Machines
WithVision

Location and positioning in the railway

Why do we need accurate location?

Accurate positional data and location is needed for superior Infrastructure Monitoring (IM), leading to gains in process, safety, productivity, time and of course cost.

Executive summary

This article addresses the key points surrounding the data and devices needed to determine accurate position.

- The issues caused by a reliance on GNSS/INS and their consequences
- How other sources of position can not only bridge the gap, but enhance your data - across the board
- The benefits of better positioned data for prediction and cross discipline analysis

The article will discuss the technologies used in railway positioning – and particularly those used in Infrastructure Monitoring. We will discuss the difficulties in positioning in the railway environment, the types of corrections that are used currently; and finally discuss what is possible with newer technologies. We will also discuss how all of these benefit and hinder IM by looking at specific use cases of some of the technologies. We welcome feedback and discussion on all of these topics, the aim is to dispel any myths, address any misconceptions and quantify the art of the possible.

How does the railway use location?

Location is needed in the railways for (amongst other things):

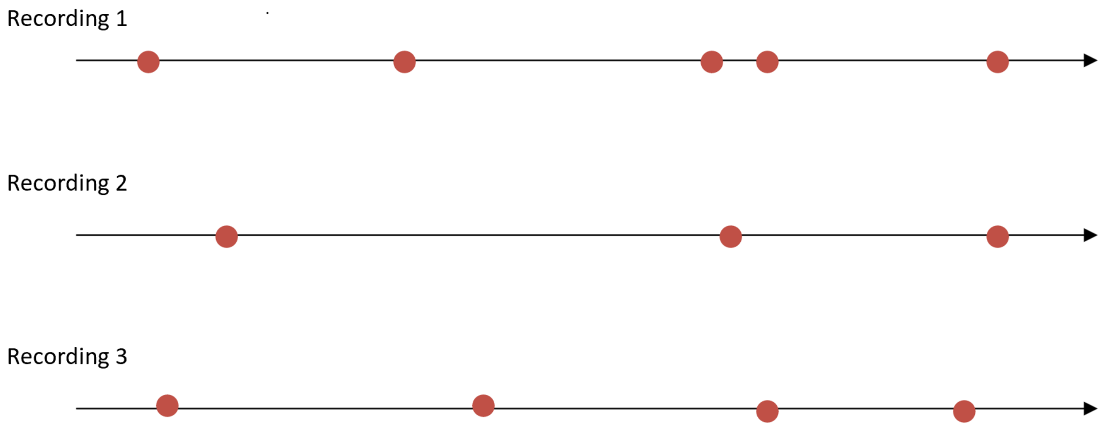
- Signalling
- Timetabling / passenger information / delays
- Infrastructure monitoring
- Maintenance work

All of these have different accuracy requirements, but they all share the same common requirement, at a very minimum the track reported must be correct. When we talk about accuracy, we are usually talking about longitudinal error, the distance up or down the track (rather than a lateral or sideways error) - the railway is in effect 1 dimensional, a train cannot move left/right away from a track unless it reaches S&C. When the wrong track is selected, the position is just plain wrong - we have moved away from the 1-dimensional map, and we can consider the error to be effectively infinite. This may be acceptable sometimes as it may well be that every inch does not need to be reported on - but on the flip side, this then implies that the incorrect position must be known as such - i.e., we must know it is incorrect. This is quite often where the issue can be, how do we deal with an incorrect position (and therefore measurement) if we don't know that it is incorrect?

With traditional fixed block signalling, the problem is solved using track circuits, i.e., we know exactly which track a train is on because it closes an electrical circuit. What we don't know (or need to know) is where along that circuit it is. If we extend this to timetabling and passenger information, we can supplement the track circuit information with an approximate position from either GNSS or some kind of wheel encoder to tell us (within a tolerance) where that train is along the circuit. This information need not be very accurate - these systems consider errors of over a minute, which is potentially hundreds of metres.

More modern approaches to signalling, such as ETCS, consume data from other sensors to provide continuous position. The same kinds of data provide position for DAS and ATO, these trains usually have some form of GNSS alongside a wheel encoder/tachometer, they take external location correction and updates from balises installed on the physical infrastructure. The locations of these balises are usually surveyed-in to high absolute accuracy and are placed frequently along the track. When coupled with an accurate timetable, this gives a location of the train to a very good accuracy. The aim is often to have this accuracy at a metre or so.

When it comes to IM, suddenly, the requirements change. We need to know where a fault is so that we can either a) monitor it or b) fix it. In the first instance this drives the accuracy requirement by saying "given a set of faults along a track, what is the largest distance apart that faults can be for me to categorically state that one is identifiable from a set of faults captured at another time". In the second instance the accuracy requirement is similar, "given a fault, what is the largest error allowable in the measurement that permits a worker to find that fault with a certain toolset (which could also have a locational error)".



