Building a Digital Twin: Testing the effectiveness of telecommunication policies in a virtual world

Edward J Oughton

University of Oxford

Acknowledgements

The research detailed in this paper was funded by the UK Engineering and Physical Science Research Council under grant EP/N017064/1: Multi-scale InfraSTRucture systems AnaLytics.

Abstract

Infrastructure decision-making is challenging due to high levels of future uncertainty. Indeed, when considering broadband coverage, we are also faced with a situation where operators are reluctant to share data, there are few existing open-source evaluation models, and most available tools cannot be used by non-technical users. Transparency is sometimes low. Often disagreements arise between operators, regulators and other actors, with little independent assessment of key issues. Consequently, this paper proposes the development of a Digital Twin as a virtual test-bed for evaluating telecommunication policies. The concept of a Digital Twin has been common for several years in aerospace engineering, since first proposed by NASA in 2010. The vision defined in this paper is for a Digital Twin to be a virtual engineering-economic representation of a real-world telecommunication network, whereby simulation techniques allow exploration of potential future states under different policy conditions. A Digital Twin is developed for the British incumbent's fixed broadband network, spanning 30 million premises and over 4.3 million geospatial telecommunication network assets. Using a network subset for Cambridgeshire, England the rollout of Fibre-To-The-Premises and Fibre-To-The-Distribution-Point upgrade options are then tested. Under different demand scenarios, market-based rollout is compared to a subsidised rollout strategy. Independently testing broadband deployment strategies in a virtual market and evaluating their effectiveness can provide greater transparency for decision-making processes. Over the long-term, a Digital Twin could help to generate new knowledge by testing experimental policy options, fostering greater innovation in how we tackle both perennial and emerging digital divide issues.

## 1 Introduction

The concept of a Digital Twin has been common for several years in aerospace engineering, since first proposed by NASA in 2010. The original intention was to create a multi-scale probabilistic simulation for a vehicle or system to mirror the life of its flying twin. However, over the past decade the concept has evolved, with one application being the development of a Digital Twin for national infrastructure systems, covering all energy, transport, digital communications, water and waste assets.

The UK's National Infrastructure Commission (2017) has pushed forward the idea of building a Digital Twin to act as a computer model of Britain's infrastructure to help plan, predict and generally improve our understanding of these systems. Indeed, it suggests that the Digital Twin concept be tested as soon as possible by setting out the challenges and security issues that need to be overcome to deliver on this vision. Firstly, this includes the impetus to obtain the necessary data required to explore the use of machine learning approaches to solving problems traditionally addressed using more simplistic forms of analysis, and secondly, demonstrate how a Digital Twin may be used for better decision making in the management of national infrastructure.

Telecommunications policy research is strongly dominated by *ex post* methodological approaches. While highly valuable, this is to the detriment of *ex ante* methodological approaches which serve as an equally valuable form of scientific philosophical enquiry, aiding in the testing of 'what if' policy scenarios. Ultimately, telecommunications policy is a classic decision-making under uncertainty problem which is amenable to scenario analysis and uncertainty quantification. Thus, in this paper the vision of a Digital Twin aims to create a virtual engineering-economic representation of a real-world telecommunications network, whereby simulation techniques allow exploration of potential future states under different policy conditions. Indeed, this approach could have significant ramifications for our *ex ante* understanding of future policy scenarios and how they may affect broadband availability, adoption and inclusion in the digital economy. Research has now begun to build a Digital Twin utilising the high-performance computing environment provided by the UK's Data and Analytics Facility for National Infrastructure (DAFNI).

The UK's first ever National Infrastructure Assessment examines future economic infrastructure needs up to 2050 for all infrastructure sectors. For digital networks, the assessment makes ambitious recommendations which include ensuring that full fibre connectivity is available to 15 million homes and businesses by 2025, 25 million by 2030 with full coverage by 2033 (National Infrastructure Commission, 2018). These numbers are important benchmarks. Currently there is disagreement between the Chancellor of the Exchequer, Philip Hammond who proposed the targets, and Openreach (the local access infrastructure owner, recently fully separated from British Telecom). Whereas

Openreach has committed to rolling out 3 million FTTP by 2020 and 10 million full FTTP lines by 2025 (Bicheno, 2018), the Chancellor has been pushing for the 15 million premises stated in the UK's National Infrastructure Plan (Reuters, 2018; Think Broadband, 2018). Hitherto, little modelling has been undertaken to explore the spatio-temporal rollout of Fibre-To-The-Premises (FTTP) across the UK, or how these numbers stack up against other technologies such as Fibre-To-The-Distribution-Point (FTTdp). We consequently put forward the following research questions:

1. How may the Digital Twin concept enhance our existing approach to testing policy options?
2. What characteristics does a Digital Twin of a broadband network require?
3. How do different policy options perform in influencing the rollout of FTTdp or FTTP?

In the following section a literature review is undertaken, followed by the paper method in Section 3. Results are presented in Section 4 and discussed in Section 5. Finally, conclusions are put forward in Section 6.

## 2 Literature Review

This review begins by surveying the existing literature to define the functionality of a Digital Twin, in the hope of helping to answer both research questions 1 and 2. This is undertaken from a critical perspective with the aim of evaluating whether the concept of a 'Digital Twin' is valuable, and has the potential to add both theoretically and empirically to the research area.

### 2.1 How do we define a Digital Twin and what properties should one have?

The concept was first mooted in the field of aerospace engineering, featuring in the US National Aeronautics and Space Administration's (2010) Modeling, Simulation, Information Technology & Processing Roadmap stating that a Digital Twin is:

> '*an integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin*.'

Since this time, the idea has developed further within aerospace to military weapons development, as well as in manufacturing ('Industry 4.0') and other related civilian societal activities. The concept is not seen as an easy development to implement due to the sheer number of lines of code required to deliver such an ambitious vision. For example, West and Blackburn (2017) put the concept into perspective by stating that the US Department of Defence's Digital Twin development programme is comparable in effort to the Manhattan Project and could take up to 100 million lines of code. Therefore, even with a potential cost estimate of $1-2 trillion and a team one-third the size of Microsoft, such an effort could still take 100-250 years to complete.

Away from the defence sector, the Digital Twin idea has been conceptually developed in the literature, even if the relatively few papers written on this topic are embryonic. One of the most comprehensive overviews of the concept is by Negri, Fumagalli, and Macchi (2017) who essentially review the roles of a Digital Twin in cyber-physical production systems. Importantly the authors evaluate the different definitions of the Digital Twin concept, concluding that essentially this idea is about providing *virtual representations of systems* along their lifecycle. Table 1 outlines a summary of the literature on this topic, stating the Digital Twin definition utilised, the aim of each paper, and any key findings or caveats.

Table 1 Survey on the literature associate with the Digital Twin concept

| Author(s) | Year | Sector | Purpose | Digital Twin definition | Paper aim | Key findings or caveats |
|---|---|---|---|---|---|---|
| Tuegel et al. | 2011 | Aerospace engineering | Structural life predictions | A Digital Twin is an ultrahigh fidelity model, which in the case of aircraft, includes all manufacturing anomalies and accepts probabilistic input of loads, environmental and usage factors. | Proposes a conceptual model for each individual aircraft to integrate computation of structural deflections and temperatures in response to flight conditions, with resulting local damage and material state evolution. | Currently a wide variety of models are used for each aircraft but are not integrated, hence the advantage of a Digital Twin is through model integration, leading to improved decisions. |
| Uhlemann, Lehmann, and Steinhilper | 2017 | Manufacturing ('Industry 4.0') | Design and production engineering | A Digital Twin is a digital shadow of a cyber-physical production system allowing centralised analysis and control of production processes. | Presents a practically feasible approach to multi-model data acquisition with the aim being to couple a production system with its digital equivalent. | Data acquisition and evaluation mostly rely on existing isolated solutions. The advantage of a Digital Twin approach is in overcoming this isolation, towards development of a comprehensive virtual production system. |
| Schleich et al. | 2017 | Manufacturing | Design and production engineering | Quoting Boschert and Rosen (2016), 'the vision of the digital twin itself refers to a comprehensive physical and functional description of a component, product or system, which includes more or less all information which could be useful in all – the current and subsequent – lifecycle phases.' | Proposes a comprehensive reference model which serves as a digital twin of the physical product in design and production engineering. | Important model properties include scalability, interoperability, expansibility, and fidelity, as well as different operations on this reference model along the product life-cycle, such as composition, decomposition, conversion and evaluation. |
| Negri, Fumagalli, and Macchi | 2017 | Manufacturing ('Industry 4.0') | Design and production engineering | Outlines 16 Digital Twin definitions but concludes that for manufacturing the value lies in 'the virtual counterpart of physical devices'. | Attempts to analyse the definition of the Digital Twin concept in the scientific literature, retracing this definition from the initial conceptualisation in the aerospace field, to the most recent interpretations in the manufacturing domain (specifically Industry 4.0 and Smart Manufacturing). | Sees a Digital Twin as a representation based on semantic data models that allow simulations to be run in different disciplines, supporting not only prognosis at design stage, but a continuous update of the virtual representation based on real time synchronisation with sensed data. |
| Ricci, Tummolini, and Castelfranchi | 2017 | Mirror worlds, AI and society | Smart environments | Utilises Gelernter's (1993) 'Mirror Worlds' idea where software models can form a chunk of reality, based on massive amounts of information. | Suggests that Mirror Worlds can be used as a conceptual blueprint to envision future smart environments in which the physical and the virtual are blended | While Mirror Worlds can allow temporal, individual, and social augmentations, they raise issues pertaining to privacy and security. |
| Haag and Anderl | 2018 | Manufacturing | Design and production engineering | Digital Twin is a comprehensive representation of an individual product that will play an integral role in a fully digitised product life cycle | Develops a Digital Twin to enable a cyber-physical environment for a piece of manufacturing equipment. | The created Digital Twin is for use in initial product development and has not been designed to show the actual state of the product all through its entire operational life time. |

Although Table 1 outlines many definitions, we focus on discussing the most relevant. For example, the definition put forward in Boschert and Rosen (2016) is that a Digital Twin 'refers to a comprehensive physical and functional description of a component, product or system, which includes more or less all information which could be useful in all current and subsequent lifecycle phases'. Hence this definition is indicative of how this approach could be used for optimising operations or predicting component failure, with the aim of enabling users (such as engineers, designers, operators and maintenance personnel) *to improve complex decision-making under uncertainty*. Herein lies the relevance to telecommunications policy, and why there may be promise in this approach for improving decision making across industry and government.

