

A DATA-INTENSIVE APPROACH TO EXPLOIT NEW GNSS SCIENCE OPPORTUNITIES

EXPLOTACIÓN DE NUEVAS OPORTUNIDADES CIENTÍFICAS DE LOS SISTEMAS DE POSICIONAMIENTO GLOBAL POR SATÉLITE (GNSS) DESDE UNA PERSPECTIVA INTENSIVA EN DATOS

Vicente Navarro*, Javier Ventura-Traveset

Navigation Science Office, European Space Agency, Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Spain.
vicente.navarro@esa.int; javier.ventura-traveset@esa.int

Abstract:

With the current GNSS infrastructure development plans, over 120 GNSS satellites (including European Galileo satellites) will provide, already this decade, continuous data, in several frequencies, without interruption and on a permanent basis. This global and permanent GNSS infrastructure constitutes a major opportunity for GNSS science applications. In the meantime, recent advances in technology have contributed "de-facto" to the deployment of a large GNSS receiver array based on Internet of Things (IoT), affordable smart devices easy to find in everybody's pockets. These devices – evolving fast at each new generation – feature an increasing number of capabilities and sensors able to collect a variety of measurements, improving GNSS performance. Among these capabilities, Galileo dual band smartphones receivers and Android's support for raw GNSS data recording represent major steps forward for Positioning, Navigation and Timing (PNT) data processing improvements. Information gathering from these devices, commonly referred as crowdsourcing, opens the door to new data-intensive analysis techniques in many science domains. At this point, collaboration between various research groups is essential to harness the potential hidden behind the large volumes of data generated by this cyberinfrastructure. Cloud Computing technologies extend traditional computational boundaries, enabling execution of processing components close to the data. This paradigm shift offers seamless execution of interactive algorithms and analytics, skipping lengthy downloads and setups. The resulting scenario, defined by a GNSS Big Data repository with co-located processing capabilities, sets an excellent basis for the application of Artificial Intelligence / Machine Learning (ML) technologies in the context of GNSS. This unique opportunity for science has been recognized by the European Space Agency (ESA) with the creation of the Navigation Scientific Office, which leverages on GNSS infrastructure to deliver innovative solutions across multiple scientific domains.

Key words: GNSS, Cloud Computing, Artificial Intelligence, Internet of Things, Open Science, Geodesy

Resumen:

Con los planes actuales de desarrollo de la infraestructura GNSS, más de 120 satélites GNSS (incluidos los satélites europeos Galileo) proporcionarán, ya en esta década, datos continuos, en varias frecuencias, sin interrupciones y de forma permanente. Esta infraestructura GNSS global y permanente constituye una gran oportunidad para las aplicaciones científicas de GNSS. Mientras tanto, avances recientes han contribuido al despliegue de una red GNSS paralela basada en la Internet de las Cosas (IoT), asequibles dispositivos inteligentes fáciles de encontrar en todos los bolsillos. Estos dispositivos, que evolucionan rápidamente con cada nueva generación, acumulan un número creciente de funcionalidades y sensores capaces de recopilar una gran variedad de mediciones. Entre estas funcionalidades, los receptores de teléfonos inteligentes de banda dual Galileo y el soporte Android para la grabación de datos GNSS sin procesar representan pasos especialmente relevantes. La recopilación de información mediante estos dispositivos, comúnmente conocida como crowdsourcing, abre la puerta a nuevas técnicas de análisis de datos en múltiples dominios científicos. Llegados a este punto, la colaboración entre diversos grupos de investigación resulta esencial para aprovechar el potencial que se esconde en los grandes volúmenes de datos generados por esta ciberinfraestructura. Las tecnologías de *Cloud Computing* extienden los límites computacionales tradicionales permitiendo la ejecución de componentes de procesamiento cerca de los datos. Este cambio de paradigma ofrece una rápida ejecución de algoritmos y análisis interactivos, omitiendo largas descargas y configuraciones. El escenario resultante, definido por un repositorio GNSS Big Data con capacidades de procesamiento acopladas, establece una base excelente para la aplicación de tecnologías de Inteligencia Artificial / Aprendizaje Automático (ML). Esta oportunidad única para la ciencia ha sido reconocida por la Agencia Espacial Europea (ESA) con la creación de la Oficina Científica de Navegación, que aprovecha la infraestructura GNSS para ofrecer soluciones innovadoras en múltiples dominios científicos.

Palabras clave: GNSS, *Cloud Computing*, Inteligencia Artificial, Internet de las Cosas, Ciencia Abierta, Geodesia

* Corresponding Author: Vicente Navarro, vicente.navarro@esa.int

1. Introduction

Over an estimated 175 zettabytes of data produced in the world, connected smart devices forming the Internet of Things (IoT) are expected to carry out 80% of the data processing and analysis by 2025 (EC 2020). This computational model close to the user ('Edge Computing') nicely complements centralised facilities ('Cloud Computing') to enable new science opportunities offered by Crowdsourcing, where citizens become distributed observatories ('Citizen Science').

Simultaneously, in the GNSS arena, Space and Ground Segments of all major players (Galileo, GPS, Beidou and Glonass) will undergo continuous upgrades leading to new features and increased levels of reliability, precision and accuracy. On the user segment side, in addition to advancements in professional GNSS receivers, an extremely dynamic IoT mass-market led by smartphones, steadily delivers technological breakthroughs in terms of chipset miniaturisation (Sony 2020), multi-constellation multi-frequency, power efficiency or data processing (Diggelen 2021).

Therefore, the scenario depicted by the all-weather, long-term, stable, high quality, worldwide GNSS physical infrastructure and the intrinsic characteristics of their signals presents a unique opportunity for science interests.

As already happened with TRANSIT, which leveraged on Doppler based navigation discovered while tracking Sputnik, in the era of Big Data and Machine Learning (ML), the continuous flow of data generated by GNSS signals from GNSS satellites offers a research value that breaks original boundaries defined by positioning, navigation and timing (PNT) services.

The ubiquity of both, GNSS signals and GNSS-enabled IoT receivers, has reached a situation where scientists devise situations where unavailability of GNSS signals represents a source of information that, properly processed, can support generation of 3D city models (UCL 2021).

Nevertheless, full realisation of the scientific potential of GNSS systems as a signal of opportunity (SOOP) demands extensions to current approaches for data processing, recording and product generation. Among these extensions, long-term preservation and systematic recording of GNSS signal digitized intermediate frequency (LaChapelle and Broumandan 2016; Navarro *et al.* 2019) represents a key aspiration to achieve fine grain information resolution and re-processing capabilities demanded for innovative science use cases.