Consider the image above, we have 3 recordings - for the first there are five faults, what we need to know is for the subsequent recordings, which faults found are the same as those on the first recording. If the positional error is large - then this could be a difficult task, especially if there are faults grouped closely (tighter than the error in the positioning system) – or even a different number of faults (new faults or faults that have been remedied).

Now, both statements come with the caveat that we still need to categorically know which track the fault was on. This is the first quandary that we are presented with in IM, not only do we need to know where we are along a track, we need to know if the track we think we are on virtually matches the physical track the train has travelled on. This is important for many reasons, not just in terms of fault finding and remedial work, but fault classification due to track category information, line speed and so on.

How has IM been positioned traditionally?

We've come a long way from synchronising mileposts by pushing a button whilst looking out of the window. GNSS and Inertial Navigation has given us the potential to locate down to a few centimetres - but never all the time (in fact, infrequently in the railway). GNSS itself has evolved in the last 20-30 years, this really started when the US removed selective availability from the signals and consumers (i.e., non-military) were able to use dual (and multi) frequency receivers to get good positional accuracy.

So where is the problem?

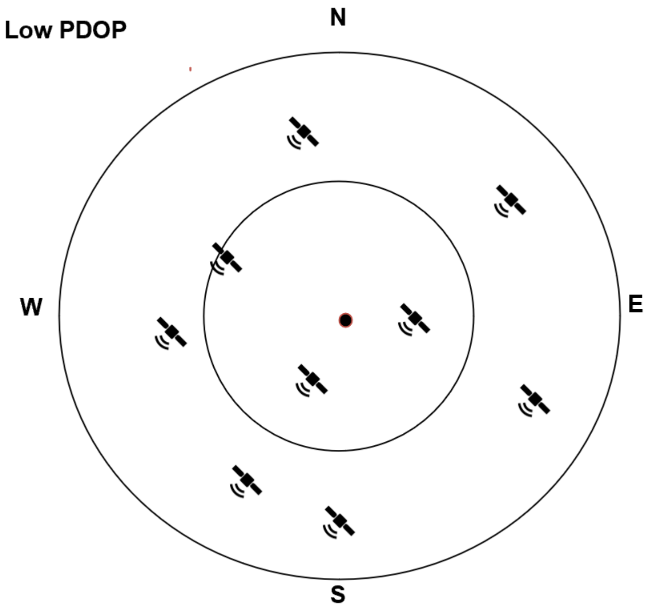
GNSS

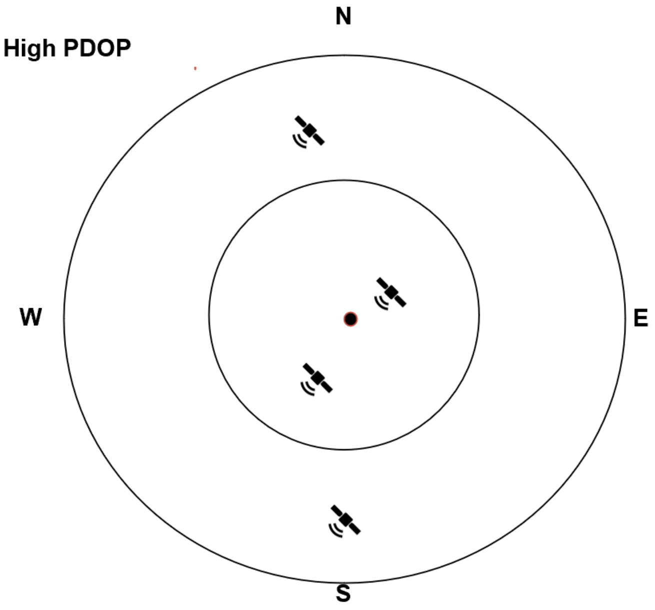
There are now multiple GNSS (Global Navigation Satellite Systems) constellations available for our use. The original one Navstar (commonly known as GPS – Global Positioning System) was developed by the US in the late ‘70s. This is now supplemented by Glonass (Russia), Galileo (Europe), Beidou (China), QZSS (Japan), IRNSS (India). In the UK for instance, we typically get good coverage from GPS, Glonass and more recently as newer receivers support it, Galileo is becoming operational.

GNSS works by having line of sight to a constellation of satellites orbiting in space, to achieve the best accuracy with GNSS, our receiver needs to be able to see a good spread of the satellites (we need some above, to the left, right, ahead and behind us ideally), this can be quantified mathematically by a value known as dilution of precision. The most used form of this is called PDOP (Position dilution of precision), this is a numerical measure of the satellite geometry (the spread of the satellites as the receiver sees them) - the smaller this value the better.

The images below give an idea of what is a good and bad PDOP, imagine the receiver is at the centre of each set of circles. The top image shows a spread of satellites in every direction - this is a good PDOP (low value), the second has satellites ahead and behind the receiver but none to the side - this will give a high PDOP. With a low PDOP there is more information that enables a receiver to calculate position with a lower ambiguity than with a high PDOP.