A similar alternative concept of 'Mirror Worlds' has been around for many years, as put forward by Gelernter (1993), where a software environment could represent a true-to-life mirror of reality, within a computer, which could be probed by citizens living in the real world. This idea has received further attention with the proliferation of interest in Artificial Intelligence and Machine Learning methods. For example, Ricci, Tummolini, and Castelfranchi (2017) examine how wearable technologies and augmented reality could enable the blending together of both virtual and physical to develop smart environments which utilise augmented or virtual reality.

Ultimately, the miniaturisation and price decline of Information Communication Technology (ICT) allows the integration of sensors, processing power, and wireless communication capabilities into virtually any product (Haag and Anderl, 2018). This is leading to a revolution in the level of data generated from human movement, traffic flows, the built environment, and industrial machinery and domestic appliances. Hence, we are now capable of collecting the data required to develop a Digital Twin, which simply was not possible a decade ago. While this type of approach is pervasive among the relatively limited research in this area, there has been a lack of focus on how this concept could be used either within telecommunications, or also the policy domain, thus motivating the purpose of this paper.

One final vision of a Digital Twin is the formation of a centralised infrastructure asset repository where data are available for selected individuals to access in a fully secure environment. Often approaches to infrastructure decision-making are piecemeal, fragmented and rely on a wide range of data sources, models, and heterogeneous assumption sets. This consequently introduces a large amount of uncertainty into the analysis undertaken to inform decisions, particularly for policy. Hence, one advantage of this central repository is that it allows access for selected analysts to the key datasets on assets, removing many barriers to research and analysis.

## 2.2 Policies options for enhancing broadband coverage

In this review we begin by reviewing supply-side policies as these are the most commonly used to address market imperfections. Supply-side factors include the use of regulation to encourage market competition, as well as subsidies, regulatory holidays and tax breaks. Much analysis has been undertaken to date, focusing on quantifying investment and adoption (penetration) under certain supply-side conditions. Falch and Henten (2017) outline three different dimensions which cut across (generally supply-side) policies pertaining to broadband services including (i) infrastructure vs. service-based competition, (ii) regulatory vs. developmental policies, and (iii) networks vs. content prioritisation. The UK has generally focused on service-based competition, a regulatory approach and an emphasis on networks, in contrast to the USA which has focused on infrastructure-based competition, a regulatory approach, and simultaneous emphasis on both networks and content. The rollout of new telecommunication technologies often rely on a mixture of 'supply push and demand pull', where market operators incrementally deliver new technologies, while business and consumer demand concurrently grow, creating favourable market conditions for further investment (Choudrie and Papazafeiropoulou, 2007).

A key area of telecommunications policy research is the effect of different regulatory regimes on broadband take-up and coverage. For example Götz (2013) uses a simulation model similar to Valletti, Hoernig, and Barros (2002), enabling various regulatory policies to be compared using coverage and penetration as key assessment metrics. The welfare performance of both supply-side and demand-side subsidies were assessed, and it was found that supply-side subsidies were an effective instrument to increase coverage and to supply less populated regions with access (whereas demand-side subsidies appear to be an expensive instrument to essentially increase penetration). In general, strong and well-informed regulatory approaches did not perform much better than the unregulated benchmark. Indeed, light-touch regulatory approaches such as uniform pricing may therefore be more appropriate given that regulators can lack desired information to thoroughly inform regulatory strategies. Due to the degree of rapid innovation in connecting technologies on the demand-side, adding uncertainty to future system states, the analysis found a lack of evidence to suggest that strong regulatory approaches were superior.

Static and dynamic model specifications are utilised by Briglauer (2014) to examine the key determinants of the adoption of fibre-based broadband services, using panel data from the European Union member states between 2004 and 2012. A key finding is that the degree of regulation placed on wholesale broadband access regulation for the incumbent's first-generation Digital Subscriber Line

(DSL) infrastructure is negatively correlated with Next Generation Access (NGA) adoption. Mobile networks also affect NGA adoption in a non-linear way. The paper concludes by suggesting that the DAE targets can be best achieved by focusing on supply-side rather than demand-side policies. This includes deregulation and encouraging favourable competitive market conditions. This judgement is informed based on previous literature, the results indicated in the paper, and historical evidence. Briglauer suggests that the supply and demand dynamic for encouraging NGA delivery is the classic 'chicken and the egg' problem. Indeed, what comes first, the infrastructure or the end-use applications and services? Historically we have seen that the market has developed and delivered end-use applications after new infrastructure has been delivered, therefore supporting this conclusion.

There has generally been less focus on demand-side policies to stimulate take-up of broadband services Kongaut and Bohlin (2015). Demand-side policies include user subsidies (broadband tax deductibles or vouchers) and educational activities to increase digital literacy and education. Uncertainty in future demand is a key constraint which has been found to affect future investment in digital communications infrastructure, spanning both fixed and wireless access (Oughton et al. 2018).

Demand-side policies have been comparatively evaluated by Kongaut and Bohlin (2015) for both Korea and Sweden, with the aim of providing insight for achieving the EU Digital Agenda. The authors conclude from this analysis that both supply and demand-side policies are important to achieve the EU Digital Agenda 2020 and that demand-side policies should be complementary, not substitutionary, for supply-side policies, favouring supply-push, demand-pull dynamics. Additionally, the demand side can be a considerable driver of adoption given the vast development in new content, applications and services often requiring high-speed broadband.

Local factors can have a very large influence over the demand for fixed access broadband, and due to the great deal of heterogeneity associated with the socioeconomic, geographic and technical characteristics of different places, the quality of broadband availability for premises has often been referred to as a 'lottery'. The expected demand for fixed broadband services influences the investment decisions of network operators, leading to disparities in market outcomes. Haucap, Heimeshoff, and Lange (2016) undertook analysis of over 1000 fixed-line broadband tariffs in a cross-country sample of 82 OECD and non-OECD states to understand how diversity can affect fixed broadband penetration. The authors report finding that price-related and socio-economic factors affect broadband demand the most. Income is the key component here, because ultimately there is an economic requirement for customers to be able to pay for desired services. This is similar to previous work undertaken by Jakopin and Klein (2011) which found that economic prosperity (measured as purchasing power parity adjusted GDP per capita across countries), and computer penetration had the largest influence on broadband take-up. Moreover, Lin and Wu (2013) identify the determinants of broadband adoption by diffusion stage in OECD countries using an Arellano-Bond GMM dynamic panel method to capture the dynamics of broadband diffusion while also solving the endogeneity problem associated with estimates in previous studies. Income, education and available content most influence early adoption. The adoption by the 'early majority' is most associated with platform competition and the previous broadband penetration. Finally, adoption by the 'late majority' and 'laggards' is associated with broadband price effects. Having completed a thorough literature review, the method will now be articulated.

## 3 Method

In developing a Digital Twin, the method consists of multiple parts as explained in the following sections. The content of the paper provides a detailed description of the national open-source fixed fibre model developed for Britain (England, Scotland and Wales) which forms a key element of the Cambridge Communications Assessment Model. The code is available for access at www.github.com/nismod/digital_comms. The model utilises very granular spatial information, including street-level data for 30 million premises which includes every building in Britain. Over 4.3 million telecommunication network assets are included in the national model.

The necessary computational requirements are therefore significant. The modelling approach is developed on a computer cluster comprised of 34 nodes, with each system node being either 8, 10 or 12 cores, typically with 128GB of RAM. Network pre-processing using all 34 nodes, takes approximately 10 hours, generating over 30GB of intermediate data for 5,383 exchange areas across Britain. A network subset for Cambridgeshire, England is utilised in this analysis for computational efficiency.

### 3.1 Synthetic network structure

In the absence of the actual network structure (due to commercial sensitivities), an estimated digital network must first be generated which broadly shares the same statistical properties as the actual network. Table 2 illustrates the statistical properties of the incumbent's network previously belonging to British Telecom (BT). A recent regulatory development means the access component is now owned by Openreach as an entirely separate company (unlike previously, where it used to exist as a division within the BT Group).

