPP and NOAA-20 satellites. All are in sun-synchronous orbits with the Metop satellites having their ascending node at 21:31 local solar time and the Suomi-NPP and NOAA-20 satellites at 13:25 local solar time. In January 2021, all four of these microwave radiance instruments collected 238,198,740 soundings, as detailed in Table 1.

Table 1. Total number of microwave radiance soundings over the month of January 2021 for each nadir-scanning microwave satellite.

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Number of MW soundings</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOAA-20</td>
<td>121,070,688</td>
</tr>
<tr>
<td>Metop-B-AMSU</td>
<td>10,265,310</td>
</tr>
<tr>
<td>Metop-C-AMSU</td>
<td>10,111,350</td>
</tr>
<tr>
<td>SNPP</td>
<td>96,751,392</td>
</tr>
<tr>
<td>All</td>
<td>238,198,740</td>
</tr>
</tbody>
</table>

For the RO sounders, we choose two contemporary RO constellations: the two-satellite constellation of Metop consisting of Metop-B and Metop-C, and the six-satellite constellation of COSMIC-2. (Note that the Metop satellites carry both nadir microwave scanners and RO instruments.) These RO satellites are characterized by high signal-to-noise ratios for signal tracking but differ substantially in their orbits. The Metop satellites fly in sun-synchronous orbits, as above, while the COSMIC-2 satellites fly in 24° inclination, 520 km altitude, rapidly precessing orbits. In January 2021, the eight RO satellites obtained 160,298 RO soundings, as detailed in Table 2. By choosing these very different RO orbits, we not only can test the rotation-collocation algorithm, we also gain some insight into the frequency of microwave-RO collocations according to orbit types. Co-hosted instruments such as on the Metop satellites can intuitively be expected to yield greater numbers of collocations daily than RO and microwave radiance instruments on different satellites with unrelated orbits – about 40% of Metop RO soundings are collocated with Metop microwave radiance soundings, whereas generally under 5% of RO soundings are collocated with microwave radiance soundings from any instrument hosted on a different satellite in an unrelated orbit.
Table 2. Total number of RO soundings over the month of January 2021 for each RO satellite. Note that the COSMIC-2 constellation contains six satellites. No Metop-C-GRAS soundings are available for January 17, 2021, so there are fewer Metop-C-GRAS soundings than Metop-B-GRAS soundings.

<table>
<thead>
<tr>
<th></th>
<th>Number of RO soundings</th>
</tr>
</thead>
<tbody>
<tr>
<td>COSMIC-2</td>
<td>125,665</td>
</tr>
<tr>
<td>Metop-B-GRAS</td>
<td>18,140</td>
</tr>
<tr>
<td>Metop-C-GRAS</td>
<td>16,493</td>
</tr>
<tr>
<td>All</td>
<td>160,298</td>
</tr>
</tbody>
</table>

All computations are done in Python version 3.9. The orbit propagator used in computing $u(t)$, $i(t)$, and $\Omega(t)$ of equation 2a is sgp4 (Vallado et al., 2006). It is initiated approximately three times daily from two-line orbit elements (TLEs). The transformation between ECF and ECI coordinate frames of equation 2b is executed using astropy, with the ECF chosen to be the International Terrestrial Reference System (ITRS) and the ECI frame the Geocentric Celestial Reference System (GCRS) (Price-Whelan et al., 2022). The conversion is executed only three times for every two-line element description in the Celestrak database in order to establish the Earth’s pole and rotation rate. Subsequent transformations between ECF and ECI are executed using the calculated pole and rotation rate.

We obtained the Metop data from Eumetsat (https://eoportal.eumetsat.int/), and the NOAA-20 and Suomi-NPP data from NOAA’s CLASS data system (https://www.class.noaa.gov/). We retrieved the RO sounding data from the COSMIC Data Analysis and Archive Center (https://data.cosmic.ucar.edu/gnss-ro/). We also retrieved historical TLEs for Suomi-NPP, Metop-B, Metop-C, NOAA-20, and the COSMIC-2 constellation from Celestrak (https://celestrak.org/) for use in the rotation-collocation method. We grouped data into folders by instrument and day, and then ran all four methods on each combination of instruments per day.