In addition to the technical challenges derived from high data volumes and processing requirements, effective scientific exploitation involves efficient collaboration across different research groups working together towards a common goal. For this purpose, complex cyberinfrastructures, commonly known as Thematic Exploitation Platform (Navarro *et al.* 2019; Nikutta *et al.* 2020,) foster the creation of essential synergies regarding data discovery, access and analysis.

2. GNSS Signal Fundamentals

GNSS signals contain all required information to continuously estimate satellite to receiver travelling time

and satellite coordinates. Methods and algorithms behind these signals are well known and ample literature is available (Sanz *et al.* 2013). Three main components define the core characteristics of GNSS signal definition and processing:

- **Carrier:** Radio frequency sinusoidal signal.
- **Ranging code:** binary sequences allowing calculation of radio signal travel time. They are known as Pseudo-Random Noise (PRN) sequences, PRN codes or spreading codes.
- **Navigation data:** binary sequence providing ancillary information on satellite ephemeris, clock bias, almanac, satellite health status, ionosphere data and other complementary information.

These main components aim to support the two key target measurements in object of GNSS signal processing, pseudorange and carrier phase measurements, which can be defined as in Eq. (1):

$$P = cd\tau = \rho + d_\rho + c(dt-dT) + d_{trop} + d_{ion} + \varepsilon_{mp} + \varepsilon_p \quad (1)$$

$$\Phi = \rho + d_\rho + c(dt-dT) + \lambda N + \lambda w - d_{ion} + d_{trop} + \varepsilon_{m\Phi} + \varepsilon_\Phi$$

Where:

P = Code (or pseudorange) measurement

Φ = Carrier phase measurement

dt = time of transit of signal τ

c = speed of light in vacuum

ρ = geometric range

d_ρ = orbital errors

dt = satellite clock offsets

dT = receiver clock offsets

d_{ion} = ionospheric delay

d_{trop} = tropospheric delay

ε_{mp} = code multipath noise

ε_p = code receiver noise

λ = frequency wavelength

N = integer ambiguity

λw = wind-up effect

$\varepsilon_{m\Phi}$ = phase multipath noise

ε_Φ = phase receiver noise

As introduced by Eq. (1), multiple sources of errors affect both measurements in different ways and orders of magnitude. Fortunately, the vast experience accumulated over years of operations in GNSS systems, has led to the development of multiple signal and data processing techniques able to provide very good estimates for the contribution of each error under most common circumstances.

Nevertheless, a closer look at the extremely high precision expected from one of the GNSS basic observables, the time of transit of signal τ , highlights the difficulty of this endeavour. As a convenient approximation, following rule of thumb provides an idea of

GNSS sensibility with regard to time alterations. Taking into account that the theoretical maximum resolution from code is about 1 percent of the spreading code chipping rate, 1 nanosecond delay would imply 30 cm for Precision code (P-code) and for Coarse/Acquisition code (C/A code), due to its x10 lower chipping rate, it would imply 3 meters.

Therefore, this landscape has triggered the creation of a large human and physical infrastructure aiming at the precise categorisation and understanding of all different elements that contribute to signal disturbances.

3. GNSS Infrastructure Fundamentals

GNSS infrastructure does not differ in its organisation from other space missions, featuring the traditional three segments split:

- *Space Segment:* it comprises the satellites in charge of transmitting carrier, ranging code and navigation data elements previously presented.
- *Ground Segment:* it carries out continuous monitoring and update to ensure overall stability and availability of the system.
- *User Segment:* it comprises the community of users making use of GNSS receivers.

The level of maturity and adoption reached by GNSS systems allows this infrastructure to feature uncommon characteristics in space missions, namely, cost-effectiveness, with satellites created in batches, excellent spatial-temporal resolution in all three segments, and public, free of charge accessibility.

When it comes to GNSS infrastructure, it is important to distinguish between global and Satellite-Based Augmentation Systems (SBAS). The latter have many similarities and even share elements with global systems, complementing them to provide higher accuracy in limited areas for a restricted number of users. This restriction makes SBAS less relevant for data crowdsourcing purposes, moving the focus to the four global constellations currently available, Galileo (Europe), GPS (USA), Beidou (China) and GLONASS (Russia).

Along the lines of data crowdsourcing, we will zoom-in to the user segment, where we can differentiate between professional and recreational groups depending on their utilisation of the GNSS signal. The first group gathers industrial and research organisations in navigation related domains. The second group represents end-users making use of GNSS for daily activities.

While not formally part of the ground segment, the first group is an essential active contributor to the scientific relevance of GNSS. More specifically, operated as a service of the International Association of Geodesy (IAG), the International GNSS Service (IGS) has been providing free and open access to high-precision GNSS data and products for twenty-five years already.

Initially focused on the delivery of the IGS core reference frame, orbit, clock and atmospheric products, over the years IGS is evolving into a multi-GNSS service as more sites expand core IGS network (500 stations approx.).

IGS operations entails data collection, archiving and dissemination of observation data sets from contributing global networks of tracking stations (Fig.1).

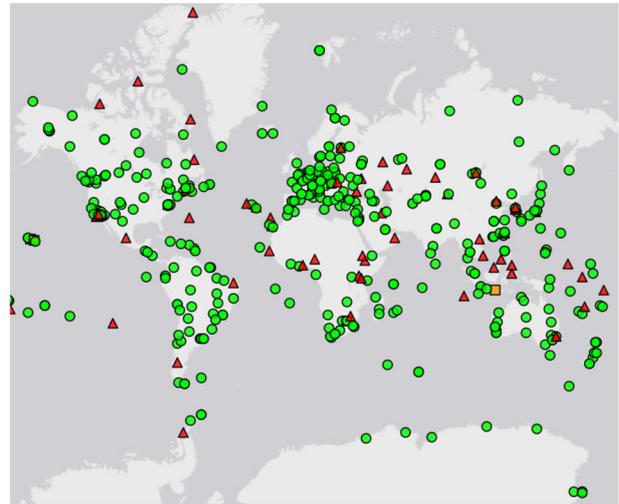


Figure 1: IGS Core Network, data received within last 10 days (green), last 30 days (orange), unhealthy (red). Source IGS.

Several Global IGS Data Centres (Navarro *et al.* 2019) replicate this data to ensure fast and fault tolerant access across the community. Furthermore, IGS coordinates and monitors the quality of a products generated by a network of Analysis Centres, combining them to produce high quality official IGS products.

IGS Final products provide the highest quality level, made available on a weekly basis with a delay up to 20 days. The IGS Final products are the basis for the IGS reference frame and they are intended for applications demanding high consistency and quality.

IGS Rapid products represent the second level in terms of quality. They become available on a daily basis with a delay of about 17 hours following the end of the observation day. Quality degradation from IGS Final to Rapid products is not significant for most use cases.