PDOP is a big topic that we won’t cover here - but the important factor in the railway is that it is very hard to achieve a low PDOP all the time, embankments, tunnels, dense urban areas, OLE and other parts of infrastructure block the signal - meaning that the PDOP usually changes frequently and is often high (more like the second image). With a high PDOP, it is unlikely that we will have a high accuracy position.





PDOP should not be used as the sole source of GNSS quality – it is a guide. Other factors can be at play, for instance, PDOP is a measure of the geometry of the satellites that the receiver is receiving signals from. This doesn't mean that all of these satellites are in the direct line of sight to the receiver. Multipath is of concern in the railway, multipath is when a signal is reflected and/or refracted from objects near the receiver. The trains we use, generally operate in high multipath areas (in effect longitudinal canyons a lot of the time). High quality (particularly multi frequency) receivers and antennas can significantly reduce multipath errors.

The spec sheet of a typical survey grade receiver will state its capabilities, something like an accuracy of 1cm + 1ppm. What does this really mean? This means that in these ideal situations where we have a very low PDOP (great view of the sky in all directions) can we expect to get a position with an accuracy of 1cm? Not quite, this also includes another figure, 1ppm - this is an additional error based on the distance of the receiver to a reference station.

Again, in here, we don't want to go into the technical details of this but to get these accuracies we also need a correction service such as RTK (Real time kinematic) or PPK (Post process kinematic), these provide correction data using ground-based reference stations that are often miles away. 1ppm means one part per million and equates to an additional error which is a function of the distance of our receiver to that (or those) reference stations(s).

There are lots of other considerations in how RTK (Real Time Kinematic) or PPK (Post Process Kinematic) work and how their accuracies are calculated and maintained, for instance, both require tracking of satellites for an amount of time (they are not instantaneous), a change in the status of satellites being tracked may change the quality of the RTK/PPK solution - and therefore the accuracy. PPK is usually more accurate than RTK as there is access to more information, for example, once an RTK fix is obtained, the accuracy obtained applies from that point in increasing time until a change in the tracking status (of satellites and/or the correction), whereas PPK can "re-wind" the fixed correction back in time because it is a post process operation. But of course, neither of these give great results in challenging GNSS areas (urban canyons, tunnels etc) where the signal may be lost all together, this is why these methods are often supplemented with an INS (Inertial Navigation System), these devices are able to bridge the gap between areas of GNSS coverage - giving continuous position where no satellite signals are being received. Depending on the grade of device, the drift (the rate the error increases) away from the truth could be a few centimetres for hundreds of metres - of course, this depends on what the truth is and whether the device is operated in real time or post process.

To decide on which positioning system should be used for IM data, it is first essential to decide what latency is required in the data. For instance, if it's acceptable to have a latency of (say) 24-48 hours after a shift, then a PPK INS solution is most likely to be suitable (given an upper limit as to what is an acceptable positional error of course). However, if results are needed in near real time (this could be within X minutes of measurement or even quickly after the end of the shift), then it is likely that a PPK INS solution will have too high a latency and as such, we will have to look for another solution, most railways now use some form of GNSS/INS either with a correction service (such as RTK or SBAS (Satellite Based Augmentation System) such as Terrastar) - or without, relying on uncorrected positions alone. Both options can and will produce erroneous positions in certain scenarios and without quality control and correction, this could cause issues for IM.

Other sources of error

The errors in GNSS/INS measurement are not the only source of positional error - a set of Lat/Long coordinates in a certain datum is probably no use to anyone, these need either attribution against a network model (to get some kind of engineering geography reference and linear measure) or a separate device to interpret them (such as a handheld GNSS device to direct maintenance workers to their target). If we start with the latter, we send a maintenance worker out with a pair of coordinates alone to fix a fault - will they be able to find it? This ultimately depends on what other information they have, the accuracy of the point we give them and the type of fault:

1. A missing fastener on a piece of single track (i.e., one line only) in a nice open sky environment - the worker will be able to get near the fault using a handheld device and find it by walking up and down the track - the fault will be visible. **Yes**

2. A top defect on a 4-track railway in a nice open sky environment - this is where it becomes tricky, armed with a long/lat (with an error), the worker can use their handheld device to position themselves nearby - with a further error from that device. The worker realistically has no idea which line the fault could be on and so may have to walk up and down each track to spot the defect (assuming they can get access). If the defect is severe enough, this may be visible while standing on the other tracks. **Maybe**

3. A minor gauge fault in a tunnel with two lines - there is no source of GNSS inside the tunnel and low levels of lighting, the worker will find it harder to navigate to the defect, which may be hard to see. Here the worker really needs more information, some kind of linear measure (metres from the start of the tunnel for example). **Unlikely**

4. An ultrasonic defect in an urban canyon area of railway with multiple tracks - because of interference such as multipath and/or possible lack of a GNSS signal at all, the error in the handheld device position will be large. Coupled with the error in the measurement position, realistically the worker needs more information (or a long time on site) to find the defect as it is most likely not visible with the human eye and requires test equipment. The worker would have to test a large area of track to find the defect meaning it would be time prohibitive. **No**

So, if the worker had information about which track the defect was on, then they would have a better chance in the last 3 of these scenarios and in all scenarios some kind of linear measure would be advantageous so that in the event of a poor position from the handheld device, they can orient themselves locally (e.g. from the start of tunnel or a set of points etc.). There is of course still a chance that they will fail to find the defect within the time allotted. So how do we know which track the defect was recorded on?