4 Analysis

We analyze the performance of the two approaches of the rotation-algorithm using signal detection theory—counting false positive and false negative rates—using the brute force algorithm as the definition of truth. Because the two approaches to the brute force algorithm are provably the same despite their different approaches to checking for temporal match-ups, they both yield precisely the same collocation pairs. In this section we present a set of case studies. In each case we choose a spatial tolerance of $\Delta s = (150 \text{ km})/R_E$, and in all cases but the last we choose a time window of $\Delta t = 600 \text{ sec}$, or ten minutes; in the fourth and final case we choose a time window of $\Delta t = 10800 \text{ sec}$, or three hours. Occultation yield can be expected to increase in direct proportion to $\Delta t$ for time windows significantly shorter than the orbital period of the nadir scanning satellites. The first case study considers collocations between COSMIC-2 RO soundings and NOAA-20 microwave radiance soundings. This is a typical case since many future RO instruments will not necessarily be co-hosted with microwave radiance sounders and will be
in different orbits. The second case study is for the co-hosted RO and microwave radiance soundings on the Metop satellites. While not many such pairings will be deployed in the future, it may suggest that RO and microwave radiance sounders be flown in tandem orbits if maximizing the collocation yield is desired. Third, the total yield of RO-microwave radiance collocations for the month of January 2021 is considered. The final case study reconsiders collocations between COSMIC-2 RO soundings and NOAA-20 microwave radiance soundings but with a time window of $\Delta t = 10800$ sec. This final case study demonstrates the excellent accuracy and efficiency of the rotation-collocation method with sub-occultations over long time windows, and documents the slight decrease in accuracy of the linearized rotation-collocation method as the curvature of the trajectory of sub-occultations in the nadir-scanner frame increases over a longer time window.

4.1 Case study: COSMIC-2 (RO) and NOAA-20 (microwave)

![Figure 2](https://doi.org/10.5194/egusphere-2022-1266)

In this case study, we examine collocations between the six-satellite COSMIC-2 radio occultation constellation and ATMS on NOAA-20, a microwave radiance sounder. In Figure 2(a), we show the collocations between COSMIC-2 and NOAA-20.
20 by day found by each of our four collocation-finding methods. The rotation-collocation algorithm with sub-occultations (orange) and the linearized rotation-collocation algorithm (light green) find slightly more collocations on each day than the brute-force algorithm (blue), but the true positive rate, defined as the number of collocations correctly predicted by the rotation-collocation method divided by the total number of predicted collocations correctly or incorrectly, is 98.9%. The time window for collocation for Figure 2 is $\Delta t = 600 \text{ sec}$. The “fraction collocated” axis on the right of Figure 2(a) is the number of predicted collocations divided by the average number of daily occultations. Notably, only 2% to 5% of COSMIC-2 RO soundings are collocated with NOAA-20 ATMS microwave radiance soundings over the month of January 2021 when $\Delta t = 600$ is used as the time tolerance for collocation, because NOAA-20 and COSMIC-2 satellites are rarely near each other.

In Figure 2(b), we show a confusion matrix for this case study. The number of sub-occultations used for this analysis is $N = 21$, and the temporal spacing between sub-occultations is $dt = 60 \text{ sec}$. In the confusion matrix, the top and bottom rows correspond to the numbers of collocations of RO soundings not found and found by brute force, respectively; and the left and right columns to the numbers of collocations of RO soundings not predicted and predicted by one of the rotation-collocation methods, respectively. The true positive rate for the rotation-collocation method with sub-occultations is $3839/3883 = 98.9\%$, and the true negative rate is $121773/121782 = 99.993\%$.

In Figure 2(c), we show the spatial distribution of COSMIC-2 soundings collocated with NOAA-20 microwave radiance soundings for January 15, 2021, found by the linearized rotation-collocation algorithm and by the brute-force method. Collocations found by the linearized rotation-collocation algorithm are shown as orange circles, and those found by the brute force algorithm are shown as blue dots. The vast majority of these collocated soundings are found by both methods. Both the brute force and the rotation-collocation algorithms found 134 collocations, 133 common to both algorithms. The rotation-collocation algorithm predicted one collocation not found by the brute force algorithm, and the brute force algorithm found one collocation not found by the rotation-collocation algorithm. For this day, the true positive rate is $133/134 = 99.3\%$.