IGS Ultra-rapid products, initially conceived as predicted products and released every six hours, they are intended for real time and near real time use cases. The Ultra-rapid products lead to significantly improved orbit predictions and reduced errors for user applications.

Moreover, relying on this very same infrastructure of stations network, data centres and analysis centres, IGS provides real time access to GNSS orbit and clock corrections, key for Precise Point Positioning (PPP).

Finally, IGS distributes key ancillary data collected from instruments co-located with GNSS receivers, such as local weather data highly relevant for troposphere post-processing use cases.

The global, coordinated effort of IGS members ensures that data and products adhere to agreed standards and strict quality checks, generating a repository readily available for science exploitation. Moreover, many other organisations contribute to create a rich ecosystem of regional data sources and GNSS receivers that complement IGS, like the Nevada Geodetic Laboratory, further increasing spatial and temporal density of GNSS data.

From the point of view of signals and infrastructure, we could conclude that GNSS sets the basis for a worldwide

large, virtual instrument and a highly distributed science data processing pipeline.

4. GNSS Science Opportunities

As presented in Eq. (1), the need for precise quantification of the different elements affecting the GNSS signal has spawned the creation multiple research fields. Moreover, quite a number of examples demonstrate GNSS far-reaching impact in terms of science opportunities.

Originally conceived for navigation on Earth, utilisation of GNSS in space has become routine in the LEO region covered by the Terrestrial Service Volume at 0 – 3.000 km altitude. Despite the small fraction of satellites using GNSS above LEO, improved coverage (Fig. 2) has increased interest in the development of the Space Service Volume (SSV) at 3.000 - 36.000 km (GEO) altitude (UN 2018).

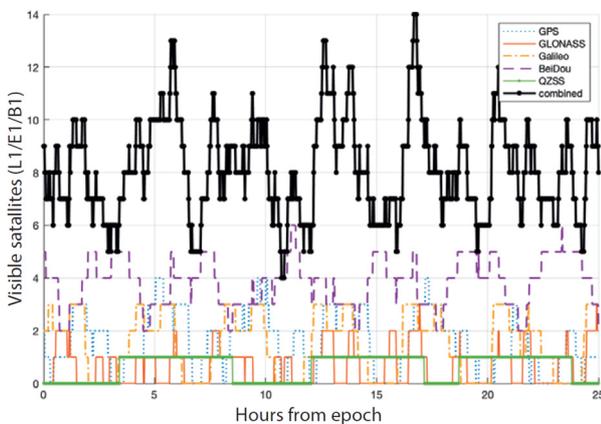


Figure 2: Estimated number of satellites visible, by individual constellation and combined, for sample L1/E1/B1 GEO user with 20 dB-Hz C/No threshold. (UN 2018).

Reached this point, travelling well beyond GEO altitude, we start devising a Lunar Navigation System where advancements in knowledge of GNSS transmit antenna patterns and promising results from experiments regarding the utilisation of a GPS-only receiver in a Moon Transfer Orbit (MTO) have demonstrated the potential of GNSS based Lunar Navigation (Delépauta *et al.* 2020).

In the field of SSV, the utilisation of GIOVE-A to demonstrate autonomous orbit determination at high altitudes with a GNSS receiver showcased, early on in its life, Galileo’s potential to go beyond its original purpose. Later on, Galileo became a precious testbed for Fundamental Physics (in addition to nominal use). GNSS satellite clocks, subject to both Special Relativity (SR) and General Relativity (GR) effects, correct these effects considering an almost perfectly circular orbit. The special characteristics of Galileo Satellites 5 and 6, located in an eccentric orbit due to a launch anomaly, combined with the frequency stability of Galileo’s Passive Hydrogen Maser (PHM) clocks and highly precise orbits, allows to verify GR predictions below current “state of the art” observing the periodic signature in their clock rate. Based on this concept, the GREAT project assessed data accumulated for over 1000 days, improving gravitational redshift determination for the very first time since Gravity Probe (Delva *et al.* 2019).

Back to Earth, multiple GNSS applications derive from its signal interaction with the environment or data fusion geolocalisation:

- Troposphere and Weather forecast: specific humidity, precipitation
- Ionosphere and Space Weather: electron density.
- Climate: sea level, sea ice coverage, ice shelf altitude, salinity
- Disasters: tsunami early warning, flood monitoring.
- Land surfaces: soil moisture, biomass, snow cover and depth, humidity content of snow

On the climate side, as global warming progresses, water vapour increases, with potential increases in intensity of precipitation and decreases in frequency (Allan and Soden 2008). GNSS complements systems such as radiosondes and remote sensing providing complementary coverage of water vapour. Remote sensing of Atmospheric Water Vapour Using the Global Positioning System (Bevis *et al.* 1992) positions GNSS at the service of weather forecast and climate change monitoring. GNSS networks like IGS provide long-term time series to support climatological requirements for global coverage, spatial resolution and homogeneity. IGS produces continuous estimates of vertically integrated water vapour content. The limited coverage over the oceans can be estimated from space-borne GNSS receivers using reflectometry techniques.

On a different topic, but still related to climate change, animal tracking with GNSS sensors provides bio-indicators of change. Animals offer in-situ stations that interact with key environments through “behaviours” like ecology of movement, identification of potential threats, conservation measures. Birds have already shown changes in reproduction and migration events linked to warming (Masello *et al.* 2021). Penguins forage in sectors of rich energy landscape where low energy is required from them. Changes in currents or temperatures affecting their energy landscape forced them to forage in expensive parts. Results show that lower foraging costs may favour a higher breeding success, explaining the positive population trend by the Gentoo penguins in the Antarctic Peninsula.

In the field of 3D city modelling, UCL (2021) has demonstrated feasibility to estimate height and some types of materials of buildings from the blockage, reflection, and attenuation of GPS (and in general Global Navigation Satellite Systems (GNSS)) data. Using semi-crowdsourced GNSS raw data from smartphone volunteers, the research has managed to extract some patterns applying machine learning (ML) techniques. This approach could provide a ubiquitous and free of charge 3D models creation/update instrument.

Travelling into space, Pajares and Moreno (2020) build on their indicator Global Navigation Satellite Systems Solar FLare (GSFLAI) to detect solar flares and quantify the associated extreme ultraviolet (EUV) solar flux rate based on over-ionization, measured from hundreds of IGS dual-frequency receivers. A generalisation of GSFLAI is presented for the much weaker stellar superflares. The new algorithm, Blind GNSS search of Extraterrestrial EUV Sources (BGEES) is able to detect EUV flares without the previous knowledge of the position of the source, which is

also simultaneously estimated. BGESS results concerning the detection and location of two stellar superflares, Proxima Centauri (18 March 2016, 08:32UT) and NGTS J121939.5-355557 (1 February 2016, 04:00UT), strongly suggest the possibility to extend the technique, also in real time.