Table 2 Network statistical properties (Analysys Mason, 2008)

| Geotype | Classification | Exchanges | Average lines per exchange | Cabinets | Average lines per cabinet | Distribution points | Average lines per distribution point | Average line length (km) |
|---|---|---|---|---|---|---|---|---|
| Inner London | Inner London | 86 | 16,812 | 2,892 | 500 | 172,118 | 8.4 | 1.24 |
| >500k pop | Major City (pop = 500k+) | 204 | 15,512 | 6329 | 500 | 376,721 | 8.4 | 1.78 |
| >200k pop | City (pop = 200k+) | 180 | 15,527 | 5,590 | 500 | 332,713 | 8.4 | 1.8 |
| >20k lines (a) | >20,000 lines, <2km from the exchange | 167 | 17,089 | 6,008 | 475 | 365,886 | 7.8 | 1.5 |
| >20k lines (b) | >20,000 lines, >2km from the exchange | 167 | 10,449 | 4,362 | 400 | 223,708 | 7.8 | 4.83 |
| >10k lines (a) | >10,000 lines, <2km from the exchange | 406 | 10,728 | 9,679 | 450 | 604,925 | 7.2 | 1.4 |
| >10k lines (b) | >10,000 lines, >2km from the exchange | 406 | 3,826 | 4,142 | 375 | 215,740 | 7.2 | 4.0 |
| >3k lines (a) | >3,000 lines, <1km from the exchange | 1,003 | 2,751 | 13,455 | 205 | 493,569 | 5.6 | 0.73 |
| >3k lines (b) | >3,000 lines, >1km from the exchange | 1,003 | 3,181 | 22,227 | 144 | 570,745 | 5.6 | 4.83 |
| >1k lines (a) | >1,000 lines, <1km from the exchange | 1,230 | 897 | 5,974 | 185 | 246,555 | 4.5 | 0.62 |
| >1k lines (b) | >1,000 lines, >1km from the exchange | 1,230 | 935 | 9,343 | 123 | 257,043 | 4.5 | 4.09 |
| <1k lines (a) | <1,000 lines, <1km from exchange | 2,302 | 190 | 0 | 0 | 130,706 | 3.4 | 0.52 |
| <1k lines (b) | <1,000 lines, >1km from exchange | 2,302 | 305 | 0 | 0 | 209,571 | 3.4 | 4.26 |
| | **National Figure:** | 5,578 | 4,886 | 90,000 | 303 | 4,200,000 | 6.5 | 2.33 |

The development of an actual network structure is treated as a pre-processing step in which key datasets are imported, mapped, and exported as a final set of shapefiles for the Digital Twin to access and utilise. The data sources employed in the analysis are outlined in Table 3.

Table 3 Utilised data sources

| Data category | Data | Source | Year |
|---|---|---|---|
| Supply-side | Postcode to exchange data | BT Openreach | 2011, 2013 |
| Supply-side | Postcode to cabinet data | BT Openreach | 2011, 2013 |
| Boundaries | Postcode polygons | Ordinance Survey | 2012 (October) |
| Supply-side | Broadband technology by postcode | Ofcom | 2017 |
| Boundaries | Local Authority District polygons | ONS | 2016 |
| Supply-side | Urban Local Authority Districts by city and size | Analysys Mason | 2008 |
| Demand-side | Premises points data | Ordinance Survey | 2016 |
| Supply-side | Exchange points | SamKnows | 2017 |

Firstly, exchange boundaries are estimated. To achieved this, data are taken for those postcodes with an explicit exchange ID association, and if no ID is available, the postcode is allocated to the nearest exchange using straight line distance. Boundary polygons can then be created for each exchange by dissolving all affiliated postcodes.

Secondly, by either summing the total number of premises, or checking the Local Authority District (LAD) the exchange resides in, exchanges can be geotyped based on the segmentation provided by Analysys Mason (2008) (see Appendix 1).

Thirdly, for those postcodes with data identifying the serving cabinet, we can dissolve the underlying shapes and estimate the cabinet location as the centroid of the new boundary. For those postcodes without cabinet information, the total number of unknown cabinets is achieved by subtracting the number of cabinets we hold data on in each exchange, from the expected average number of cabinets based on the geotype. The number of unknown cabinets is then used to cluster the remaining premises into estimated cabinet areas, with the cabinet location being placed using a k-means clustering algorithm. This same clustering process is repeated for premises, based on the average number of premises served by each distribution point by geotype. Boundaries are then generated from all point locations for cabinet and distribution assets using a Voronoi tessellation approach.

Fourthly, we construct the edges of the network by calculating the shortest path along the road network using an OpenStreetMap topology layer and the OSMNX library. This is carried out for all exchange-to-cabinet links and cabinet-to-distribution links using walking routes for the calculation, as this avoids road traffic restrictions which would skew the estimation process (e.g. one-way roads, bollards etc.), given that trenching is normally laid on or near the proximate pavement. Finally, for distribution-to-premises connections, these are aerially deployed.

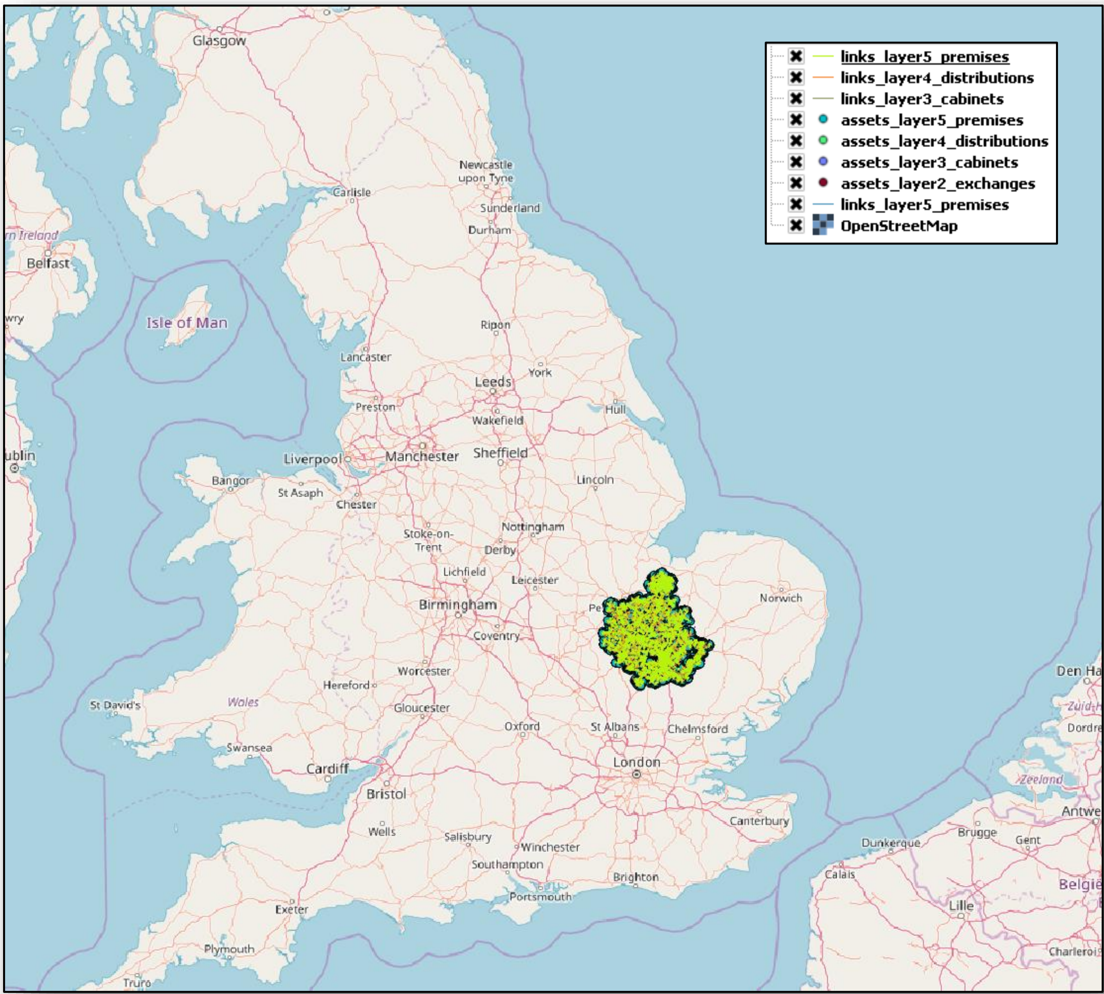
Finally, the technologies already present in the broadband network are identified by using Ofcom (2017). We focus on the incumbent's network (hence, excluding Virgin Media's cable network). All postcodes are assumed to have access to legacy ADSL2+ basic broadband services (<24 Mbps). Premises in postcodes are then allocated technologies based on the modem sync speeds reported as follows:

- >30 Mbps as FTTC VDSL2 availability ('superfast broadband')
- >80 Mbps and <300 Mbps as G.fast pod availability (from the existing cabinet)
- >300 Mbps as FTTP availability ('ultrafast broadband')

Importantly, if FTTC is present, the connected between the exchange and cabinet is therefore determined to be fibre (>90% of premises passed). Additionally, if FTTP is present at the premises, the edges between the exchange and distribution point are assumed to be fibre, as well as the final drop to that premises. All premises, distribution points, cabinets and exchanges are exported as shapefiles, along with the associated boundaries.