In Figure 2(d), we show a confusion matrix for collocations found by the linearized rotation-collocation algorithm between COSMIC-2 RO soundings and NOAA-20 ATMS soundings for the month of January 2021. The linearized rotation-collocation algorithm finds the same collocations as the rotation-collocation method with sub-occultations in this case, and so the true positive rate for the linearized rotation-collocation method is $3839/3883 = 98.9\%$ and the true negative rate is $121773/121782 = 99.993\%$.

Many of the misclassified COSMIC-2 RO soundings (44 out of 53 total) are false positives, predicting a collocation when one does not exist. We found that 6 out of 44 (13.6%) of the false positives result from unavailable NOAA-20 data. For these soundings, no NOAA-20 ATMS sounding exists within eight seconds of the predicted collocation time. Of the remaining false positives, 11 (25% of total) are soundings that fall just beyond the maximum $\delta u_{\text{max}}$ when compared to NOAA-20’s orbit, thereby falling just outside the time window $\Delta t$. The remaining 27 false positives (61.4% of total) are soundings that fall just outside of the maximum scan range $\delta s$ of the NOAA-20 ATMS instrument. All of the false positive and false negative cases found here are associated with failures of the first assumption of the rotation-collocation algorithm, namely, that all of the nadir scanner soundings fall perfectly on an unbroken line at $\delta u = 0$ in the rotated frame as illustrated by Figure 1(c).
In summary, both versions of the rotation-collocation algorithm are correct on 98.9% of the occasions for which a collocation between COSMIC-2 RO soundings and NOAA-20 ATMS soundings is predicted and incorrect only 0.007% of the time when a COSMIC-2 RO sounding is not found to be collocated with a NOAA-20 ATMS sounding. Over the course of January 2021, the true number of collocated soundings between COSMIC-2 RO and NOAA-20 ATMS soundings within a time window of 10 minutes is 3,848. The yield as a fraction of total COSMIC-2 RO soundings is 3.0% over the month. On a daily basis, the fraction ranges from 2.0% to 5.0%; see Figure 2(a).

4.2 Case study: Metop-B (RO) and Metop-B (microwave)

![Figure 3](https://example.com/fig3)

**Figure 3.** (a) Daily collocations for Metop-B GRAS RO soundings and Metop-B AMSU-A microwave radiance soundings for the month of January 2021. (b) Confusion matrix for the Metop-B GRAS RO soundings and Metop-B AMSU-A ATMS microwave soundings for January 2021 for the rotation-collocation method with sub-occultations. (c) Map of collocations on January 15, 2021, for Metop-B AMSU-A microwave soundings and Metop-B GRAS RO soundings. (d) Confusion matrix for the Metop-B GRAS RO soundings and Metop-B AMSU-A microwave soundings for January 2021 for the linearized rotation-collocation method.

In this case study, we examine collocations between two instruments cohosted on a satellite, the GRAS RO instrument and the AMSU-A nadir-scanning microwave radiance instrument. Figure 3(a) is the same as figure 2 but for this case study.

In this case study, we examine collocations between two instruments cohosted on a satellite, the GRAS RO instrument and the AMSU-A nadir-scanning microwave radiance instrument. Figure 3(a) is the same as figure 2 but for this case study.

Co-hosting instruments greatly increases the collocation yield, with around 38–46% of Metop-B RO soundings collocated with Metop-B microwave soundings, in comparison to under 3% of COSMIC-2 RO soundings collocated with NOAA-20
microwave soundings. The intuition for this is straightforward. If a setting RO sounding is obtained at a time $t_{RO}$, then it is very likely that the satellite had flown over that same location earlier by $L/v_{leo}$ in which $L$ is the limb distance for the RO sounding and $v_{leo}$ is the low-Earth orbiting satellite’s orbital velocity. Typically, $L \approx 3000$ km and $v_{leo} \approx 7.5$ km s$^{-1}$, meaning a collocated microwave radiance sounding may have been swept out by the scanner approximately 400 seconds prior. The temporal collocation check is always satisfied for co-hosted RO and nadir scanning instruments as long as the spatial window is greater than 400 seconds ($\Delta t > 400$ s). For the collocation to be found, though, the boresight angle of the RO sounding with respect to the satellite’s velocity vector must be less than the angle corresponding to the sweep of the AMSU-A scan $\delta s_{\text{max}}$ as viewed at limb distance $L$. Maximum boresight angles for RO instruments typically lie around $60^\circ$, but the nadir scan of AMSU-A corresponds to a maximum boresight of approximately $27^\circ$ at limb distance. As a consequence, instead of all RO soundings by Metop-B being collocated with a Metop-B microwave sounding, approximately only 40% are. This corresponds to the spatial check for collocation only being met 40% of the time.