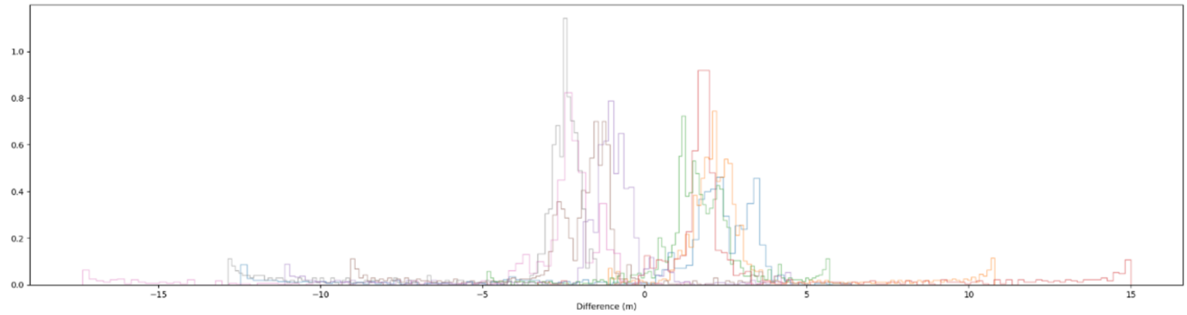
Map matching

Map Matching is the process of calculating the path traversed by a train through a network model, this could be done by finding the nearest line to the long/lat - but we already know that we have a potential positional error in the measurement point, the nearest track might not be the right one. We could also have not just a positional error in the network model (i.e., track geography incorrect or an offset), but we could also have topological errors (i.e., tracks ordered incorrectly, missing tracks, additional tracks), as such, we must be careful as to how that attribution is done. When doing this with IM data, sophisticated algorithms are employed, these don't just return the closest line, but they return the best fit path through a section in time - which is a much better measure. But of course, this could still be wrong, how do we tell? The answer is - we don't. The maintenance worker (potentially) goes out and can't find the fault, it isn't fixed - worse still a speed restriction could be applied to the wrong line (and NOT applied to the right one).

The point is that whatever happens with IM data currently - some of the results are just plain wrong. This leads to an additional question in choosing the positioning system - what percentage of time is it acceptable for the position to be incorrect? This is the fundamental part that is missing in standards and spec sheets today. Yes, system X can give you 1cm + 1PPM in clear open sky, yes, the drift of the data in the absence of GNSS can be 10cm over 1km (given certain criteria and caveats – not least the speed of the train) – but what percentage of the data points produced by the overall system have the incorrect attribution? Currently, the latter is a number that is relatively large – we are probably talking somewhere between 1 and 10 % for the better systems. This may seem like a small number, but for an IM shift covering 200 miles, this means between 2 and 20 miles of the track has an ambiguous (or worse – incorrect) status.

The answer to the question - what percentage of time is it acceptable for the position to be incorrect, is quite easy to write. The answer is – given sufficient measurement runs within the maximum allotted time frame for recording, provided that at least one of these measurement runs is correct – then every piece of track must have a valid measurement (and therefore position). Basically, the percentage of time answer is zero – or we must introduce counter measures (such as speed restrictions etc.).

So, what is the mitigation for this? Record it again, and again, and again and hope that it is right at some point? So how do you detect when the position is wrong? This is done by user validation, so not only are multiple redundant shifts being run, but users trawl through positional (and measurement) data to validate it. This is of course starting to sound like we are no longer in the age of AI!



As we have said, the errors in GNSS depend upon a number of factors - not least satellite visibility, they also depend on environmental factors (ionospheric events for example) as well as the type of equipment used (and the maintenance of that equipment) and can even depend on the time of day. All of this leads to a realisation that the worst part of the error with a GNSS based system is that the errors are not predictable. Seen above is a typical histogram of the error spread of a typical GNSS receiver, as we can see the error varies and is not well distributed.

Other sources of position

So, what can be done to correct/supplement the position to ensure correctness? There are a number of existing methods out there. The first one we have discussed is a correction service to try and remove GNSS errors (RTK, PPK, Terrastar) - there are constraints with these, availability of correction service, time constraints and satellite availability. There is the option to use physical infrastructure such as balises, RFID tags - these rely on physical installation (and survey), maintenance and are not available throughout the network. Care must also be given to ensuring that the events generated by these are consumed at the correct level by dependent systems (so not to introduce a time-based error). These also must be “mapped” to the network model being used (this not only involves traditional survey, but a potential shift and attribution to the network model so that these “speak” the same language as the positioning and measurement systems).