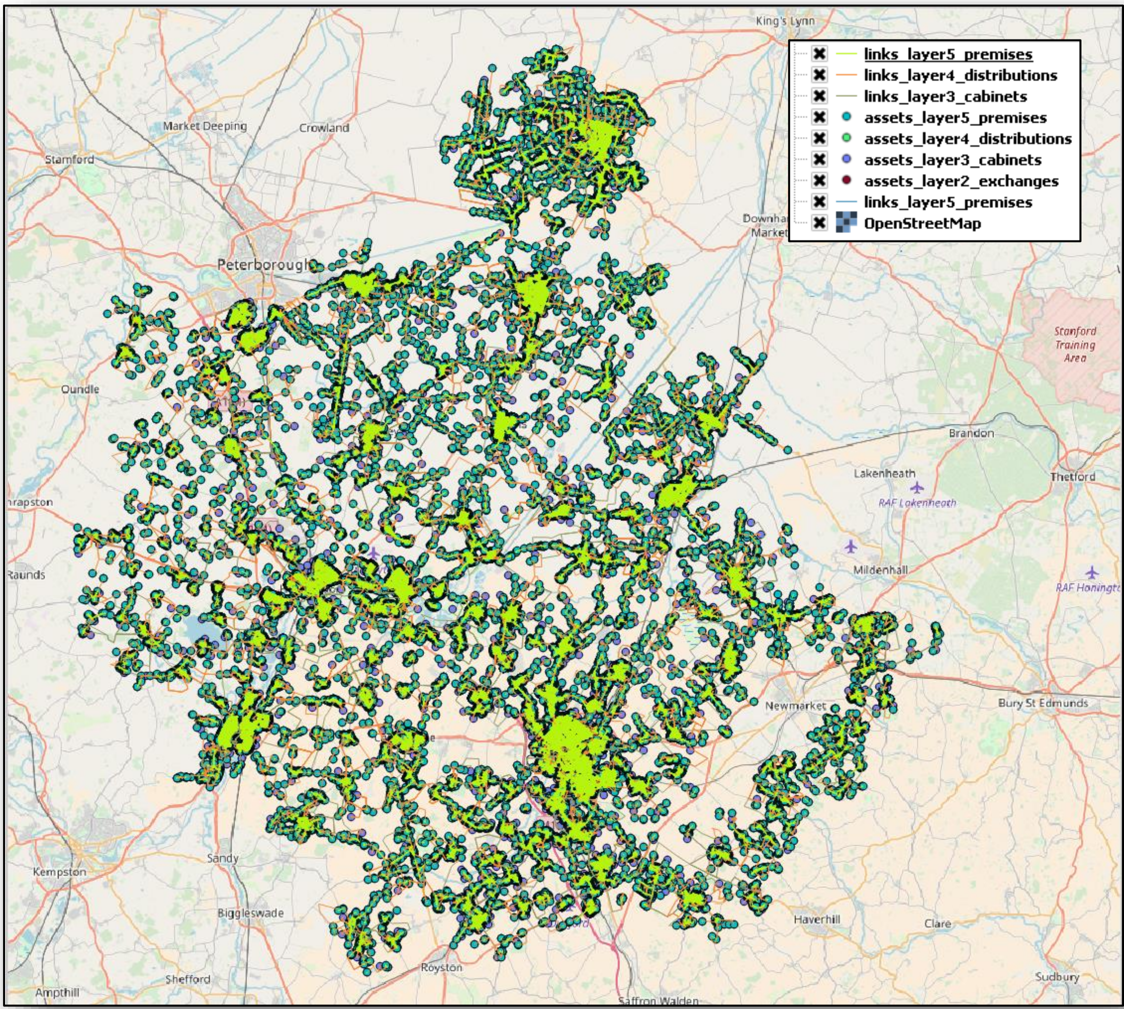
Figure 1 illustrates the geographic location of the Cambridgeshire network subset (in lime green), in the East of England, approximately 80 kilometres north of London. The network is mapped onto layers provided by the open-source OpenStreetMap service.

<u>Figure 1 Geographic location of the network subset for Cambridgeshire, England</u>



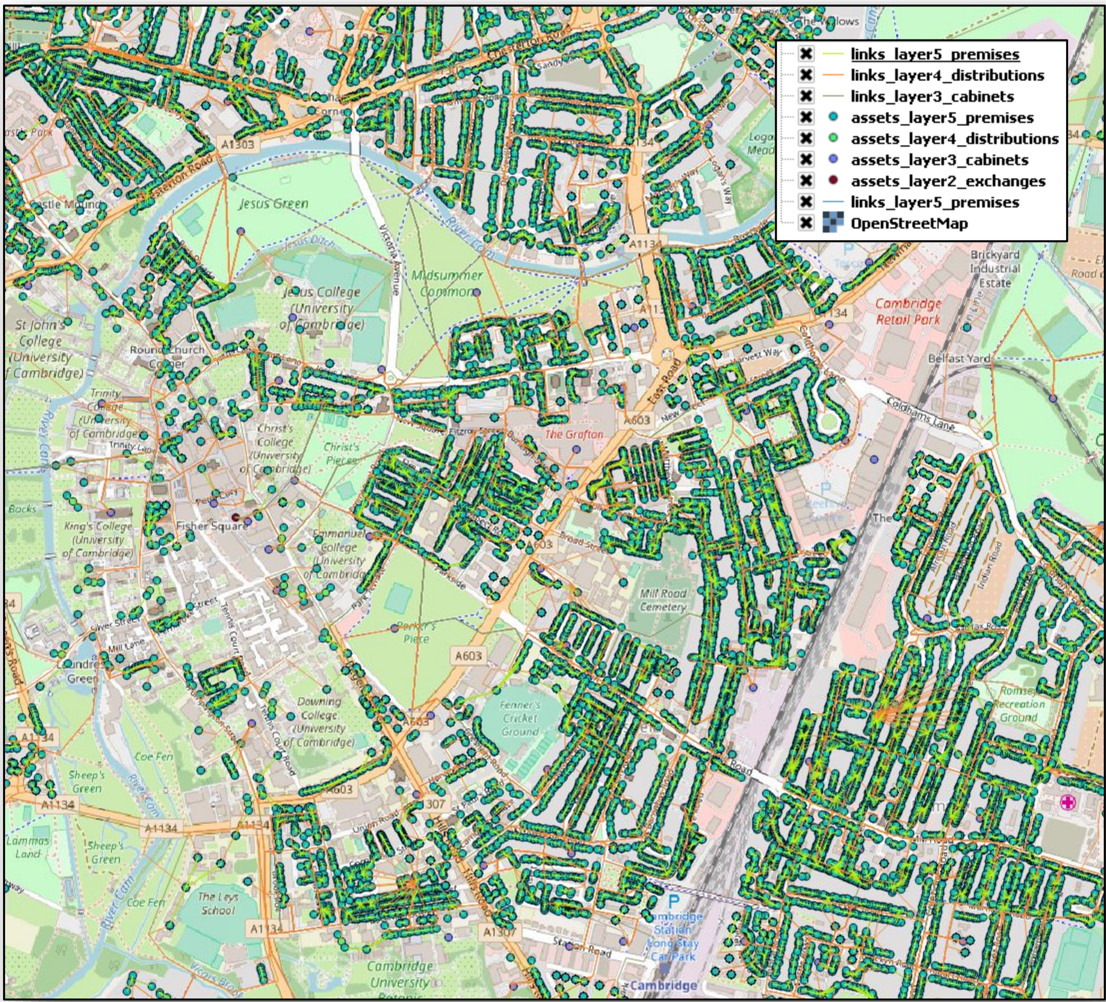

The Cambridgeshire network consists of 290,831 premises, 36,325 distribution points, 1819 cabinets and 71 exchanges across 3,389 square kilometres, as illustrated in greater detail Figure 2.

Figure 2 Synthetic network for Cambridgeshire, England




The local detail of the network structure for the city of Cambridge central exchange is visualised in Figure 3.

Figure 3 Local network structure in the city centre of Cambridge, England




## 3.2 High-resolution demand data

A microsynthesis technique is applied to Britain which combines census data on occupied households, communal residences, and unoccupied dwellings to generate a synthetic population of dwellings classified in several categories. This approach can generate a realistic population of dwellings in a region at a specified geographical resolution, by converting aggregate census data into a synthetic population of individual households that is consistent with census data at the specified resolution. The method is based on quasi-random sampling to directly generate non-fractional populations very efficiently. The method and software implementation of this geographic microsynthesis is open-source and available from [www.github.com/nismod/household_microsynth](www.github.com/nismod/household_microsynth). The geographical units applied using this technique include Middle Super Output Areas (MSOA) and Output Areas (OA).

Data are utilised for individuals based on age, gender at the MSOA level, and for households based on socio-economic status at the OA level. Adoption probabilities for each characteristic are then taken from the Ofcom Technology Tracker H2 2017 survey data (QE21) based on the type of adopted fixed broadband service by household. Data are used for the take-up of fixed broadband services over 30 Mbps. Adoption probabilities derived from this survey data are stated in Table 4.