Figure 3(b) shows the performance of the rotation-collocation method with sub-occultations on collocations between Metop-B-GRAS and Metop-B-AMSU throughout the month of January 2021, using 21 sub-occultations, or a sixty-second spacing between sub-occultations. The true positive rate for the rotation-collocation method with sub-occultations is $7179/7224 = 99.4\%$ and the true negative rate for the rotation-collocation method with sub-occultations is $10894/10918 = 99.8\%$.

Figure 3(d) shows the performance of the linearized rotation-collocation method on collocations between Metop-B-GRAS and Metop-B-AMSU throughout the month of January, 2021. Most of the misclassified Metop-B-GRAS soundings are false positives. We have found that 2 of 40 (5\%) of the false positives are due to unavailable Metop-B-AMSU data – for these predicted collocations, there is no Metop-B-AMSU sounding data available within eight seconds of the predicted collocation time. The remaining 38 (95\% of total) false positives are soundings that fall just outside of the maximum scan range $\delta u_{\text{max}}$ of the Metop-B-AMSU instrument. The true positive rate for the linearized rotation-collocation method is $7179/7219 = 99.4\%$, but excluding incorrect predictions that occur due to missing data, the true positive rate is $7179/7213 = 99.5\%$. The true negative rate for the rotation-collocation method is $10894/10921 = 99.8\%$.

4.3 Full analysis: COSMIC-2 and Metop (RO); S-NPP, NOAA-20, Metop (MW)

In this section, we examine collocations between COSMIC-2 and Metop-B and Metop-C radio occultations and Metop, S-NPP, and NOAA-20 microwave soundings. In Figure 4, we show the collocations by day found by each of our four collocation-finding methods, as well as the fraction of radio occultations that are collocated with microwave soundings, using the daily average number of radio occultations as the denominator. Metop-C-GRAS data was missing for January 17, which explains the steep drop in total collocations found on January 17. As before, the time window for collocation is $\Delta t = 600s$.

Over all satellite combinations, only between 16\% and 23\% of RO soundings are collocated with any MW soundings; ideally, as many RO soundings would be collocated with MW soundings as possible. It is clear that there is room for improvement in the percentage of soundings that are collocated, and cohosting instruments leads to a large increase in collocations, as shown in §4.2.
Figure 4. Daily collocations for all satellite combinations for the month of January 2021. “Fraction collocated” refers to the ratio of collocations each day to the average daily number of RO soundings over all RO satellites.

Figure 5. Map of collocations on January 15, 2021, for all satellite combinations.

In Figure 5, we show all the collocations on January 15, 2021. These collocations occur all over the globe. Collocations in the Tropics, between 23.43°S and 23.43°N in latitude, are most important for profiling water vapor in the planetary boundary.
layer (Wang et al., 2017). Future satellite missions with GNSS-RO payloads should consider cohosting microwave radiometer payloads or launching into low-inclination orbits in order to meet the need for collocations in the Tropics.

Overall, the linearized rotation-collocation method found 29944 collocations and correctly identified 159852 RO soundings as not collocated. There were 170 missed predictions, or occultations for which the brute-force method found a collocation but the linearized rotation-collocation method did not. There were 276 incorrect predictions, which are occultations where the linearized rotation-collocation method found a collocation, but the brute-force method did not. Out of these 276 incorrect predictions, 25.4% were caused by missing microwave data, 31.5% were soundings that fall just outside of the maximum scan range of an microwave instrument, and the remaining 43.1% were soundings that fall just beyond the maximum delta argument of latitude when compared to an MW satellite’s orbit. The linearized rotation-collocation method had a \( \frac{29944}{30220} = 99.0\% \) true positivity rate and a \( \frac{159852}{160022} = 99.9\% \) true negative rate. Excluding incorrect predictions resulting from missing or corrupted microwave radiance data, the true positive rate is \( \frac{29944}{30150} = 99.5\% \).

