Open Spatial Dataset for GNSS and Autonomous Navigation

T.G. Pelham[1], Y. Chen[2], S. Liu[2], Z. Ji[2], E. Anyaegbu[3], R. Wong[3], and R. Grech[3]

[1]University of Bristol, UK, t.g.pelham@bristol.ac.uk
[2]Cardiff University, UK
[3]Spirent Communications, UK

***Abstract*—Global Navigation Satellite System (GNSS) data has become a stable of modern life, whether delivered through a smartphone, vehicle, or as part of an industrial dataset. The expectation of reliable position and time information, irrespective of the environment, becomes challenging in tunnels and canyons with significant multipath, whether natural or urban. An open dataset of Lidar, Computer Vision, Inertial measurement, and GNSS data is presented, combined with preliminary analysis of the dataset on a selected recording, considering GPS coarse acquisition, lidar and computer vision based navigation together with model based channel reconstruction. Lidar and video based techniques demonstrate positioning error as low as 3.8m root mean square error, together with GNSS acquisition and signal blockage alignment with computer vision and lidar maps.**

***Index Terms*—GNSS, GPS, LIDAR, Antennas, Beamforming, Computer Vision**

I. Introduction

The advent of autonomous vehicles and advanced driver-assistance systems (ADAS) has revolutionized the transportation landscape, promising a future of safer and more efficient mobility. However, the realization of this vision hinges on the development of robust positioning systems that can operate reliably in challenging environments. Global Navigation Satellite System (GNSS) signals, the mainstay of modern positioning systems, are susceptible to various impairments, including signal blockage, multipath propagation, and interference. These impairments can significantly degrade positioning accuracy, posing a critical challenge for autonomous systems [1].

In order to address these challenges, there are two complementary approaches to the development of robust positioning in the emerging electromagnetic environment addressed in this work. The first is the radio domain; considering techniques based upon radio propagation, beamforming, or machine learning based signal processing. The second is based upon an understanding of the spatial domain, provided by sensors such as LIDAR or computer vision, considering different techniques in sensor fusion in order to provide a robust understanding of position and environment to the host system such as an autonomous vehicle.

A. Radio Domain

The radio domain approaches can be subdivided further into, exploitation of the spatial component of the signal, in order to provide a reliable position from GNSS sources. Beamforming approaches can be used with antenna arrays in order to provide additional directivity and selectivity [2], this can be further enhanced using inertial navigation to reduce the acquisition time in a noisy environment [3]. The signal domain can be exploited with RF Fingerprinting approaches to provide protection against spoofing or source replication degrading the GNSS signals [4]. All of these approaches can be considered using the presented dataset, in addition to the use of the synchronised spatial information to the development of reflectometry and survey methods based upon GNSS signals [5].

B. Spatial Domain and Information Fusion

For autonomous robots and vehicles, state estimation, mapping, and localisation are functions which also rely on path planning, control, and many other capabilities. The vision-based and lidar-based methods are widely used in state estimation for online simultaneous localisation and mapping (SLAM), which can provide the robot with six DOF state estimations. However, single-sensor-based systems are not reliable, lidar-based methods have motion point cloud distortion; the vision-based methods have interference caused by illumination changes and dynamic surroundings. Sensor fusion is commonly adopted for more recent SLAM algorithms, merging multiple types of data from different sensors to reduce the uncertainty of the system.

The dataset can be applied to sensor fusion-based SLAM, such as lidar-based methods, including lidar inertial odometry via smoothing and mapping (LIO-SAM) [6], and vision-based methods such as VINS-MONO [7] and VINS-FUSION [8]. LIO-SAM is a well-known tightly coupled lidar SLAM system which fuses the lidar and imu in the system. VINS-MONO and VINS-FUSION are examples of visual-inertial odometry (VIO) systems. The difference between the two algorithms is the former one adopted a mono-camera, and the latter one adopted a pair of stereo cameras. Both algorithms have the best performance in underground scenarios where the moving objects are minimal as the relative motion of plants and vehicles leads to an accumulated error [9]. These algorithms all have their limitations, which may be addressed by utilising more sensors for better performance. The algorithms proposed in [10] and [11] fuse IMU, lidar, and camera to overcome the limitations of the lidar-based and visual-based methods. Alternatively the GPS message can be included, which is also supported by the presented dataset given the synchronised lidar, IMU, and RF recordings [12].