Table 4 Adoption probabilities by individual and household characteristics

| **Age** ($a_i$) | **Socio-economic status** ($b_i$) | **Nation** ($c_i$) | **Settlement type** ($d_i$) | **Gender** ($e_i$) |
|---|---|---|---|---|
| 16-24 = 36%, | AB = 42%, | England = 39%, | Urban = 40%, | Male = 40%, |
| 25-34 = 41%, | C1 = 40%, | Scotland = 44% | Rural = 34%. | Female = 38% |
| 35-54 = 43% | C2 = 38%, | Wales = 32% | | |
| 55+ = 34% | DE = 32% | | | |

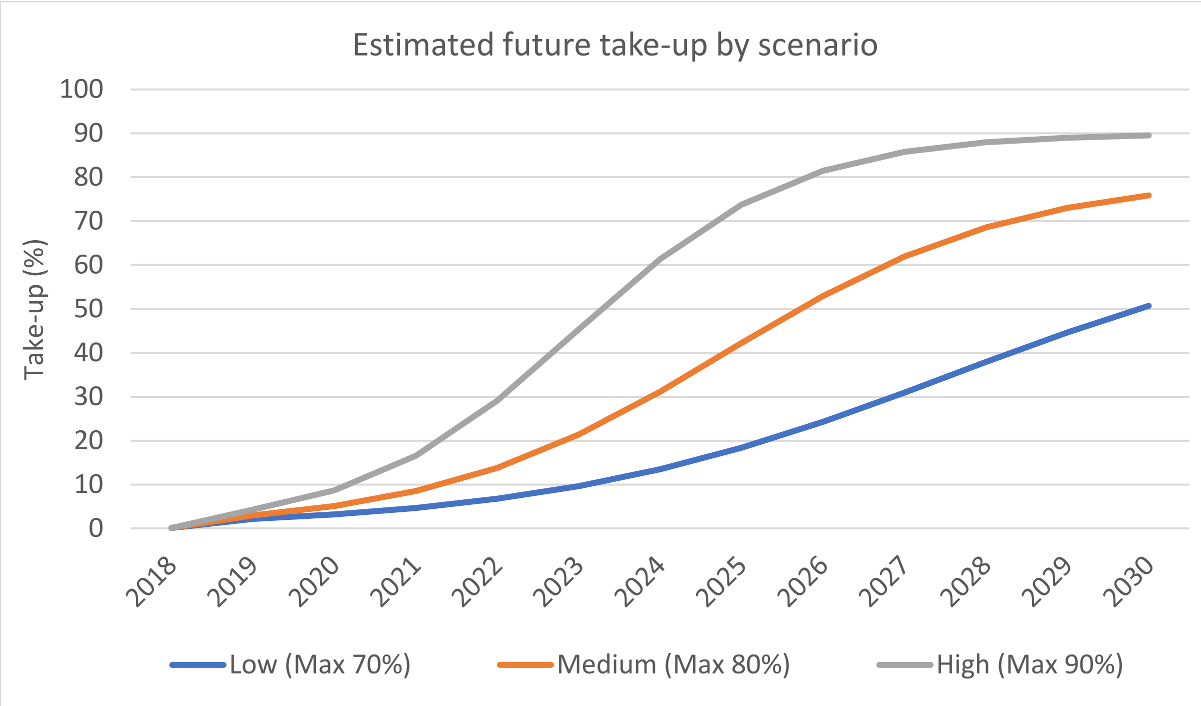
Based on the logic that the multiplication product forms a probability indicator for household propensity to adopt high capacity (>30 Mbps) fixed broadband services, the following equation (1) is applied to each individual:

$$p_i = a_i \cdot b_i \cdot c_i \cdot d_i \cdot e_i \tag{1}$$

After multiplying across all adoption factors, the product provides a propensity to adopt score per geographically located household where $a_i$ represents age, $b_i$ represents socio-economic status, $c_i$ represents nation, $d_i$ represents urban-rural indictor and $e_i$ represents gender. This product per person ($p_i$) is aggregated to the household level, as broadband adoption decisions and the final connection normally serves a single household.

Finally, all households are ranked based on propensity to adopt. Take-up is very hard to predict, hence justifying a scenario-based approach to quantify the uncertainty associated with different levels of adoption. Figure 4 illustrates the logistic curve approach utilised to estimate take-up. Baseline take-up is calibrated based on the historical adoption rate for FTTC taken from the Ofcom (2011:2018) Infrastructure Report series (later renamed 'Connected Nations'), whereby between 2010-2016 adoption reached 31% of premises. Using each scenario forecast, the percentage of adoptees (reported as 'adoption desirability') is selected from the ranked list of households using the propensity to adopt metric.

Figure 4 Estimated future take-up by scenario

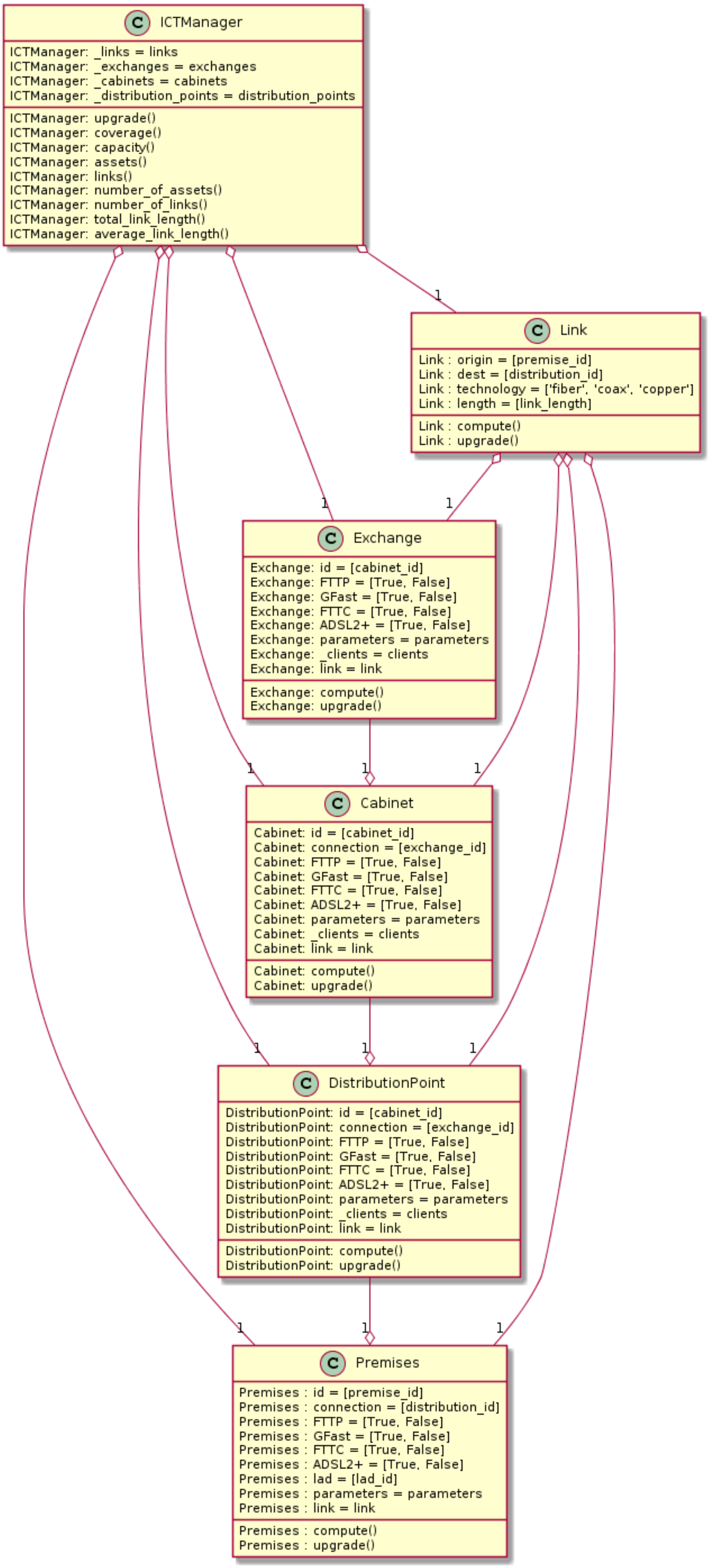



Finally, a distribution is fitted which represents the household Willingness To Pay for broadband. Willingness To Pay is parametrised using Ofcom (2018) data, where the current mean ($\mu$) expenditure is £40. The minimum and maximum values of this distribution are assumed to be £20 and £60 respectively, based on the uSwitch price comparison website (www.uSwitch.com) as of 8th August 2018 using fixed data and voice options only (excluding other services such as sports content).

## 3.3 Assessment model

The shapefiles of the synthetic network can be accessed and manipulated using a GeoJson format. An object-oriented approach is taken to modelling the network nodes and links, as this allows simple but elegant specification of model classes, their features and the interactions. Prior to implementing the code in Python, an object diagram is created in Unified Modeling Language (UML) reflecting the multi-level network structure of the data, as illustrated in Figure 5.