### 4.4 Number of collocations by day, by satellite combinations

Table 3 shows the number of collocations per day over the month of January 2021 by satellite combinations. Metop-B-AMSU and Metop-B-GRAS generate a large yield of collocations. These instruments are co-hosted, which allows a high percentage of occultations to be collocated with microwave soundings. For the same reason, Metop-C-AMSU and Metop-C-GRAS share a high number of collocations. There are no collocations between Metop-B-AMSU and Metop-C-GRAS, and non between Metop-C-AMSU and Metop-B-GRAS. Metop-B and Metop-C have co-planar orbits but are approximately half an orbit apart within their orbital plane. As such, their trajectories never intersect or get sufficiently close for measurements from their instruments to be collocated.

Table 3. Number of collocations by day, by satellite combinations. The first row in each cell shows the average number of collocations per day found by both brute-force methods (recall that both brute-force methods yield an identical list of collocations), with the standard deviation of the number of collocations per day in parantheses. The second row shows the same metrics for the rotation-collocation method with sub-occultations, and the third shows the same metrics for the linearized rotation-collocation method.

<table>
<thead>
<tr>
<th>Collocations</th>
<th>NOAA-20</th>
<th>Metop-B-AMSU</th>
<th>Metop-C-AMSU</th>
<th>SNPP</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COSMIC-2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force</td>
<td>124.1 (25.4)</td>
<td>105.9 (24.5)</td>
<td>100.3 (21.4)</td>
<td>124.5 (28.2)</td>
<td>454.8 (71.6)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>125.3 (25.5)</td>
<td>106.1 (24.5)</td>
<td>101.4 (22.0)</td>
<td>124.0 (28.1)</td>
<td>456.7 (72.2)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>125.5 (25.5)</td>
<td>106.0 (24.5)</td>
<td>101.2 (22.0)</td>
<td>123.8 (28.1)</td>
<td>456.3 (72.3)</td>
</tr>
<tr>
<td><strong>Metop-B-GRAS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force</td>
<td>10.6 (17.0)</td>
<td>232.2 (9.9)</td>
<td>0.0 (0.0)</td>
<td>26.7 (26.5)</td>
<td>269.7 (26.4)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>10.7 (17.1)</td>
<td>233.1 (9.6)</td>
<td>0.0 (0.0)</td>
<td>26.9 (26.8)</td>
<td>270.7 (26.5)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>10.7 (17.1)</td>
<td>232.9 (9.6)</td>
<td>0.0 (0.0)</td>
<td>26.9 (26.8)</td>
<td>270.4 (26.5)</td>
</tr>
<tr>
<td><strong>Metop-C-GRAS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force</td>
<td>24.0 (26.0)</td>
<td>0.0 (0.0)</td>
<td>218.0 (44.5)</td>
<td>13.9 (20.0)</td>
<td>246.9 (72.8)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>24.2 (26.3)</td>
<td>0.0 (0.0)</td>
<td>219.1 (44.2)</td>
<td>14.2 (20.5)</td>
<td>248.5 (73.0)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>24.2 (26.3)</td>
<td>0.0 (0.0)</td>
<td>218.8 (44.2)</td>
<td>14.2 (20.5)</td>
<td>248.2 (73.0)</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force</td>
<td>157.2 (35.1)</td>
<td>338.4 (26.9)</td>
<td>311.3 (66.0)</td>
<td>164.6 (36.1)</td>
<td>971.4 (124.1)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>158.6 (35.2)</td>
<td>339.2 (27.1)</td>
<td>313.5 (66.1)</td>
<td>164.6 (36.3)</td>
<td>975.8 (124.7)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>158.6 (35.2)</td>
<td>338.8 (27.2)</td>
<td>313.0 (66.1)</td>
<td>164.5 (36.3)</td>
<td>974.8 (124.8)</td>
</tr>
</tbody>
</table>
### 4.5 Computational expense analysis and accuracy

Table 4 shows the core-minutes per day of RO data required to compute collocations for different combinations of satellites on an 8-core 2020 MacBook Pro with an M1 chip and 16 GB of RAM. The fastest method, the linearized rotation method, takes on average less than one core-minute per day to compute collocations for all satellites and achieves a 328-fold acceleration over the sorted brute-force method. The acceleration by the linearized rotation-collocation method varies depending on the time tolerance and computational hardware used but in general ranges between 40-fold and 400-fold over conventional brute force algorithms.