5. Machine Learning for GNSS

In traditional programming, developers build systems writing sequences of instructions to describe very carefully what to do to achieve a goal. Machine learning changes this in a fundamental way, enabling computers to learn from data to achieve a goal without being explicitly programmed (Samuel 1959). While it has been around for quite some years already, substantial increase in data generation rates along with advancements in processing power and software libraries has led to a new golden age of machine learning. Commonly associated to computer vision, speech recognition or games, there is a large portion of complex use cases where it is significantly easier to define a behaviour in terms of input and output data.

The popular term Software 2.0 (Karpthy 2017), adopted for the machine learning programming paradigm, reflects its revolutionary impact in software development. Machine Learning fundamentally departs from Software 1.0 “classical stack”, where programmers use languages such as Java, Python, C++, etc. to reach some desirable behaviour point in the program space (Fig. 3). Now the quest is delegated to algorithms designed to autonomously reach the aforementioned desired state through a training process that makes use of abstract weights and hyperparameters to govern the behaviour of a neural network. Human intervention shifts towards understanding those features driving the expected behaviour and the subsequent selection of the artificial model that best suits observed characteristics like for example temporal correlation.

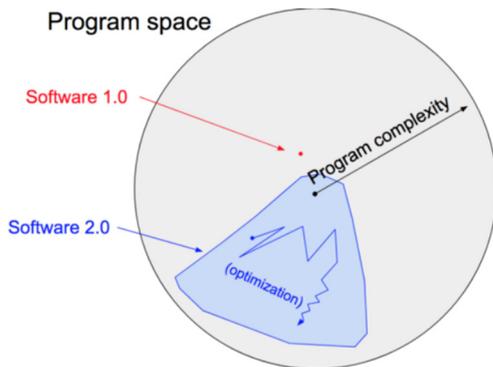


Figure 3: Software 2.0 space.

In fact, many well-known problems in the GNSS arena, traditionally formulated as mathematical expressions, fit within the premise defined by the problem of improving some measure of performance P when executing some task T , through some type of training experience E . (Mitchell 1997).

In analogy with biology, the basic element of a Neural Network is a neuron, which can be defined as in Eq. (2):

$$a = g(W_x x + b) \quad (2)$$

Where:

a = neuron output vector

g = activation function

W_x = neuron weight

x = neuron input vector

b = bias term

Neural networks present different characteristics depending on activation functions and architecture. For example, Recurrent Neural Networks (RNN) are known for their ability to account for temporal correlation in the data. However, in their first implementation, they are not efficient memorizing long-term information. The Long Short Term Memory (LSTM) adds an internal memory concept, allowing information to flow further into the network. Many more efficient architectures exist. There are many alternatives and architectures, for a complete review on one could refer to (Goodfellow *et al.* 2016).

A review of literature focused on the application of machine learning to GNSS data shows significant utilisation of neural networks, time series, and classification technologies. Areas of interest go from the core GNSS goal, improving localization (Hosseinyalamdary 2018; Diggelen 2019) to all other bias effect presented in Eq. (1) like multipath (Diggelen and Wang 2020; Quan *et al.* 2018), ionosphere (Orus 2018; Jiao *et al.* 2007; Liu *et al.* 2020; Linty *et al.* 2019) and troposphere (Benevides *et al.* 2019) effects. These works demonstrate how ML can contribute to address problems addressed in a different way.

ML principle lays behind the so-called “hidden layers”. During training, weights defined for each neuro-to-neuron connection evolve to maximise task performance. When applied to GNSS the basic inputs, usually the residuals, are complemented with “fused” ones relevant for the task like, C/N0, 3D models (Diggelen 2021), constellation type, signal type, etc. The architectural support for data fusion presents ML as a very promising alternative to other analytical solutions with rigid definitions of their input parameters. False ionospheric problems (Benton and Mitchell 2011), would greatly benefit from ML capability for seamless integration of data. ML algorithms can work at signal level (Navarro *et al.* 2019) to assess automatically the quality of the received signal.

In order to realise ML benefits, explainability, reliability, security and privacy challenges (EC 2020) require a systemic approach throughout the ML lifecycle comprising data preparation, feature extraction, training and validation.

- **Data Preparation:** this is the first and most critical one as good data represents the basis for all ML subsequent steps. It is in charge of filtering, sampling, cleaning and basic transformation.
- **Feature Extraction:** this step is dedicated to the extraction of information, which will be provided as input to the Machine Learning Modules.
- **Training:** during this step, the Machine Learning Modules embedding the algorithms construct the model.

- **Validation:** it implies comparing real with predicted data in two steps. The first one refers to data used during the model training while the second refers to data absent during model construction.

Assessment of GNSS ML algorithms implies processing simulated or real data collected via GNSS receivers for a set of scenarios. Outputs compared with a “ground truth” derive ML performance metrics driving an iterative process.

6. Raw GNSS Data, Internet of Things and Crowdsourcing

On the 14th of April 2021, EUSPA announced the mark of two billion Galileo-enabled smartphones in the market, increasing the dominance of smartphones in the GNSS user segment. As we will see, this is particularly relevant when it comes to the utilisation of the second frequency, E5. A few months earlier, Sony (2020) had announced de release of a multi-constellation, dual-band GNSS Receiver for wearable devices with the industry’s lowest power consumption to date. Besides, since Android 7, the possibility to access GNSS raw measurements on smartphones fosters development of new processing techniques and optimisations (GSA 2017). Hence, fast-paced advances in mass-market technology provide, “de-facto”, a large GNSS receiver network of affordable, smart and ubiquitous devices, sometimes referred as the Internet of Things (IoT).

Science initiatives to leverage on the potential represented by this GNSS-enabled “virtual instrument” expand to domains with a long tradition of dedicated instruments like Astronomy (CRAYFIS 2021) or 3D mapping (UCL 2021). These initiatives rely on a crowdsourcing process, characterised by a collaborative effort to obtain the required data or service from a large group of people (usually undefined volunteers). This process is becoming increasingly popular as it provides a cost-effective solution to collect a large volume of data. There are a variety of applications, such as image labelling, object counting, translation or slogan design (Marcus and Parameswaran 2013).

Few references in GNSS literature devise crowdsourcing approaches for GNSS science. De Oliveira *et al.* (2020), proposed a crowdsourcing concept to estimate troposphere water vapour distribution from GNSS using a simulated smartphone network. However, one of the key feasibility problems is the definition of the effective minimum sample for crowdsourced data (Meng 2018). Without careful assessment of data quality, quantity and problem complexity, the apparently large crowdsourced data may effectively account for a small fraction of the full scenario.