Figure 5 UML diagram of classes within the broadband network structure

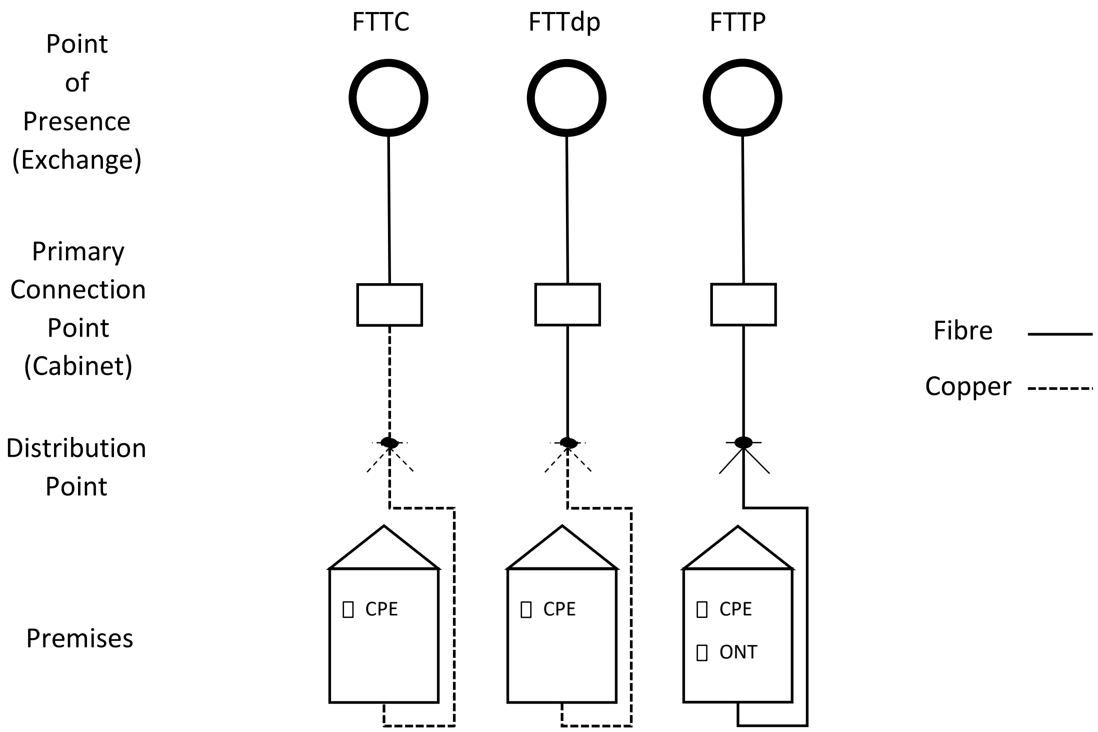


The ICTManager class holds all attributes and methods for the entire hierarchy enabling a model run script to receive and provide necessary data flows. For each time step, the calculation of broadband capacity, coverage and cost can then be estimated.

## 3.4 Network dimensioning

We utilise the technologies outlined within the UK's National Infrastructure Commission's analysis titled 'Costs for Digital Communications Infrastructure' (produced by Tactis and Prism, 2017). Infrastructure reuse is assumed to take place. Firstly, G.Fast is a variant of DSL technology developed to deliver high-speed data transfer over very short distances using existing copper infrastructure. Connections from numerous subscribers are multiplexed via a Digital Subscriber Line Access Multiplexer (DSLAM) into a port connected at the local exchange. Openreach is currently deploying G.Fast from the cabinet aiming to connect 10 million homes by 2020, however, in this analysis we only consider the deployment of this technology from the distribution point for future proofing purposes. For FTTdp (with G.Fast), fibre is laid between the cabinet and distribution point (as illustrated in Figure 6), and a single G.Fast active unit is placed at the distribution point per 8 premises. Premises within 300m of the distribution point can achieve up to 300 Mbps, with the real capacity being dependent on the line length.

Figure 6 Network architecture for FTTC, FTTdp and FTTP



Secondly, for FTTP a GPON Point-to-Point architecture is deployed utilising 32 or 64-way splitting determined economically depending on the number of premises being served. This approach allows speeds of up to 1 Gbps to be delivered, with little impact arising due to the distance between serving assets and the final premises.

3.4 Infrastructure upgrade and policy options

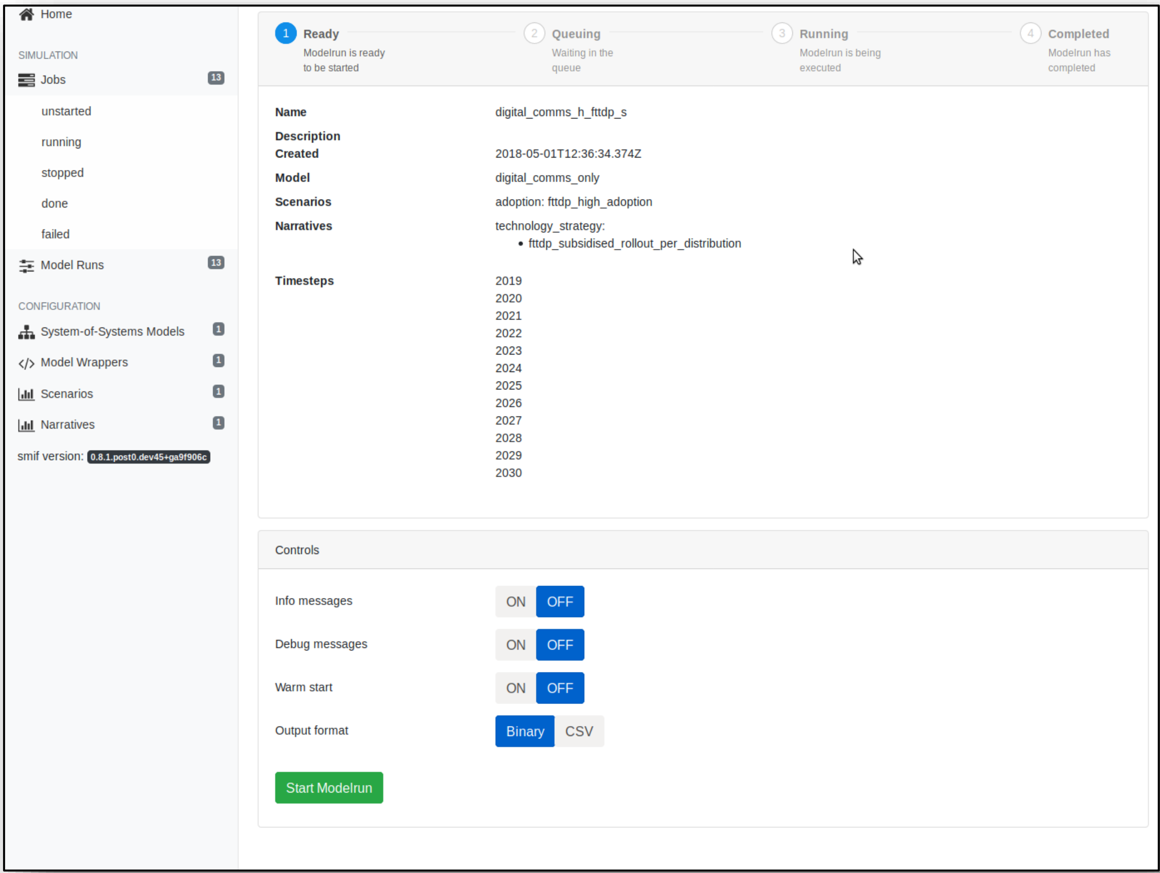
Demand is based on the previously articulated propensity to adopt score, and the subsequently allocated Willingness To Pay quantity for each household. The revenue opportunity available when someone wants to adopt is the monthly Willingness To Pay, for the length of the period in which operators usually expect to make a return from an investment, which is approximately four years. Cost is based on the economics of upgrading an area. The Benefit Cost Ratio ($BCR_{it}$) creates an investment attractiveness index ($BCR_{it} = B_{it} - C_{it}$) which can be ranked nationally, where $B_{it}$ is the potential revenue benefit for the $i$th asset option at time $t$, where $C_{it}$ is the subsequent cost for the $i$th asset option at time $t$. Investments are made rationally if the BCR exceeds one, with the most attractive areas being selected first. Given that the UK's National Infrastructure Assessment recommends that the UK Government should part subsidise the rollout of FTTP, beginning by 2020, a subsidy can consequently be applied to unviable investments to ensure rollout takes place. A matched funding approach is assumed, whereby £5 million from each party can achieve an annual fund of £10 million for the study area, per year. This roughly equates to £500 per premises for the bottom 100,000 least viable premises in Cambridgeshire.

Given the commercial sensitivity of equipment and deployment costs, assumed cost structures are utilised from the UK National Infrastructure Commission's report by Tactis and Prism (2018). The mean cost is utilised for fibre laid across all underground surfaces and aerial deployment options resulting in an approximate installation cost of £10 per meter. Physical infrastructure access to existing assets is assumed, allowing infrastructure reuse of existing ducts and poles. An Optical Connection Point is included per premises at £45, along with a customer equipment totally £74. All civil works and materials are included in these costs. For FTTdp, we assume a G.Fast 8 port distribution point (£200) combined with customer equipment costing (£37).

3.5 Graphical User Interface

Analysts in government and industry can benefit greatly from being able to test their own scenarios and strategies, as this can improve existing understanding and help to make better decisions. A key part of a Digital Twin is the ability to provide a simple user interface for decision-makers to engage with the virtual representation of the telecommunication network. This has been achieved using the React JavaScript library for building user interfaces. Figure 7 illustrates the model run interface, which can run batches of user-defined scenarios for different broadband policy options.