II. Measurement

The measurement campaign for this dataset was planned with the consideration of a number of challenging problems in reliable GNSS positioning and autonomous navigation. Natural and Urban canyons and parking structures present both significant attenuation and multipath, and in order to address this in the dataset for future research groups the Clifton Gorge was selected as a significant natural canyon, also providing both an overhead concrete canopy and the Clifton Suspension Bridge as a potential source of multipath interference (Figure 3 & 2). The Ladies Mile on the Clifton downs was selected as a suitable reference source, providing a straight tree lined road, without major buildings. Lewins Mead and Rupert Street at the heart of urban Bristol were selected as ideal measurement paths for their significant high rise buildings. The final recording was performed descending into the underground carpark of the Merchant Venturers Building, providing a transition measurement from outdoors with no significant obstructions to GNSS signals, to two floors underneath the building. A list of each recording, and their start and end points is included in Table I.

A. Recording Setup

The mobile recording setup used for the data collection consisted of a mounting board, with the cameras, Novatel receiver, antennas, and cameras mounted directly, together with a frame supporting the LIDAR, microstrain IMU, and Ublox GPS. The board was then attached securely to the roof rack of the car (Toyota Auris), and the power and data connections are provided to a Robot Operating System (ROS) recorder mounted internally, together with the GSS6450 GNSS Signal recorder. Antenna 1 is connected directly to the recorder, while Antenna 2 & 3 are connected both to the recorder, and as the primary and secondary antenna for the Novatel Receiver, with locations given in the open source dataset [14].

- Ouster OS0-128 Lidar
- Microstrain 3dm-gx-45 IMU
- Ublox evk-m8c GPS
- Spirent GSS6450 GNSS Signal Recorder
- Novatel GNSS-850 GPS Receiver
- Luxonis S2 W OAK Cameras, arranged at -30^o,0^o,30^o in azimuth with respect to the direction of travel.

For IMU data, one Microstrain 3dm-gx-45 IMU collects the 9-axis-data with the frequency of 100Hz. The lidar generates a 3D point cloud at 10Hz with a range of 100m, and 128 vertical lines. The computer vision component is provided by three Luxonis S2 W OAK cameras, each comprised of three individual cameras, one RGB camera and a pair of stereo cameras; all three lenses are set with a resolution of 720P on 30Hz. All the data is labelled and synchronised through the Robot Operating System (ROS) built-in time. The ROS time is aligned with the GPS time through the GPS receiver's messages. The ground truth positioning is provided by the Novatel SPAN-CPT system. The recorded IQ data

Fig. 1. Measurement Setup mounted on car, with Lidar, Camera Array, and three GPS antennas

was then stored as IQ data within a Hierarchical data format (HDF5), providing an open source format with compression, and parallel IO access [13]. A single HDF5 file was created for each measurement within the dataset, to be paired with a rosbag file for each. The rosbag contains the timestamped IMU, Position, Camera, and LIDAR data. In order to support the widest array of considered applications, the measurement timestamped position and inertial measurement data is also replicated within the RF HDF5 file, together with the solution information from the Novatel GPS receiver, and the measurement start and finish times from the GSS6450 recorder, allowing coarse time synchronisation between the RF and spatial datasets. The dataset is available in full under a creative commons license for both the RF and spatial portions [14, 15].

Name	Start	Finish	Duration (s)	Bandwidth (MHz)
Ladies Mile (Southwest)	51.473546,-2.622330	51.462820,-2.626245	187.5	10
Clifton Gorge (South)	51.457493,-2.629310	51.450292,-2.624581	105.3	10
Lewins Mead	51.455012,-2.596772	51.458161,-2.592167	120.8	10
Rupert Street	51.459073,-2.590861	51.455363,-2.596414	107.9	30
Clifton Gorge (North)	51.448524,-2.618736	51.461431,-2.628727	176.9	30
Ladies Mile (Northeast)	51.464722,-2.626950	51.472554,-2.620549	140.4	30
Underground Carpark	51.455746,-2.602437	51.455788,-2.602907	64.5	30

TABLE I
Dataset Measurement Locations, with description, start and finish locations (latitude,longitude), and duration (seconds)

III. Results

The analysis of the full dataset presents a significant challenge, with the unprocessed dataset comprising 631GB. In order to consider the applications of interest however, initial processing is presented on the Clifton Gorge (South) recording, both in terms of the wider terrain using available digital surface models, and in terms of the Lidar, computer vision, and RF measurements.