Finally, the aforementioned dual-frequency represents a game changer for smartphone accuracy (Fig. 4). On one hand, the dispersive nature of the ionosphere allows to remove the ionospheric effect using two-frequency measurements. On the other hand, the narrower correlation peak L5/E5 signals improves accuracy and multipath resilience.

Moreover, exploitation of dual-frequency and ancillary data, like 3D maps, enables unprecedented solutions for urban canyons accuracy like 3DMA GNSS (Diggelen and Wang 2020).



Figure 4: Dual-frequency test at the European Space Technology Centre (ESTEC). Samsung S8 (left), Xiaomi Mi-8 (right). Source ESA.

7. Ionosphere Science Case

The ionosphere is a dispersive medium that extends from about 50km up to more than 1300 km in terrestrial atmosphere. As presented in Eq. (1) the ionosphere plays a fundamental role in GNSS, as signal propagation speed varies with its electron density, measured as Total Electron Content (TEC). Perturbations in the ionosphere go from global to local effects (a few hundreds of km range) and their temporal range varies between seconds and days. Physical background differs and impact on GNSS may vary from a slight delays causing accuracy degradation to Loss-of-Lock (LoL). Hence, beyond the great interest for scientific research, monitoring and understanding of the ionosphere is critical for a large number of daily life applications that rely on radiosignals.

For multi-frequency GNSS receivers under nominal conditions, ionosphere-free combination allows to remove the ionospheric effect. Single frequency receivers depend on ionospheric model information broadcasted in the navigation message (Klobuchar, NeQuick). These models present a correction capability from 50% to 70%. Furthermore, IGS, in coordination with its Analysis Centres, distributes post-processed Global Ionosphere Maps (GIM). These maps are more precise than the broadcast model by approximately 20%, and they can be used as references for ionospheric studies (Orus 2018).

Monitoring of ionospheric parameter like TEC, combined with indexes like ROTI, S4 or σ_ϕ , acting as proxy of ionospheric perturbations (Borries *et al.* 2020) provide a solid dataset for Machine Learning predicting models.

Traditionally, GNSS methods to deal with scintillations events rely on thresholds applied to amplitude and phase scintillation indices S4 and σ_ϕ . In presence of noise and disturbances, this method suffers from high number of false alarms and missed detections.

Figure 5 shows an example where threshold discrimination only some values would be part of the scintillation events. However, human inspection or sophisticated ML models (SVM) identify a longer duration of the event (Savas and Dosis 2019).

Therefore, the goal of Machine Learning in this case is to avoid such missed detections, identifying (similarly to humans) contribution to the same scintillation event, or excluding outliers.

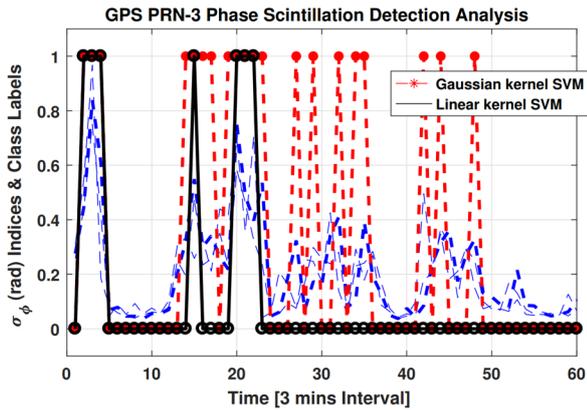


Figure 5: Linear and Gaussian Kernel SVM Result, Data Set: June 6, 2016, SANAE IV.

Despite the nature of the scintillation events does not make straightforward the identification of correlations (Jiao *et al.* 2014), purely data-driven approaches are competing with state-of-the-art, model-based methods. Moreover, the flexibility of neural networks permits further refinements using additional inputs, such as solar imagery (Boulch *et al.* 2018).

Liu *et al.* (2020) LSTM model forecasts spherical harmonic coefficients used for GIM. The model is trained with Dst index, EUV flux and SH coefficients time-series. Reported results showed that the model was competitive against empirical models (IRI-2016 and NeQuick2) performing well in quiet and storm geomagnetic conditions.

Orus (2018) shows promising results in the application of DNN to estimate the ionosphere mapping function. The study highlights NN capability to derive STEC for 150 stations on Earth.

However, current ground station networks spatial resolution limits data availability. Exploiting GNSS Big Data crowdsourcing would allow to “close the gaps”.

The availability of dual band smartphone and the possibility of extracting ionospheric delay from their GNSS raw data (GSA 2017) opens the door to a substantial increase in current spatial resolution by combination of the dual-frequency carrier phase observations (e.g. GPS L1 and L5, or Galileo E1 and E5a). While Orus (2018) delivers a global model of the ionosphere based on historical data, this approach provides a distributed model based on real-time data collected from a network of smartphones.

8. Troposphere Science Case

As presented in Eq. (1) tropospheric delay is one of the biases affecting GNSS code and phase measurements (Sanz *et al.* 2013). This delay has wet and hydrostatic components. The hydrostatic component represents approximately 90% of the tropospheric delay. This part has slow variability and its behaviour is quite predictable based on local temperature and atmospheric pressure. Models mainly rely on pressure data providing estimates in the order of millimetres. The estimation of the wet delay presents more difficulties (Bevis *et al.* 1992).

To address this problem GNSS community offers solutions based on the IGS network of geodetic antennas

and receivers, capable of carrying out precise carrier phase GNSS measurements on multiple frequencies. IGS Data Centres provide post-processed troposphere products like total zenith path delay (ZPD) and north/east troposphere gradient. Besides, availability of surface pressure and temperature measurements allow further extraction of precipitable water vapour.

However this solution provides a limited spatial resolution derived from the costs associated to the high precision GNSS receivers required. Crowdsourced ZPD estimates would provide the opportunity to increase the spatial resolution of troposphere maps, thus allowing for improved estimation of the drag coefficient important in the reentry of a spacecraft, climate characterization and severe weather events forecasting. The feasibility of ZPD retrieval from GNSS receivers and antennas embedded in smartphones was demonstrated by (Tagliaferro *et al.* 2019).

9. GNSS Science Support Centre (GSSC), Thematic Exploitation Platform

GSSC Thematic Exploitation Platform brings together GNSS science use cases, machine learning frameworks and IoT streaming requirements under an n-tier big data architecture for open science. Organised around several system domains (Fig. 6), it relies on a complex technology stack (Navarro *et al.* 2019) that spans across Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) computing spaces.