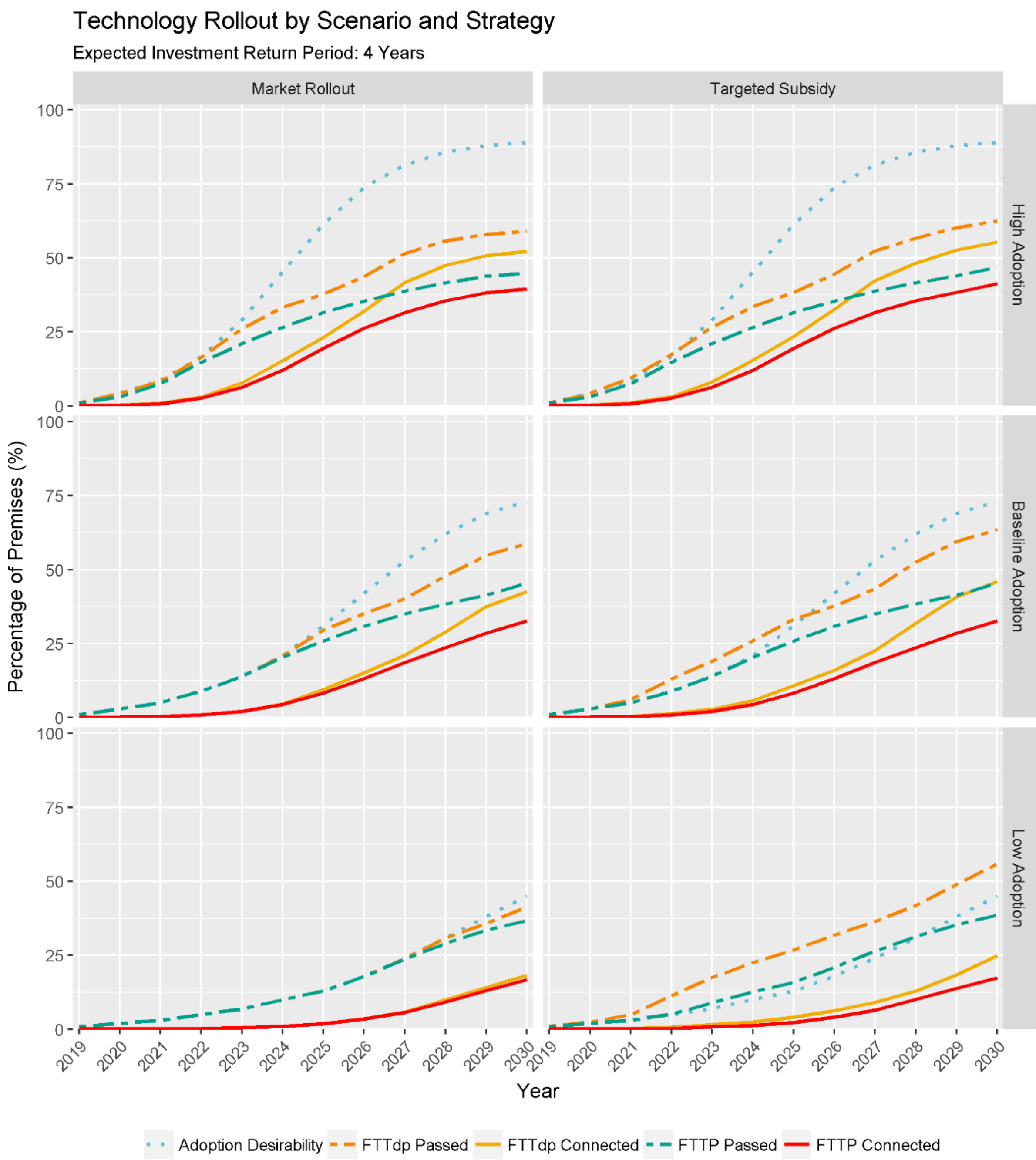
Figure 7 Research software Graphical User Interface



4 Example results

Annual metrics are written out for each year over the study period. These results can either be aggregated or visually mapped to show the key rollout dynamics of the technologies under different policy conditions. This section reports the results for the study area, in support of research question 3. Figure 8 illustrates adoption desirability, as well as the rollout of FTTP and FTTdp according to premises passed and premises connected.

Figure 8 Technology rollout by scenario, strategy and deployed technology



Adoption desirability represents the exogenous take-up specified for each scenario, as illustrated previously in Figure 4. Due to the rational cost-benefit logic of the model, premises passed and premises connected are always lower in the market-based approach. In contrast, in the targeted subsidy strategy, premises passed does exceed adoption desirability in the low adoption scenario, resulting from government support (as has happened with the UK's FTTC rollout due to coverage targets).

Premises passed indicates the percentage of premises with access to either FTTdp or FTTP. Despite adoption desirability being much more substantial in the baseline and high demand scenarios, premises passed remains fairly similar across all scenarios. This is indicative of the economics being reasonably viable for the first third but becoming more and more challenging as the model moves to deploy new infrastructure in suburban and rural areas. As FTTdp is cheaper to rollout, it can be delivered to more premises, reaching 59% by 2030 in the baseline market rollout, when compared to FTTP reaching only 45%.

Premises connected represents those premises with both a desire to adopt and access to the desired service. In the baseline scenario only 32% of premises connect utilising FTTP by 2030, as opposed to 43% able to connect with FTTdp. Importantly, the difference between the level of adoption desirability and premises connected represents the number of premises unable to connect to their desired technology. Hence, in the baseline scenario, adoption desirability is 73% by 2030, meaning 41% of premises wanting to connect to FTTP are unable, opposed to only 30% unable to connect to FTTdp. This illustrates that the modelling approach produces realistic results whereby different users are heterogeneously distributed across space. Location, network effects and economics leads to some users being left unserved due to market failure. Therefore, early adopters are only likely to get the services they desire if they are co-located with enough similar users for the operator to justify the investment upgrade risk. Figure 8 illustrates that the targeted subsidy strategy exhibits little difference in the pace of rollout when compared with the market-based rollout strategy, which is due to the cost component of the cost-benefit equation being substantial in low population density areas.

## 5 Discussion

In this discussion the research questions are returned to and reviewed based on the evidence presented in this paper.

*How may the Digital Twin concept enhance our existing approach to testing policy options?*

Having undertaken a thorough literature review on the purpose and capabilities of a Digital Twin, the evidence suggest that this approach could be utilised by developing a virtual representation of dynamically evolving telecommunication networks. This method could enable improved estimation of factors affecting the rollout of digital technologies which provide new connectivity services. This process is ultimately determined based on the cost of the required digital infrastructure investment, in relation to the potential revenue opportunities new demand segments can offer. Hence, being able to model the interaction between these determining factors can help to increase our current understanding of how adoption desirability, network characteristics and the Willingness To Pay of consumers affect market outcomes.

Moreover, a key point arising from reviewing the telecommunication policy landscape is the underrepresentation in the literature of deductive methodological analyses. Thus, a Digital Twin can help analysts address questions which are highly amenable to this philosophical approach, such as the quantification of uncertainty for decision-making, by providing a platform which allows the testing of 'what if' questions. Importantly, building a Digital Twin in a centralised repository allows key datasets to be collected, stored, maintained and available from a single, trusted and fully-secure source, which is capable of providing decision-support analytics to policy makers.

*What characteristics does a Digital Twin of a broadband network require?*

This paper has proposed several key elements appropriate to deliver on the vision of a Digital Twin. If the definition is to be kept a close as possible to the aim of the original NASA concept, then it is necessary to create a virtual broadband network topology which is spatially-explicit, existing within the structure of the built environment, and which exhibits the same statistical properties as the actual network. Given the hierarchical nature of the underlying broadband network, this motivates the need for a multilevel network structure, which has initially been developed here purely from a topological perspective, but which could also be developed to include traffic routing in the future. While this initial network modelling effort in developing a Digital Twin has been substantial, the incorporation of household-level demand has equally been ambitious. Premises-level estimates of demand based on disaggregated demographic and socio-economic forecasts enable the key characteristics driving broadband adoption to be captured. In this paper a single policy option was tested to illustrate the capability of the Digital Twin, however future research must enhance the number of options available in the model for testing in this virtual environment. Finally, a key contribution is the development of a Graphical User Interface for the Digital Twin, allowing users to utilise its capabilities as an Application-as-a-Service via a web browser. The next steps must include development of a dashboard of key performance metrics to help improve the user evaluation of broadband availability and adoption.