<table>
<thead>
<tr>
<th>Collocations</th>
<th>NOAA-20</th>
<th>Metop-B-AMSU</th>
<th>Metop-C-AMSU</th>
<th>SNPP</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COSMIC-2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force #1</td>
<td>13.0 (1.7)</td>
<td>7.3 (0.9)</td>
<td>7.4 (1.3)</td>
<td>35.7 (4.9)</td>
<td>63.4 (8.6)</td>
</tr>
<tr>
<td>Brute-force #2</td>
<td>11.5 (1.5)</td>
<td>2.6 (0.3)</td>
<td>4.0 (8.2)</td>
<td>33.6 (4.5)</td>
<td>51.7 (11.6)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>0.3 (0.0)</td>
<td>0.3 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.2 (0.0)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.2 (0.0)</td>
</tr>
<tr>
<td><strong>Metop-B-GRAS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force #1</td>
<td>1.9 (0.0)</td>
<td>0.9 (0.0)</td>
<td>1.1 (0.1)</td>
<td>5.2 (0.2)</td>
<td>9.1 (0.3)</td>
</tr>
<tr>
<td>Brute-force #2</td>
<td>1.7 (0.0)</td>
<td>0.3 (0.0)</td>
<td>0.4 (0.0)</td>
<td>4.9 (0.2)</td>
<td>7.3 (0.2)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.4 (0.0)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td><strong>Metop-C-GRAS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force #1</td>
<td>1.9 (0.1)</td>
<td>1.0 (0.1)</td>
<td>0.9 (0.2)</td>
<td>4.8 (0.9)</td>
<td>8.2 (2.2)</td>
</tr>
<tr>
<td>Brute-force #2</td>
<td>1.7 (0.1)</td>
<td>0.4 (0.0)</td>
<td>0.3 (0.1)</td>
<td>4.6 (0.9)</td>
<td>6.6 (1.8)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.3 (0.1)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brute-force #1</td>
<td>16.7 (1.9)</td>
<td>9.1 (1.0)</td>
<td>9.3 (1.4)</td>
<td>45.5 (5.4)</td>
<td>80.7 (9.5)</td>
</tr>
<tr>
<td>Brute-force #2</td>
<td>14.7 (1.7)</td>
<td>3.2 (0.4)</td>
<td>4.7 (8.2)</td>
<td>42.9 (5.0)</td>
<td>65.6 (12.1)</td>
</tr>
<tr>
<td>Rot.-coll. w/ sub-occ</td>
<td>0.5 (0.0)</td>
<td>0.5 (0.0)</td>
<td>0.6 (0.1)</td>
<td>0.6 (0.1)</td>
<td>2.2 (0.2)</td>
</tr>
<tr>
<td>Linear. rot.-coll.</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.1 (0.0)</td>
<td>0.2 (0.0)</td>
</tr>
</tbody>
</table>

### 4.6 Longer timescale analysis

Recall the second assumption outlined in §2.2: that the apparent position of an RO sounding in a nadir-sounder frame forms a linear trajectory. Over longer timescales, this trajectory elongates and its curvature becomes more apparent. To test the validity of this assumption, we applied all four collocation-finding methods to finding collocations between NOAA-20 ATMS and COSMIC-2 with $\Delta t = 3 \text{ hours}$, a time window 18 times longer than that used for §4.1–§4.5. Increasing the time tolerance in this way greatly increases the number of possible collocations. For the rotation-collocation method with sub-occultations, we used $N = 7$ sub-occultations, or a spacing of $dt = 3600 \text{ s}$ between sub-occultations.
Figure 6. (a) Daily collocations for COSMIC-2 RO soundings and NOAA-20 ATMS microwave radiance soundings for the month of January 2021, with a 3-hour time tolerance. (b) Confusion matrix for the COSMIC-2 RO soundings and NOAA-20 ATMS microwave soundings for January 2021 for the rotation-collocation method with sub-occultations. (c) Map of collocations on January 15, 2021, for NOAA-20 ATMS soundings and COSMIC-2 RO soundings. (d) Confusion matrix for the COSMIC-2 RO soundings and NOAA-20 ATMS microwave soundings for January 2021 for the linearized rotation-collocation method.