The signalling system itself could also be used, as a train travels through the network it passes through signal berths and this information is made available via a service. There are issues with this:

- The mapping between berths and the network model may not be accurate
- The time and location resolution of this is not precise (i.e., we get an approximate time a train passed a location and the location could also be approximate)
- It is not trivial to work out which train is which and we often need a lag of time to calculate this
- The coverage is not necessarily granular enough to be used everywhere

Other on train sensors can be used, Lidar has been used to work out which direction a train travels through a switch for instance. This can work well but does not always work through complicated areas of infrastructure and requires the network model to always match reality. It also requires the position (and track attribution) before the switch to be correct.

Lineside fibre optic cable has also been used to position trains; this obviously requires the installation of the fibre but also requires calibration. For IM, the drawbacks around timing and synchronisation would prove an issue.

It is worth revisiting the problem, positioning of IM data is more than a pair of Long/Lat coordinates, it is also a problem of time. If we imagine a system that is capable of synchronisation to another with an error of 10 milliseconds, this at face value sounds great. However, if we consider a train running at 125mph - this equates to an error of nearly 56cm. Add this to our positional error and it’s clear this is not ideal. With this in mind, how systems are synchronised is paramount as we can quickly introduce huge positional errors if it is not done correctly. It’s very easy to get this wrong and consideration of, not only computer clocks, but also stamping of data correctly from input sensors, must be paramount.

Ideally, we would like to keep boots off ballast and avoid installation of physical assets, we would also like a solution that works everywhere with maintenance of, at most, one system (keeping multiple systems in line is difficult - especially when dealing with data for IM which requires up to date information at all times).

A note on relative and absolute accuracy

What does having accurate position actually mean – there are two metrics to quantify?

Absolute accuracy - i.e., where the position is in absolute space. A good question would be - is this important? Absolute accuracy is only ever really important if we want to measure the same thing (or locate the thing) with two different methods - i.e., absolute accuracy gives us the mapping between the two, if we go to the same place with different systems/devices or at different times, do we get the same result.

Relative accuracy is often interchanged with repeatability, if we measure something with a device, can we measure it again with the same device and get the same answer. Consider how Track Geometry (TG) is currently validated (and therefore often positioned). A reference run is recorded, subsequent runs are then recorded and “aligned” to the initial run. This works well if the track geometry doesn’t change much between runs and the measurement system has no (or few) erroneous measurements. However, there are assumptions made here, namely that the base run is correct - both in terms of measurement AND position, so that the position of subsequent runs will be the same as the base run. If the geometry changes (either through physical track change or erroneous measurement) then this will introduce an error into the alignment/position of the runs. The biggest drawback of this approach is its lack of relatability back to an absolute truth, we potentially have no (easy) way of directing a maintenance worker to the exact location of a defect and/or correlating other measurements (e.g., an OLE measurement from a contact-based system) to the same data as there is no absolute positional reference - that we can quantify.

RailLoc

Traditionally other sources of position have been used to supplement GNSS (including using Inertial Measurement units for navigation), if we had a more consistent version of the truth, we could flip this on its head and use this as the source and supplement with GNSS (and inertial) where this information is not available.

As it turns out, information is present and available all over the network which enables us to do this. By treating the problem as one of computer vision rather than positioning, we can take a completely new approach to the problem. By looking down at the 4-foot, it is possible to uniquely position ourselves by using the random structure of the infrastructure. This could be ballast structure, orientation of bolts, cracks and marks on sleepers - or all of these, basically anything visible. Obviously, these things change over time - but not as frequently as you might think, and so given sufficient running, it's possible to keep these up to date. For additional redundancy, fallback to traditional GNSS/Inertial gives a truly robust positional system - everywhere, even in GNSS denied areas.

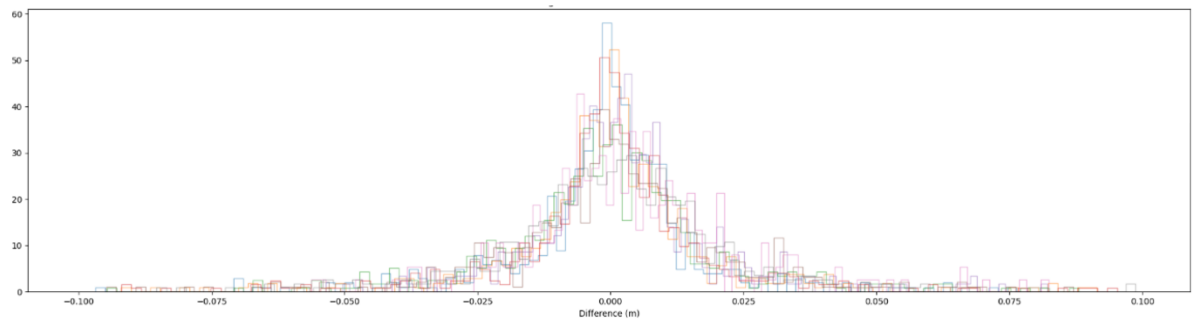
What are the advantages of this approach?