A. Environment Modelling

The measurement environment was populated using the National Lidar Program digital surface model for the area, creating a surface mesh for alignment with Lidar measurements from within the dataset [16]. This digital recreation of the terrain was implemented within LyceanEM, an open source electromagnetics package for rapid virtual prototyping of antenna arrays on complex platforms, and channel modelling in the time and frequency domain [17, 18]. This package

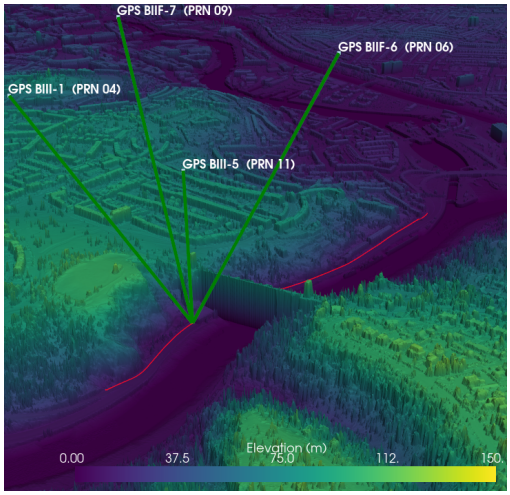

Fig. 2. Clifton Gorge Recording Path - Southbound. The ground truth path is shown in red, with GPS satellite direction lines and labels for point B.

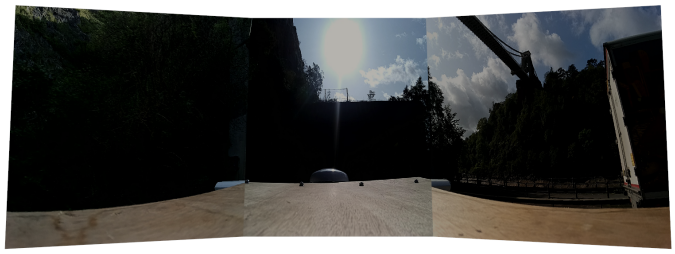

Fig. 3. Merged Visualisation of the primary RGB camera on each Luxonis S2 W camera at point B (Figure 2), showing the natural canyon wall (Left, the concrete canopy, and the Clifton Suspension Bridge (above right).

allows for the creation of digital twins of the GPS satellites, their antennas, and the scattering environment between each considered satellite and the antenna array mounted on the car while in motion through the environment.

Algorithm	**RMSE (m)**	σ
LIO-SAM	3.77	1.51
VINS-MONO	149	78.8
VINS-FUSION	474	263.7

TABLE II
COMPARISON OF ROOT MEAN SQUARE ERROR AND STANDARD DEVIATION IN POSITION FOR LIO-SAM, VINS-MONO, AND VINS-FUSION ALGORITHMS.

Three of the discussed sensor fusion based algorithms were used with the Clifton Gorge dataset in order to compare both lidar and vision based SLAM with the ground truth GPS track. The resultant comparison is included in Table II. All three methods considered suffered from the relative motion in the environment, with the vision based approaches demonstrating excessive, progressive errors, while the lidar based approach is still to large to be acceptable in a truly autonomous vehicle.

B. Coarse Acquisition Signal Detection

A parallel code phase search algorithm (Figure 4) was implemented using python to provide an initial coarse acquisition of the dataset [19]. Based upon a selected coherent integration time of 1ms, and the sampling rate of 10.3057MHz, frequency bins of 700Hz from -10kHz to 10kHz were selected. In order to align with the Lidar sampling rate, 1054 sample points were considered through the recording (Clifton Gorge-South, Table I).

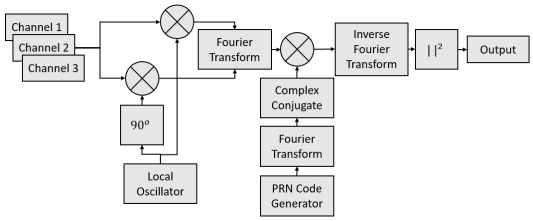

Fig. 4. Parallel Code Phase Search Algorithm implemented for GPS coarse acquisition

The course acquisition metrics relate the detected peak in the search algorithm to the second largest peak, as shown in Figure 5, 6, & 7. Point A corresponds to the start of vehicle motion from the initial position, the concrete canopy shown in Figures 2 & 3 is marked by point B at the start, and point C as the vehicle exits the canopy in each figure.