- *User Layer*: this layer provides the Human-Machine Interface (HMI) for users and administrators to access all functionalities. The objective of this layer is to decouple presentation logic from business logic implemented by other layers. This layer allows smooth integration of HMI functionalities into a homogeneous look & feel through the provision of extension points.
- *Exploitation and Preservation Layer*: this layer groups domains implementing generic and user specific analysis functionalities. These domains provide access to information and processing assets integrated in GSSC. The system relies on executable modules and data at different levels of processing which are natively stored or federated to other systems. These two types of assets are combined in the exploitation layer to deliver more complex services and products.
- *Support Layer*: this layer groups domains implementing components that include libraries providing common features required across the whole system. These libraries act as glue-code to adapt to component needs.
- *Infrastructure Layer*: this layer provides basic support for the implementation of the preservation and support layers. It is based on COTS that can be reused “as-is” with an integration pattern mainly based on the configuration of a set of parameters to adapt the behaviour of the COTS to the specific needs.

This architecture supports innovative GNSS Science Use cases through seamless integration of remote, edge-

based GNSS Intermediate Frequency Recording Stations (LaChapelle and Broumandan 2016; Navarro *et al.* 2019).

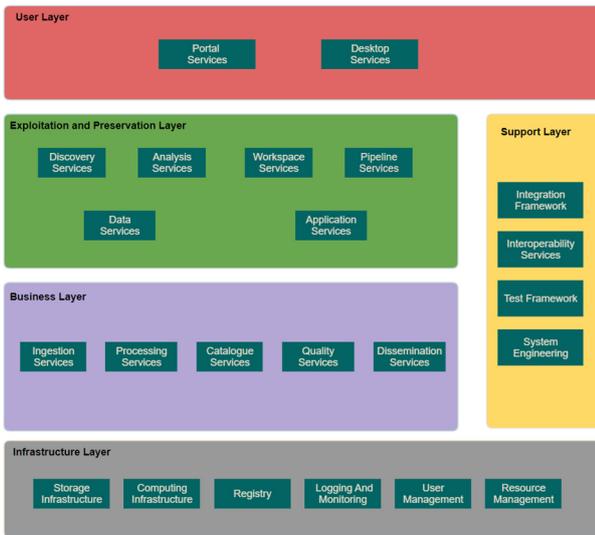


Figure 6: N-Tier Architecture and System Domains.

10. First Results

GSSC’s repository further extends IGS Global Data Centre with ESA’s GNSS related data from projects (GREAT, GESTA) and missions. Taking note of the increasing trend in space missions to embark GNSS receivers, the GSSC repository already integrates SVS data from ESA (SWARM, GOCE) and non ESA missions (SAC-C, JASON, CHAMP, ICESAT). This approach aims to promote and maximise scientific return from this highly relevant data, taking advantage of ESA’s privileged position in the field.

Furthermore relevant data from other organisations is also available, like Environmental Monitoring Unit data from GSAT0207 and GSAT0215 satellites.

First internal tests have demonstrated GSSC’s suitability to act as a catalyser for GNSS research, students and external organisations. These groups can build their own analysis tools, sitting on top of this repository, to address scenarios like:

- Fast development of prototype ideas requiring initial provisioning of complex systems.
- Software-as-a-Service delivery of desktop applications and systems.
- Training framework for coding competitions, trainees, small external developments.
- Outreach and demonstration of systems requiring complex initial configuration.

GSSC platform is already available at gssc.esa.int for selected private beta users, having successfully completed its core development phase. It features core analysis capabilities and a preliminary catalogue of GNSS data collections.

This platform, characterised by the move of processing components to data, provides native support for open science. It evolves from the traditional FTP solution to provide an integrated environment with advanced services for data discovery (Fig. 7) and analysis (Fig. 8).

This cyberinfrastructure puts the focus on the science community, promoting user contributions to the platform as data and computing extensions.

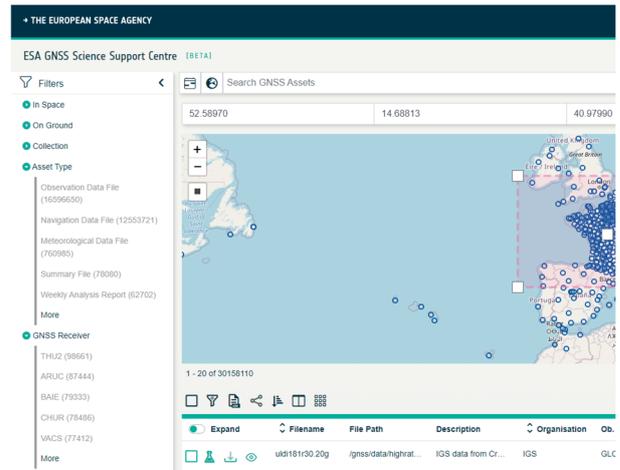


Figure 7: GSSC Discovery Service.

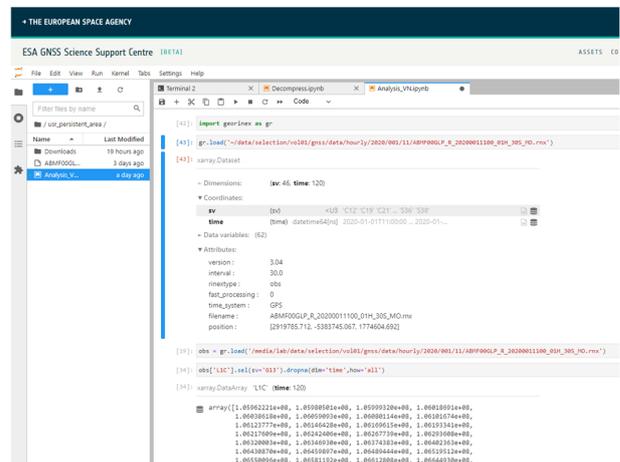


Figure 8: GSSC Analysis Service – Custom JupyterLab.

GSSC Analysis Services provide on-demand web-access to exploitation tools. These tools range from general domain ones like Octave to GNSS-specific ones designed to tackle a particular problem. Prominent example of these general domain systems is JupyterLab. In addition to the vanilla versions of JupyterLab, GSSC provides become notebooks further customised by ESA for specific GNSS analysis.

GNSS Big Data infrastructure (Navarro *et al.* 2019), deployed at three sites, has reached stability during the operations campaign allowing the evaluation of target experiments.

Finally, first results shown in Fig. 9 from dual-frequency Xiaomi Mi8 smartphone measurements show promising results to derive ionospheric delay and related TEC from its geometry-free observable.

11. Conclusions and Future Work

In this work, we have presented GNSS science achievements and opportunities. We have provided up-to-date status information on several key activities regarding challenges and trade-offs for the exploitation of crowdsourced data, applying machine learning to GNSS

science use cases. The described GSSC architecture has successfully demonstrated its suitability as a Thematic Exploitation Platform.