- Works everywhere - including GNSS denied areas like tunnels
- Requires no boots on ballast, no installation or maintenance of the physical infrastructure
- Works in near real time (we don't have to wait for the end of shift to start processing)
- Small footprint on the IM vehicle - simple installation
- Can be transformed to any datum or representation, e.g., ETRS89 and/or linear references
- Very high repeatability (relative accuracy)
- Can inherit the absolute accuracy of an initial survey
- Gives unambiguous position - this approach can definitively state that this is the correct track

There are potential disadvantages of this approach - but these can be mitigated relatively easily.

- Infrastructure can change - this means that we may not be able to determine the track when it changes. This can be mitigated through the addition of a fallback to GNSS/Inertial, multiple systems running over the same piece of track to detect change as it gradually happens and therefore maintain the model
- Requires installation on the train body - this is a small and simple installation and any vehicle being used for IM will require other systems and sensors to also be installed
- Requires additional mapping - a recording to map the visual features to the network model is required. This can either be a bespoke run or could piggyback from a survey. For an initial IM run on a new track, the data can be positioned with a lag allowing for the correct attribution of the data. As such, this need not be a drawback, it can be done via existing processes and surveys

One of the additional benefits of this approach is how the errors manifest themselves, in effect, there are thousands of points of interest everywhere in the 4-foot, the more of these that we can "match" to, the lower the error. We don't need to see all of the features every time - a few features can give us a position. When we don't see any matching features, the gap can be bridged by using an IMU - or even GNSS if the gap is big enough to warrant. The errors also behave in a much more normal way, when we match features, the error is low and generally the more features the lower the error. This gives us something that we can't definitively have with GNSS/INS – assurance of position, as the track effectively emits a unique signature, it is possible to state categorically that the vehicle being positioned is on track X.



Fault Navigator

This potentially still leaves an issue, we may have both an absolute reference for our IM data and also be able to understand the error bound in position between runs, but our maintenance worker still needs to be able to navigate to the faults. Let's consider the four scenarios we covered earlier but this time using RailLoc as the positioning source:

1. A missing fastener on a piece of single track (i.e., one line only) in a nice open sky environment - the worker will be able to get near the fault using a handheld device and find it by walking up and down the track - the fault will be visible. **Yes**

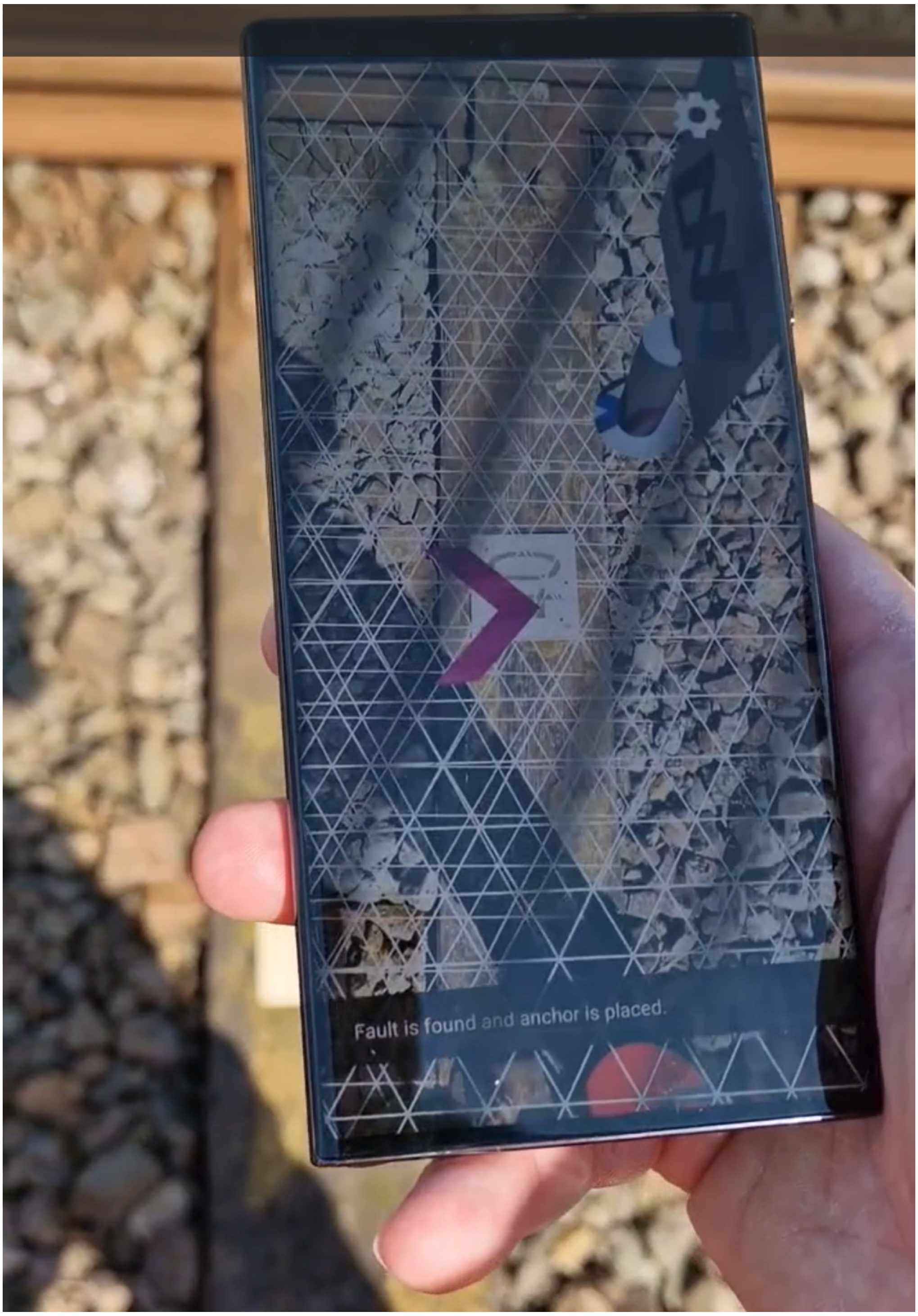
2. A top defect on a 4-track railway in a nice open sky environment - armed with a long/lat (with an error), the worker can use their handheld device to position themselves nearby - with an error. However, with RailLoc, the worker now knows for definite which line the fault was recorded on and so can derive the correct line and look for the defect (assuming it's visible). **Yes**