During this measurement period, there were four GPS satellites above 40^o in elevation from the measurement path, with pseudo-random-noise codes (PRN code) 4, 6, 9, and 11. PRN 4 which is occluded most significantly by the canyon wall (Figure 2) is only weakly visible. PRN 9 can clearly be seen through the recording, aside from the portion underneath the concrete canopy, while PRN 6 is more intermittent. PRN 11 was not observed in the dataset. Between points A and B, the two channel Novatel receiver maintained between 13 and 15 satellites in the positioning solution, degrading to 0 satellites between point B and C, and then returning to 10 satellites after 1.6 seconds, and a stable 14 satellites after 17.6 seconds.

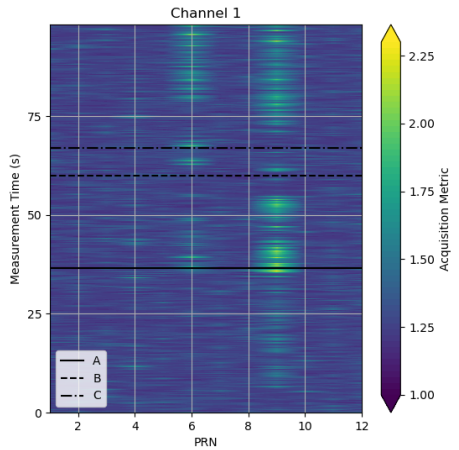

Fig. 5. Coarse Acquisition Metrics for Channel 1, with the start of mobile recording shown as A, the beginning of concrete canopy as B, and the exit from the canopy as C

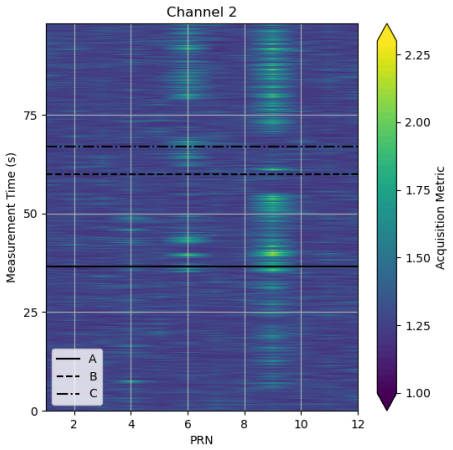

Fig. 6. Coarse Acquisition Metrics for Channel 2, with the start of mobile recording shown as A, the beginning of concrete canopy as B, and the exit from the canopy as C

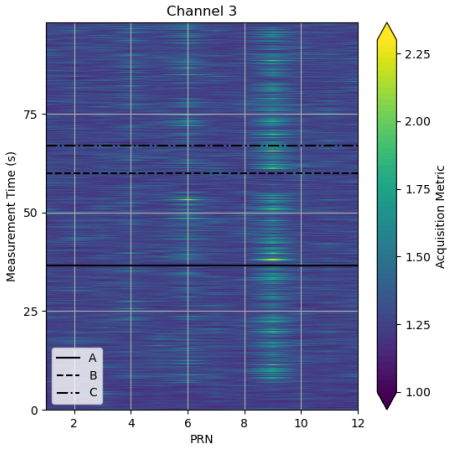

Fig. 7. Coarse Acquisition Metrics for Channel 3, with the start of mobile recording shown as A, the beginning of concrete canopy as B, and the exit from the canopy as C

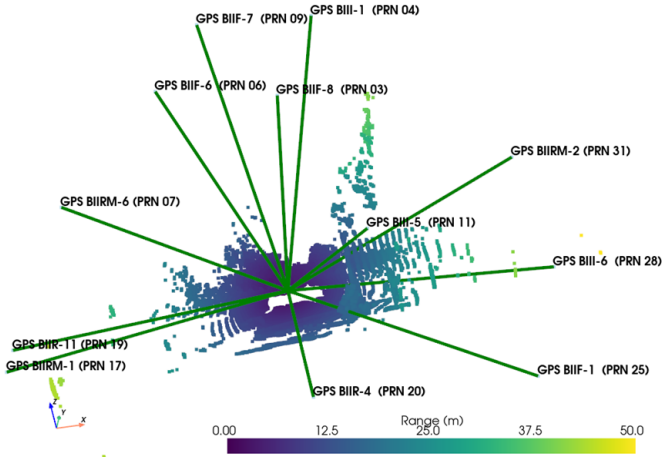

Fig. 8. Lidar Snapshot for approach to concrete canopy, with satellite bearings denoted by labelled green lines.