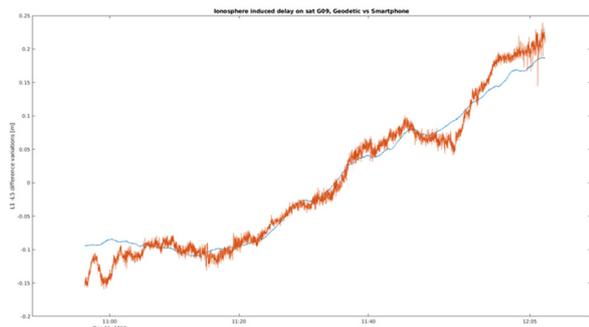


Figure 9: Geometry-free L1-L5 combination obtained from a Xiaomi Mi8 smartphone (red) and a co-located geodetic station (blue).

GSSC next steps gear towards the evolution from currently restricted, beta system into a public environment for open-science and innovation across multi-disciplinary GNSS research communities. Moreover, throughout 2021, new ESA missions and projects are to contribute to build up GSSC catalogue with new applications, notebooks and data collections. Machine Learning exploitation of GNSS crowdsourced data for ionosphere and troposphere use cases the will harness the full potential of this platform.

Key activities related to these use cases encompass:

- Exploitation of Space Service Volume data. Assessment of data fusion for ionosphere modelling (Zakharenkova *et al.* 2016).

- Smartphone data pre-processing improvements to deal with specially noisy measurements.
- Smartphone GNSS antenna phase robot calibration to correct for antenna impact on L1 and L5 phase observations.
- Engagement of third parties for data crowdsourcing required to move away from simulated data.
- Deployment of data processing infrastructure for real time acquisition and distribution of ML-based ionosphere and troposphere corrections.
- Refine approach towards model governance and data provenance (FAIR GNSS). These are two hot topics for implementing AI solutions. In this area, ESA studies on blockchain technology have defined potential contributions.

Acknowledgements

This work was supported by the European Space Agency as part of Research and Development Programmes under Science and Navigation Directorates.

The authors would like to thank the GNSS Science Advisory Committee and ESA Navigation Support Office for their support and suggestions. We also thank our Industrial partners, involved in science use cases assessment and implementation. Thanks also to the Science and Operations technical IT Unit at ESAC supporting the deployment of the GSSC Thematic Exploitation Platform. We would like to thank all data collection providers, with special thanks to IGS, ILRS, CDDIS, BKG and IGN for their sustained and remarkable support making possible the creation of the GSSC Repository at the core of this work.

References

- ALLAN, R., and SODEN, B., 2008. Atmospheric Warming and the Amplification of Precipitation Extremes. *Science*, **321**, Issue 5895, pp. 1481-1484. DOI: 10.1126/science.1160787
- BENEVIDES, P., CATALAO, J., and NICO, G., 2019. Neural Network Approach to Forecast Hourly Intense Rainfall Using GNSS Precipitable Water Vapor and Meteorological Sensors. *Remote Sens.* **2019**, **11**(8), p. 966. DOI: 10.3390/rs11080966
- BENTON, C., and MITCHELL, C., 2011. GPS satellite oscillator faults mimicking ionospheric phase scintillation. *GPS Solut*, **16**, pp. 477–482. DOI 10.1007/s10291-011-0247-3
- BEVIS, M., BUSINGER, S., HERRING, T. A., ROCKEN, C., ANTHES, R. A., and WARE, R. H. (1992). GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System. *Journal of Geophysical Research: Atmospheres*, **97**(D14), pp. 15787-15801.
- BOULCH, A., CHERRIER, N., and CASTAINGS, T., 2018. Ionospheric activity prediction using convolutional recurrent neural networks. arXiv preprint arXiv:1810.13273.
- BORRIES, C., WILKEN, V., JACOBSEN, K.S., GARCIA, A., DZIAK-JANKOWSKA, B., KERVALISHVILI, G., ZAKOWSKI N., TSAGOURI, I. PAJARES, M., FERREIRA, A., and HOQUE, M., 2020. Assessment of the capabilities and applicability of ionospheric perturbation indices provided in Europe. *ScienceDirect. Advances in Space Research*, **66** (2020), pp. 546–562.
- CRAYFIS. 2021. The app that turns your phone into a cosmic ray detector. No joke. Available: <https://crayfis.io/> [5/22, 2021].
- DE OLIVEIRA, T., MÄKELÄ, M., MONTLOIN, L., LEHTOLA, T., THOMBRE, S., and LEHTOLA, V., 2020. Towards tropospheric delay estimation using GNSS smartphone receiver network. *Science Direct, Advances in Space Research*, **2020**. DOI: 10.1016/j.asr.2020.09.041