3. A minor gauge fault in a tunnel with two lines - with low levels of lighting and no source of position, the worker will find it harder to navigate to the defect, which, in itself, may be hard to see. Here the worker really needs more information, the worker knows which line the fault is on from the RailLoc data and using other sources (a distance from the tunnel portal could be derived from the RailLoc position for example) can estimate a linear distance to find the fault. **Potentially**

4. An ultrasonic defect in an urban canyon area of railway with multiple tracks - because of interference such as multipath and/or possible lack of a GNSS signal, the error in the handheld device position will be large. The worker now knows which track they are on but realistically must test a large section of track to find the defect. They may be able to orient themselves with additional context around the RailLoc position (e.g., sets of points, other physical infrastructure) This may or may not be possible depending on the initial localisation that they can have (i.e., does their handheld device put them within range of the defect) and the time available to them on site. **Unsure**

An absolute reference can mean different things to different people, it is really a feature or monument that is distinguishable from others by using some kind of measurement device (or by eye!). So typically, we think of an absolute measurement as a long/lat pair referenced to some datum (usually ETRS89 in the UK), this is easily transformable to a local grid reference. We need some other device to interpret this - usually some kind of GNSS receiver (with the errors having been discussed). A point can though, have multiple absolute references, as above an ETRS89 reference as well as a reference to the OS national grid; this could also be an offset along a bearing to a monument (such as a trig pillar). In the case of data captured alongside RailLoc, a further absolute reference is available, i.e., that of the underlying features that were captured by the system. As it turns out, we can make use of these features using a handheld device to help navigate a track worker to the same location.

This works in the same way as RailLoc, we use the device’s (e.g., mobile phone) camera to match features on the track to those stored in the map, we can then use this to localise ourselves in the feature reference space which in turn has a mapping back to absolute (ETRS89/OS) space.



Fault is found and anchor is placed.

So, using this app (Fault Navigator) the answer to our four problem statements now becomes **YES** for each one! In fact, if we think about the hardest problem of the four, the ultrasonic measurement, then rather than test sections within a large range, the maintenance worker only need test within the range of relative error:

$$R_e = T_e + F_e$$

Where R_e is the relative error, T_e is the timing error between RailLoc and the Ultrasonic recording system and F_e is the feature matching error (from RailLoc and Fault Navigator combined). In practice, this number will be relatively small (a few cm), meaning that the worker can test a very short section to find the suspect/defect which obviously allows much more focussed testing over a much shorter time. The likelihood of them rejecting a suspect or confirming a defect is greatly increased with more accurate positioning. By reducing the area required for manual testing (or walkout), a worker is more likely to increase the amount of test passes (with ultrasonic testing equipment), this ultimately means a more definitive conclusion will be drawn from the testing - whether rejection of a suspect or validation of a defect.