*How do different policy options perform in influencing the rollout of FTTdp or FTTP?*

A key issue is whether technology rollout should focus on coverage over capacity (in the case of FTTdp), or capacity over coverage (in the case of FTTP). The rollout efficiency of FTTdp is attractive, especially given the current lack of applications for FTTP-like speeds (e.g. 1Gb/s), while the future proofing of FTTP (and the reliability an FTTP connection can provide) is also widely beneficial (albeit at higher cost). The baseline results produced from the study area indicate that left fully to the market, approximately two thirds of premises will be viably passed using FTTdp and just under half using FTTP by 2030, under the current business model. Hence, this is indicative of the need for a change in the investment return horizon to improve the Benefit-Cost Ratio which drives rational investment decision-making. Currently, UK telecom operators desire an investment return within at least four years, which does not seem to support the fibre economics illustrated in this paper, based on current capital expenditure costs. The difference in the expected investment return horizon is most noticeable when comparing traditional operators with the Alternative Network Providers (the 'Alt Nets'), who take advantage of investment capital from longer-term investors (pension funds, venture capitalists etc.) who look at FTTP investment over a return horizon of decades rather than years (hence, in line with other infrastructure investments in power plants, toll roads and water processing facilities). Now a substantial proportion of the development of the Digital Twin has taken place, further research needs to explore the business model implications of fibre investment to provide greater understanding of market-based deliver methods, and indeed how the costs of delivery could be reduced.

## 6 Conclusion

This paper undertook an evaluation of the Digital Twin approach for developing a virtual representation of a fixed broadband network. Key elements of a Digital Twin developed for this purpose are identified, including a spatially-explicit multilevel broadband network topology, and a set of premises-level demand forecasts for demographic and socio-economic broadband adoption factors. These building blocks can provide new insight into the effectiveness of a range of supply- and demand-side policy options that can then be tested utilising this virtual representation, via a Graphical User Interface for non-technical decision-makers. Hence, a key contribution of this paper is to demonstrate that this can be achieved and made available via an open-source research software codebase. Much of the debate associated with disparities in broadband deployment pertains to market failure and the lack of adequate connectivity in certain areas. Often the costs of deployment are disputed between the incumbent and other operators, and there are few tools that help with this problem. Indeed, over the long-term a Digital Twin could help to improve transparency, generate new knowledge of market-based deployment methods, and allow the testing of more experimental policy options to foster innovation in tackling perennial and emerging digital divide issues. The conclusion reached is that a Digital Twin can aid in improving existing estimates of those factors affecting the rollout of digital technologies, therefore this framework should be pertinently developed to test the effectiveness of a wider range of policy options available to decision-makers.

## 7 References


Analysys Mason. 2008. 'The Costs of Deploying Fibre-Based next-Generation Broadband Infrastructure'. Cambridge: Analysys Mason. http://www.analysysmason.com/PageFiles/5766/Analysys-Mason-final-report-for-BSG-(Sept2008).pdf.

Bicheno, Scott. 2018. 'Openreach Makes Big FTTP Statement, but with Strings Attached'. Text. Telecoms.Com. 1 February 2018. http://telecoms.com/487539/openreach-makes-big-fttp-statement-but-with-strings-attached/.

Boschert, Stefan, and Roland Rosen. 2016. 'Digital Twin—The Simulation Aspect'. In *Mechatronic Futures*, 59–74. Springer, Cham. https://doi.org/10.1007/978-3-319-32156-1_5.

Briglauer, Wolfgang. 2014. 'The Impact of Regulation and Competition on the Adoption of Fiber-Based Broadband Services: Recent Evidence from the European Union Member States'. *Journal of Regulatory Economics* 46 (1): 51–79. https://doi.org/10.1007/s11149-013-9237-4.

Choudrie, Jyoti, and Anastasia Papazafeiropoulou. 2007. 'Assessing the UK Policies for Broadband Adoption'. *Information Systems Frontiers* 9 (2–3): 297–308. https://doi.org/10.1007/s10796-006-9022-3.

Falch, Morten, and Anders Henten. 2017. 'Dimensions of Broadband Policies and Developments'. *Telecommunications Policy*, November. https://doi.org/10.1016/j.telpol.2017.11.004.

Gelernter, David. 1993. *Mirror Worlds: Or the Day Software Puts the Universe in a Shoebox...How It Will Happen and What It Will Mean*. Oxford University Press.

Götz, Georg. 2013. 'Competition, Regulation, and Broadband Access to the Internet'. *Telecommunications Policy* 37 (11): 1095–1109. https://doi.org/10.1016/j.telpol.2013.03.001.

Haag, Sebastian, and Reiner Anderl. 2018. 'Digital Twin–Proof of Concept'. *Manufacturing Letters* 15: 64–66.

Haucap, Justus, Ulrich Heimeshoff, and Mirjam R. J. Lange. 2016. 'The Impact of Tariff Diversity on Broadband Penetration—An Empirical Analysis'. *Telecommunications Policy*. https://doi.org/10.1016/j.telpol.2015.09.005.

Jakopin, N. M., and A. Klein. 2011. 'Determinants of Broadband Internet Access Take-up: Country Level Drivers'. *Info* 13 (5): 29–47. https://doi.org/10.1108/14636691111160626.

Kongaut, Chatchai, and Erik Bohlin. 2015. 'Towards Broadband Targets on the EU Digital Agenda 2020: Discussion on the Demand Side of Broadband Policy'. *Info* 17 (3): 1–15. https://doi.org/10.1108/info-02-2015-0017.

Lin, Mao-Shong, and Feng-Shang Wu. 2013. 'Identifying the Determinants of Broadband Adoption by Diffusion Stage in OECD Countries'. *Telecommunications Policy* 37 (4–5): 241–51. https://doi.org/10.1016/j.telpol.2012.06.003.

National Aeronautics and Space Administration. 2010. 'Modeling, Simulation, Information Technology & Processing Roadmap: Technology Area 11'. Washington DC: National Aeronautics and Space Administration.

National Infrastructure Commission. 2017. 'Data for the Public Good'. London: National Infrastructure Commission. https://www.nic.org.uk/publications/data-public-good/.

———. 2018. 'National Infrastructure Assessment: An Assessment of the United Kingdom's Infrastructure Needs up to 2050'. London: National Infrastructure Commission. https://www.nic.org.uk/assessment/national-infrastructure-assessment/.

Negri, Elisa, Luca Fumagalli, and Marco Macchi. 2017. 'A Review of the Roles of Digital Twin in CPS-Based Production Systems'. *Procedia Manufacturing* 11: 939–948.

Ofcom. 2017. 'Connected Nations 2017'. London: Ofcom. http://stakeholders.ofcom.org.uk/market-data-research/market-data/infrastructure/connected-nations-2015/.

Oughton, Edward J., Zoraida Frias, Mischa Dohler, Jason Whalley, Douglas Sicker, Jim W. Hall, Jon Crowcroft, and David D. Cleevely. 2018. 'The Strategic National Infrastructure Assessment of

Digital Communications'. *Digital Policy, Regulation and Governance* 20 (3): 197–210. https://doi.org/10.1108/DPRG-02-2018-0004.
Reuters. 2018. 'Hammond Sets Target for 15 Million "full-Fibre" Broadband...' *Reuters*, 22 May 2018. https://uk.reuters.com/article/uk-britain-politics-hammond/hammond-sets-target-for-15-million-full-fibre-broadband-connections-by-2025-idUKKCN1IN29Z.
Ricci, Alessandro, Luca Tummolini, and Cristiano Castelfranchi. 2017. 'Augmented Societies with Mirror Worlds'. *AI & SOCIETY*, December, 1–8. https://doi.org/10.1007/s00146-017-0788-2.
Schleich, Benjamin, Nabil Anwer, Luc Mathieu, and Sandro Wartzack. 2017. 'Shaping the Digital Twin for Design and Production Engineering'. *CIRP Annals* 66 (1): 141–144.
Think Broadband. 2018. 'Chancellor Sets 15 Million Full Fibre Premises Target'. 2018. https://www.thinkbroadband.com/news/8054-chancellor-sets-15-million-full-fibre-premises-target.
Tuegel, Eric J., Anthony R. Ingraffea, Thomas G. Eason, and S. Michael Spottswood. 2011. 'Reengineering Aircraft Structural Life Prediction Using a Digital Twin'. Research article. International Journal of Aerospace Engineering. 2011. https://doi.org/10.1155/2011/154798.
Uhlemann, Thomas H.-J., Christian Lehmann, and Rolf Steinhilper. 2017. 'The Digital Twin: Realizing the Cyber-Physical Production System for Industry 4.0'. *Procedia Cirp* 61: 335–340.
Valletti, Tommaso M., Steffen Hoernig, and Pedro P. Barros. 2002. 'Universal Service and Entry: The Role of Uniform Pricing and Coverage Constraints'. *Journal of Regulatory Economics* 21 (2): 169–90. https://doi.org/10.1023/A:1014387707275.
West, Timothy D., and Mark Blackburn. 2017. 'Is Digital Thread/Digital Twin Affordable? A Systemic Assessment of the Cost of DoD's Latest Manhattan Project'. *Procedia Computer Science* 114: 47–56.

Appendix 1 Urban area look-up table

| Major Agglomerations | Minor Agglomerations |
|---|---|
| Bristol | Aberdeen |
| Glasgow | Aldershot |
| Manchester | Birkenhead |
| Liverpool | Blackpool |
| Nottingham | Bournemouth |
| Sheffield | Brighton |
| Newcastle | Cardiff |
| Birmingham | Coventry |
| Leeds | Barnsley |
| | Derby |
| | Edinburgh |
| | Kingston |
| | Leicester |
| | Luton |
| | Northampton |
| | Norwich |
| | Plymouth |
| | Portsmouth |
| | Preston |
| | Reading |
| | Southampton |
| | Southend |
| | Swansea |
| | Middlesbrough |
| | Gillingham |
| | Stoke-on-Trent |