- DELÉPAUTA, A., GIORDANO, P., VENTURA, J., BLONSKIA, SCHÖNFELDT, M., SCHOONEJANS, P., AZIZA, S., and WALKERA, R., 2020. Use of GNSS for lunar missions and plans for lunar in-orbit development. *21 May 2020 Science Direct*. DOI: 10.1016/j.asr.2020.05.018
- DELVA, P., PUCHADES, M., SCHÖNEMANN, E., DILSSNER, F., COURDE, C., BERTONE, S., GONZALEZ, F., HEES, A., LE PONCIN, C., MEYNADIER, F., PRIETO, R., SOHET, B., VENTURA, J., and WOLF, P., 2019. A new test of gravitational redshift using Galileo satellites: The GREAT experiment. *March-April 2019 Science Direct*. DOI: 10.1016/j.crhy.2019.04.002
- DIGGELEN, F., 2021. Keynote Presentation at the ION ITM Meeting, Jan 2021. Available: <https://sites.google.com/corp/view/frankvandiggelen/videos> [5/22, 2021].
- DIGGELEN, F., 2019. *A-GPS: Assisted GPS, GNSS, and SBAS*. Artech House, London, 2019
- DIGGELEN, F., and WANG, J., 2020. *Improving urban GPS accuracy for your app*. Available: <https://android-developers.googleblog.com/2020/12/improving-urban-gps-accuracy-for-your.html> [5/22, 2021].
- EC. 2020. European Commission - White Paper on Artificial Intelligence. A European approach to excellence and trust. Brussels, 19.2.2020, COM (2020) 65 final.
- GOODFELLOW, I., BENGIO, Y., and COURVILLE, A., 2016. *Deep Learning*. MIT Press, 2016. Available: <http://www.deeplearningbook.org> [5/22, 2021].
- GSA. 2017. White Paper Using GNSS Raw measurements on Android devices. Available: https://www.euspa.europa.eu/system/files/reports/gnss_raw_measurement_web_0.pdf [5/22, 2021].
- HOSSEINYALAMDARY, S., 2018. Deep Kalman Filter: Simultaneous Multi-Sensor Integration and Modelling; A GNSS/IMU Case Study. *Sensors (Switzerland)*, **18**(5), p. 1316,
- JIAO, Y., MORTON, Y., and TAYLOR, S., 2014. Comparative studies of high-latitude and equatorial ionospheric scintillation characteristics of GPS signals. *IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, pp. 37-42.
- KARPATY, A., 2017. Software 2.0. Available: <https://karpaty.medium.com/software-2-0-a64152b37c35> [5/22, 2021].
- LACHAPELLE, G., and BROUMANDAN, A., 2016. Benefits of GNSS IF data recording. *European Navigation Conference (ENC). 2016*. DOI: 10.1109/EURONAV.2016.7530544
- LINTY, N., FARASIN, A., FAVENZA, A., and DOVIS, F., 2018. Detection of GNSS ionospheric scintillations based on machine learning decision tree. *IEEE Transactions on Aerospace and Electronic Systems*, **55**(1), pp. 303-317.
- LIU, L., ZOU, S., YAO, Y., and WANG, Z., 2020. Forecasting Global Ionospheric TEC Using Deep Learning Approach. *Space Weather*, **18**(11). DOI: 10.1029/2020SW002501
- MARCUS, A., and PARAMESWARAN, A., 2013. Crowdsourced Data Management: Industry and Academic. *Perspectives Foundations and Trends in Databases*, **6**(1-2), pp. 1–161. DOI: 10.1561/19000000044
- MASELLO, J. F., BARBOSA, A., KATO, A., MATTERN, T., MEDEIROS, R., STOCKDALE, J. E., and QUILLFELDT, P., 2021. How animals distribute themselves in space: energy landscapes of Antarctic avian predators. *Movement ecology*, **9**(1), pp. 1-25.
- MENG, X., 2018. STATISTICAL PARADISES AND PARADOXES IN BIG DATA (I). *The Annals of Applied Statistics 2018, Institute of Mathematical Statistics*, **12**(2), pp. 685–726. DOI: 10.1214/18-AOAS1161SF
- MITCHELL, T. 1997. *Machine Learning* (McGraw Hill, 1997), ISBN 978-9384761165
- NAVARRO, V., ÁLVAREZ, R., DEL RÍO, S., GUERRA, R., MARTÍN-RUBIO, V., DIEGO, M. Á., and MARSEILLE, M., 2020. ESA Datalabs: Multi Mission Science Exploitation and Preservation Platform. In *Astronomical Society of the Pacific Conference Series*, **527**, p. 287.
- NAVARRO, V., DEL RIO, S., CASTILLO, M., MARTIN, F., and VENTURA, J., 2019. GNSS Science Support Centre. *International Colloquium on Scientific and Fundamental Aspects of GNSS*, **2019**. DOI: 10.13140/RG.2.2.24797.23525.
- NAVARRO, V., DITTRICH, R., SKABURSKAS, K., YING, Y., BÉGIN M.E., and PEREZ, F., 2019. Big Data GNSS for Intermediate Frequency Recording Stations. *Big Data in Space*, **2019**. DOI: 10.13140/RG.2.2.19920.64005
- NIKUTTA, R., FITZPATRICK, M., SCOTT, A., and WEAVER, B.A., 2020. DataLab — A community science platform. NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Ave, Tucson, AZ 85719, USA: DOI: 10.1016/j.ascom.2020.100411
- ORUS, R., 2018. Using TensorFlow-based Neural Network to estimate GNSS single frequency ionospheric delay (IONONet). ScienceDirect, Nov 2018. *Advances in Space Research*, **63**, pp. 1607–1618
- PAJARES, M., and MORENO, D., 2020. Real-Time Detection, Location, and Measurement of Geoeffective Stellar Flares From Global Navigation Satellite System Data: New Technique and Case Studies. *Space Weather*, **18**, e2020SW002441, 2020. DOI: 10.1029/2020SW002441

- QUAN, Y., LAU, L., ROBERTS, G.W., MENG, X., and ZHANG, C., 2018. Convolutional Neural Network Based Multipath Detection Method for Static and Kinematic GPS High Precision Positioning. *Remote Sens.* 2018, **10**, p. 2052. DOI: 10.3390/rs10122052.
- SAMUEL, A. L., 1959. Some studies in machine learning using the game of Checkers (1959). *IBM Journal of Research and Development*, pp. 71-105. DOI: 10.1147/rd.33.0210
- SANZ, J., JUAN, J.M., and HERNANDEZ, M., 2013. *GNSS Data Processing*, Vol. I: Fundamentals and Algorithms (ESA TM-23/1, May 2013), ISBN 978-92-9221-886-7
- SAVAS, C., and DOVIS, F., 2019. Comparative Performance Study of Linear and Gaussian Kernel SVM Implementations for Phase Scintillation Detection. *2019 International Conference on Localization and GNSS (ICL-GNSS)*. DOI: 10.1109/ICL-GNSS.2019.8752635
- SONY. 2020. Sony to Release High-Precision GNSS Receiver LSIs for IoT and Wearable Devices. Available: <https://www.sony.com/en/SonyInfo/News/Press/202008/20-069E/> [5/22, 2021].
- TAGLIAFERRO, G., GATTI, A., and REALINI, E., 2019. Assessment of GNSS Zenith Total Delay Estimation using Smart Devices. *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, Miami, Florida, September 2019, pp. 3879-3891. DOI: 10.33012/2019.17074
- UCL. 2021. Indicative Data: Extracting 3D Models of Cities from Unavailability and Degradation of Global Navigation Satellite Systems (GNSS). Available: <https://gtr.ukri.org/projects?ref=MR%2FS01795X%2F1> [5/22, 2021].
- UN. 2018. The Interoperable Global Navigation Satellite Systems Space Service Volume, United Nations Office for Outer Space Affairs, October 2018. ISBN 978-92-1-130355-1
- JIAO, Y., HALL, J. J., and MORTON, Y. T., 2017. Performance evaluation of an automatic GPS ionospheric phase scintillation detector using a machine-learning algorithm. *NAVIGATION, Journal of the Institute of Navigation*, **64**(3), pp. 391-402.
- ZAKHARENKOVA, I., ASTAFYEVA, E., and CHERNIAK, I., 2016. GPS and in situ Swarm observations of the equatorial plasma density irregularities in the topside ionosphere. *Earth Planets and Space, Springer/Terra Scientific Publishing Company*, **68**, pp.120.