Figure 6(a) shows the collocations by day on the left vertical axis and fractional yield of collocations on the right vertical axis for NOAA-20 and COSMIC-2 over January, 2021. With $\Delta t = 10800$ s (3 hours), the linearized rotation-collocation method (light green) finds many more collocations than the brute-force algorithm (blue) and the rotation-collocation method with sub-occultations (orange). It is also apparent that with a time window of 3 hours, around half of all COSMIC-2 RO soundings are collocated with NOAA-20 soundings, many more than with a time window of 10 minutes.

Figure 6(b) shows the performance of the rotation-collocation method with sub-occultations on collocations between COSMIC-2 and NOAA-20 for January, 2021. This method has a true positive rate of $63341/63430 = 99.9\%$ and a true negative rate of $62218/62235 = 99.97\%$. Figure 6(d) shows the performance of the linearized rotation-collocation method on collocations between COSMIC-2 and NOAA-20 for January 2021. This method has a true positive rate of $63315/66221 = 95.6\%$ and a true negative rate of $59401/59444 = 99.93\%$.

Although the linearized rotation-collocation method has many more incorrect predictions than the rotation-collocation method with sub-occultations, it retains a 95.6% true positive rate. This illustrates that even over a three-hour period, the
Table 5. Total number of incorrect predictions (collocations identified by rotation-collocation method but not by the brute-force method), total number of missed predictions (collections missed by rotation-collocation method but found by the brute-force method), and total number of correct predictions (collocations found by both methods) for collocations between NOAA-20 ATMS soundings and COSMIC-2 RO soundings over the month of January 2021 with a 3-hour time tolerance for collocation, for the rotation-collocation method evaluated with a varying number of sub-occultations.

<table>
<thead>
<tr>
<th>Number of sub-occultations</th>
<th>Time between sub-occultations</th>
<th>Incorrect predictions</th>
<th>Missed predictions</th>
<th>Correct predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (same as linearized)</td>
<td>6 hours</td>
<td>2906</td>
<td>43</td>
<td>63315</td>
</tr>
<tr>
<td>3</td>
<td>3 hours</td>
<td>364</td>
<td>49</td>
<td>63309</td>
</tr>
<tr>
<td>4</td>
<td>2 hours</td>
<td>168</td>
<td>35</td>
<td>63323</td>
</tr>
<tr>
<td>5</td>
<td>90 minutes</td>
<td>113</td>
<td>21</td>
<td>63337</td>
</tr>
<tr>
<td>6</td>
<td>72 minutes</td>
<td>95</td>
<td>19</td>
<td>63339</td>
</tr>
<tr>
<td>7</td>
<td>60 minutes</td>
<td>89</td>
<td>17</td>
<td>63341</td>
</tr>
</tbody>
</table>

The apparent computational efficiency gains come about because the brute-force methods are decelerated more rapidly than \((\Delta t)^{-1}\) with longer time tolerance \(\Delta t\). Brute-force methods only do the spatial check for nadir scan soundings that match in time, and so when many more soundings match in time, many more spatial checks are performed, which can be quite slow. This problem is particularly acute for the sorted brute-force method, which is actually the slowest method for a time window of 3 hours. The key advantage of the sorted brute-force method is that it considers many fewer nadir scanner soundings for each
RO sounding than brute-force method #1 does. When the time window is long, this advantage evaporates but the time taken to search for the start and end of the time window in the sorted list of soundings remains, making the time taken by the sorted brute-force method similar to that taken by the brute-force method #1.

Additionally, because some RO soundings may occur at the very beginning or very end of a day, the brute-force methods must consider 30 hours of nadir scanner sounding, beginning 3 hours before the start of the day and ending 3 hours after the end of day, in order to find all collocations for a single day. With a 10-minute time tolerance for collocations, the brute-force methods only need consider 24 hours and 20 minutes of microwave soundings, speeding up the search for collocations. As a result, the acceleration provided by the rotation-collocation method is much more dramatic with $\Delta t = 3$ hours than with $\Delta t = 10$ minutes.