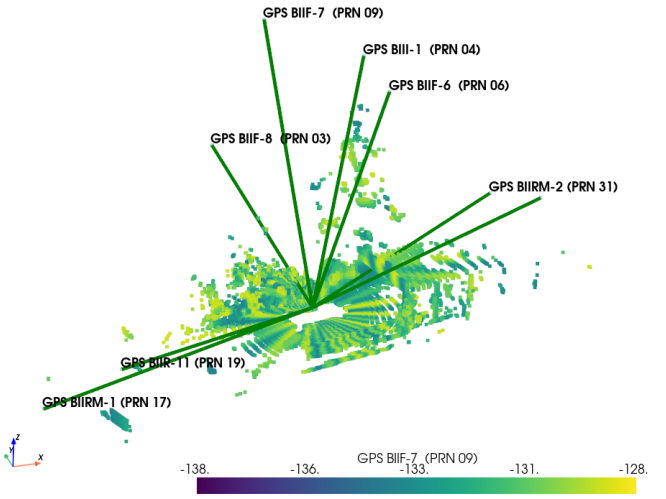

Fig. 9. Beamforming map from the LyceanEM channel reconstruction using LIDAR snapshot, modelling the expected environment scatter from GPS BIIF-7 (PRN 09)

This is consistent with the measurement vehicle moving out from the steeper parts of the gorge and towards Bristol Floating Harbour, with less obstructions close to the Portway. LyceanEM was used to perform channel reconstruction based upon a Lidar snapshot collected before moving underneath the concrete canopy (Figure 8), using the receiving array to beamform within the computational environment to identify significant scattering within the local environment (Figure 9).

IV. Conclusion

The presented combination of open electromagnetics model, open dataset, software defined GPS acquisition and SLAM algorithms provides an initial look at the utility of the open spatial dataset. While the initial software defined GPS acquisition is limited to signal detection, further development is planned to support research into the merging of the spatial and RF domains via LyceanEM. This package uses the CUDA architecture for acceleration, and has been run on mobile devices such as the Nvidia AGX Orin. This provides the opportunity for development of sensor fusion algorithms which can convert the spatial domain sensed via Lidar or computer vision into a radio channel model in a mobile device. The challenge in the creation of an accurate and robust positioning strategy with interrupted GNS systems has been demonstrated using the dataset with published algorithms, with root mean square errors approaching 500m. Future work will focus on the use of this and future datasets to develop robust approaches, merging the use of spatial intelligence and onboard channel models with sensor fusion algorithms to provide reliable, accurate positioning information.

ACKNOWLEDGMENT

This project was supported by the Royal Academy of Engineering, the Office of the Chief Science Adviser for National Security under the UK Intelligence Community Postdoctoral Research Fellowship programme, and Spirent Communications.

V. REFERENCES

[1] Rigas T. Ioannides, Thomas Pany, and Glen Gibbons. "Known Vulnerabilities of Global Navigation Satellite Systems, Status, and Potential Mitigation Techniques". In: *Proceedings of the IEEE* 104.6 (June 2016), pp. 1174–1194. ISSN: 1558-2256. DOI: 10.1109/jproc.2016.2535898. URL: http://doi.org/https://doi.org/10.1109/jproc.2016.2535898.

[2] Yuting Ng and Grace Xingxin Gao. "GNSS Multireceiver Vector Tracking". In: *IEEE Transactions on Aerospace and Electronic Systems* 53.5 (Oct. 2017), pp. 2583–2593. ISSN: 0018-9251. DOI: 10.1109/taes.2017.2705338. URL: http://doi.org/https://doi.org/10.1109/taes.2017.2705338.

[3] Qiang Li et al. "A Robust Anti-Jamming Navigation Receiver with Antenna Array and GPS/SINS". In: *IEEE Communications Letters* 18.3 (Mar. 2014), pp. 467–470. ISSN: 1089-7798. DOI: 10.1109/lcomm.2014.012314.132451. URL: http://doi.org/https://doi.org/10.1109/lcomm.2014.012314.132451.