With RailLoc, rather than the initial position being arbitrary or uncontrolled, we are able to survey it. This means that we can either piggyback on a Lidar (or similar) survey and wait for the fully processed positional output (this could take some time) or we use an inertial navigation system with PPK and match the position to a network model (either provided or this could be made as part of the process) which could have a relatively quick turnaround (within a day or two following the shift). This allows us to ensure (through a level of QC) that the path chosen through the model is correct and correctly attributed, but also understand (through statistics generated) the absolute accuracy of the resulting data. The benefit is that this is a one-time operation, any subsequent runs match to this one and inherit the absolute accuracy. For a typical survey of this kind, we can expect the absolute accuracy of the data to be a few centimetres - with outliers in very challenging GNSS denied environments. The important fact is that subsequent runs are placed in the same place every time - irrespective of any errors in the measurement (TG, OLE etc.) system. In fact, this approach makes it possible to place different datasets from different trains in the same position, time and time again, without the ambiguity of using an arbitrary reference run.

Additional possibilities with RailLoc

Change detection

As the infrastructure changes, the features used in the RailLoc map do too. This could involve a feature gradually moving (as bits of ballast shift and settle) or a feature no longer being present. Both can lead to additional benefits:

- Tamping - the change after tamping allows RailLoc to provide information about the extent that was tamped (exactly) and where it was NOT tamped
- Changes due to renewals - where a section of track has been renewed, this can be picked up straight away using RailLoc and fed back into the Network Model as well as being used to invalidate previous geometry recordings (for example - as the geometry has changed, previous defects may no longer be valid)
- Monitoring of ballast - for gradual changes of ballast, either through reduction in features or perhaps movement of features, RailLoc may be able to provide insight where these phenomena are unwanted (e.g., voiding)

Input into signalling systems

As mentioned before, systems such as DAS/ATO/ETCS often consume data from balises installed on the physical infrastructure. This is a potentially costly exercise requiring time on site (boots on ballast) as well as planning, surveying and maintenance. These are safety critical systems relying on the availability of all the sensor inputs, additional redundancy can be gained through consuming data from RailLoc - this could be continuous or on a low frequency basis (acting as a virtual balise for instance). Given sufficient frequency of operation, RailLoc could be used to increase the spacing of balises (i.e., reduce the number required and therefore reduce the installation and maintenance costs). This would be done through fixed updates (i.e., virtual balise) but also through higher quality continuous data (supplementing or replacing tacho and GNSS) into these systems.

Summary & conclusions

In this article we have discussed the typical approaches to railway positioning, and particularly those used for positioning Infrastructure Monitoring data.

The devices and systems we have discussed have stood us in good stead, but there is more we can do. By embracing newer technologies, we can replace manual processes, retain data and be more efficient. Ever increasing demand on the railway means that the need for better data and automation has never been higher, increasing capacity and frequency of services means that we need to keep boots off ballast where possible. The safety implications of workers being on a live railway mean that we should only send them to site where it is absolutely essential.

We have discussed inertial navigation systems (combinations of GNSS and IMU) which are now commonplace, these can give excellent results, but not everywhere in the railway. The railway does not lend itself to perfect GNSS positioning because of the large number of occlusions of the signals (embankments, OLE, tunnels, station canopies etc.), this means that we either accept a degraded or incomplete dataset or we look to supplement or replace this approach.

GNSS based systems can produce higher quality data through additional post processing, but this causes delays to the solution that may not be acceptable for IM data. This approach still doesn’t give us the answer everywhere. We have also discussed further sources of error such as the underlying network model (both topology and geometry) which can lead to attributional errors.

Additional sources of positional correction are available, but these are likely to either require installation to physical infrastructure (as well as maintenance) or offer a lower (spatial and temporal) resolution. In fact, RailLoc can be used to supplement or even replace these.

The advent of computer vision has unlocked more innovative approaches to positioning, such as RailLoc. Rather than supplement GNSS based systems with external information, RailLoc provides the base truth and uses GNSS/INS for infill. Unlike the other systems discussed, RailLoc works everywhere on the railway and requires no installation to physical infrastructure. RailLoc can not only give excellent relative positional accuracy but can provide high absolute accuracy through initial survey. RailLoc is also able to provide certainty that a defect (or full path) was captured on a specific track with categorical assurance.

Away from positioning data on vehicles, we have also discussed the issue of track worker navigation to defects and suspects captured by these systems. Standard handheld GNSS devices can work well, but struggle in challenging GNSS areas and their errors can be large (or worse misunderstood). Building on the RailLoc functionality, Fault Navigator allows navigation of a maintenance worker to the precise location of a fault, significantly reducing time on site and repeat visits.

Terms & Acronyms

Acronym	**Description**
ATO	Automatic Train Operation
CM	Centimetre
DAS	Driver Advisory System
ETCS	European Train Control System
ETRS89	European Terrestrial Reference System 1989
GNSS	Global Navigation Satellite System
IM	Infrastructure Monitoring
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
KM	Kilometres
MPH	Miles Per Hour
MWV	Machines With Vision
NR	Network Rail
OLE	Overhead Line Equipment
OS	Ordnance Survey
PDOP	Position Dilution of Precision
PPK	Post Process Kinematic
PPM	Parts Per Million
QC	Quality Control
RFID	Radio Frequency Identification
RTK	Real Time Kinematic
SBAS	Satellite Based Augmentation System
TG	Track Geometry