In conclusion, the rotation-collocation method retains remarkable accuracy when the time spacing between sub-occultations is 60 minutes or less. Even with a 3-hour spacing between sub-occultations, the rotation-collocation method retains accuracy above 95%. The time taken by the rotation-collocation method only scales with number of RO soundings and number of sub-occultations, whereas the time taken by the brute-force method scales with time tolerance. This makes the rotation-collocation method an excellent choice for finding collocations with time tolerances of 3 hours or more.

5 Conclusions

The rotation-collocation method has great potential to quickly find collocations between RO soundings and nadir-scan soundings. In fact, the rotation-collocation method generalizes easily, and can be applied to any set of sparsely sampled satellite data and any set of continuously sampled data from a nadir-scanning satellite. When applied to a month’s worth of RO soundings from COSMIC-2, Metop-B-GRAS, and Metop-C-GRAS and a month’s worth of MW soundings from Metop-B-AMSU, Metop-C-AMSU, SNPP, and NOAA-20 with a time tolerance of ten minutes, the linearized rotation-collocation method finds 29994 collocations with a 99.0% true positive rate and a 99.9% true negative rate and has a 328-fold acceleration over the fastest brute-force method. Furthermore, when incorrect predictions that result from missing microwave are held out, the linearized rotation-collocation method achieves a true positive rate of 99.5%. This indicates that when the time tolerance for collocation is low, the linearized rotation-collocation method achieves near perfect accuracy and does so hundreds of times faster.

When applied to a months’ worth of COSMIC-2 RO soundings and NOAA-20 microwave soundings with a three-hour time tolerance for collocation, the rotation-collocation method with sub-occultations spaced sixty minutes apart achieves 99.9% true positive and 99.97% true negative with a 918-fold acceleration over the fastest brute-force method. The linearized rotation-collocation method achieves 95.6% true positive and 99.93% true negative with a 3124-fold acceleration over the brute-force method. This demonstrates that the rotation-collocation method maintains a near perfect accuracy with sub-occultations up to an hour apart, and that the rotation-collocation methods offer an improvement in speed over brute-force methods as the time tolerance for collocation is increased.

Finally, the rotation-collocation method shows that with a ten-minute time tolerance and 150 km spatial tolerance, there were an average of nearly 1000 collocated RO soundings each day of January 2021, or around 16% of all unique RO soundings.
from Metop-B-GRAS, Metop-C-GRAS, and COSMIC-2. Around 40% of Metop-B-GRAS soundings were collocated with Metop-B-AMSU soundings, and around 40% of Metop-C-GRAS soundings were collocated with Metop-C-AMSU soundings. Cohosted instruments on Metop-B and Metop-C greatly increase the percentage of soundings that are collocated, and that cohosting MW and RO instruments is a powerful tool for increasing the number of collocations.

5.1 Future Work and Applications

At present, the rotation-collocation algorithm identifies RO soundings which are collocated with nadir-scanner soundings, but does not identify the specific nadir-scanner sounding associated with each collocation. In the future, the authors plan to extend the rotation-collocation algorithm to identify the specific nadir-scanner soundings associated with each collocated RO sounding, and to integrate this extended version of the rotation-collocation algorithm into NASA’s existing earth science data management software in order to speed up collocation-finding and assimilation of RO data into numerical weather prediction models.

In addition, the rotation-collocation method can be easily modified to identify collocations between two different nadir-scanning satellites. It can also be extended to predict collocation yield for satellite missions with nadir-scanning payloads in different orbits. In this way, the rotation-collocation method can be used as a constellation planning tool and a mission planning tool in order to select collocation-maximizing orbits for nadir-scanning satellites.

Code and data availability. The code associated with this paper will be made available at https://github.com/alexmeredith8299/ro-nadir-collocation. All of the data associated with this paper is freely available from UCAR, NOAA, and EUMETSAT. The code repository contains instructions on how to download the data.

Author contributions. AM performed data analysis and contributed to the manuscript. SL designed the study, provided advice on data analysis, and contributed to the manuscript. KC provided advice on data analysis and on the manuscript. All authors contributed to the interpretation of the results and edited the manuscript.

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

Disclaimer. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.
Acknowledgements. This work was supported by the National Science Foundation under grant GEO-1850089, the NASA Decadal Survey Incubator Program under grant 80NSSC22K1103, and the NSF Graduate Research Fellowship Program under Grant No. 2141064.
References