[4] Debashri Roy et al. "GANSAT: A GAN and SATellite Constellation Fingerprint-Based Framework for GPS Spoof-Detection and Location Estimation in GPS Deprived Environment". In: *IEEE Access* 10 (2022), pp. 45485–45507. ISSN: 2169-3536. DOI: 10.1109/access.2022.3169420. URL: http://doi.org/https://doi.org/10.1109/access.2022.3169420.

[5] Joan Francesc Muñoz-Martín et al. "Stokes Parameters Retrieval and Calibration of Hybrid Compact Polarimetric GNSS-R Signals". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–11. ISSN: 1558-0644. DOI: 10.1109/tgrs.2022.3178578. URL: http://doi.org/https://doi.org/10.1109/tgrs.2022.3178578.

[6] Tixiao Shan et al. "LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping". In: (Oct. 2020). DOI: 10.1109/iros45743.2020.9341176. URL: http://doi.org/https://doi.org/10.1109/iros45743.2020.9341176.

[7] Tong Qin, Peiliang Li, and Shaojie Shen. "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator". In: *IEEE Transactions on Robotics* 34.4 (Aug. 2018), pp. 1004–1020. ISSN: 1941-0468. DOI: 10.1109/tro.2018.2853729. URL: http://doi.org/https://doi.org/10.1109/tro.2018.2853729.

[8] Tong Qin et al. *A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors*. 2019. arXiv: 1901.03642 [cs.CV].

[9] Li-Ta Hsu et al. "UrbanNav:An Open-Sourced Multisensory Dataset for Benchmarking Positioning Algorithms Designed for Urban Areas". In: *Proceedings of the Satellite Division's International Technical Meeting*. GNSS 2021 (Oct. 2021). ISSN: 2331-5954. DOI: 10.33012/2021.17895. URL: http://doi.org/https://doi.org/10.33012/2021.17895.

[10] Tixiao Shan et al. "LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping". In: (May 2021). DOI: 10.1109/icra48506.2021.9561996. URL: http://doi.org/https://doi.org/10.1109/icra48506.2021.9561996.

[11] Shibo Zhao et al. "Super Odometry: IMU-centric LiDAR-Visual-Inertial Estimator for Challenging Environments". In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Sept. 2021). DOI: 10.1109/iros51168.2021.9635862. URL: http://doi.org/https://doi.org/10.1109/iros51168.2021.9635862.

[12] Shengyu Li et al. "Multi-GNSS PPP/INS/Vision/LiDAR tightly integrated system for precise navigation in urban environments". In: *Information Fusion* 90 (Feb. 2023), pp. 218–232. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2022.09.018. URL: http://doi.org/https://doi.org/10.1016/j.inffus.2022.09.018.

[13] Sandeep Koranne. "Hierarchical data format 5: HDF5". In: *Handbook of Open Source Tools*. Springer, 2011, pp. 191–200.

[14] Timothy Pelham et al. *GNSS Combined Vehicular Measurements - RF*. Jan. 2024. DOI: 10.5523/bris.33bzdi4wbwhhg2cznap5gl3hns. URL: https://data.bris.ac.uk/data/dataset/33bzdi4wbwhhg2cznap5gl3hns.

[15] Timothy Pelham et al. *GNSS Combined Vehicular Measurements - ROS*. Mar. 2024. DOI: 10.5523/bris.2s2dhb906jzdb2jt1xwltvkppx. URL: https://data.bris.ac.uk/data/dataset/2s2dhb906jzdb2jt1xwltvkppx.

[16] Environment Agency. *National LIDAR Programme*. HMG. Oct. 2022. URL: https://www.data.gov.uk/dataset/f0db0249-f17b-4036-9e65-309148c97ce4/national-lidar-programme.

[17] T. G. Pelham. "Rapid antenna and array analysis for virtual prototyping". In: (2022). DOI: 10.1049/icp.2022.2330. URL: http://doi.org/https://doi.org/10.1049/icp.2022.2330.

[18] Timothy Pelham. "LyceanEM: A python package for virtual prototyping of antenna arrays, time and frequency domain channel modelling". In: *Journal of open source software* 8.86 (June 2023), p. 5234. ISSN: 2475-9066. DOI: 10.21105/joss.05234. URL: http://doi.org/https://doi.org/10.21105/joss.05234.

[19] Kai Borre et al. "A Software-Defined GPS and Galileo Receiver". In: *Birkhäuser Boston eBooks* (2007). DOI: 10.1007/978-0-8176-4540-3. URL: http://doi.org/https://doi.org/10.1007/978-0-8176-4540-3.