Programme Area: Smart Systems and Heat

Project: EnergyPath

Title: EnergyPath Networks Functional Specification V2.1

Abstract:

The Tool Kit Development Project sits within Work Package 2, and supports the Development of the EnergyPath Networks Tool. This Deliverable comprises a report which sets out the Functional Specification for Version 2.1 of the EnergyPath Networks Tool. This report describes the functional requirements and specification which have driven the development of the tool.

Context:

Energy consultancy Baringa Partners were appointed to design and develop a software modelling tool to be used in the planning of cost-effective local energy systems. This software is called EnergyPath and will evolve to include a number of additional packages to inform planning, consumer insights and business metrics. Element Energy, Hitachi and University College London have worked with Baringa to develop the software with input from a range of local authorities, Western Power Distribution and Ramboll. EnergyPath will complement ETI's national strategic energy system tool ESME which links heat, power, transport and the infrastructure that connects them. EnergyPath is a registered trade mark of the Energy Technologies Institute LLP.

Energy Systems Catapult
Smart Systems and Heat Programme
The EnergyPath™ Networks (EPN) Tool
Version 2.1

(D1) Functional Specification

CLIENT: Energy Systems Catapult

DATE: 10/05/2017

www.baringa.com

Version History

Version	Date	Description	Prepared by	Approved by
V0_1	19/03/2014	Initial draft	JG, PT, LH, EP, FL, NAM, FT	OR
V1_0	23/4/2014	Final	JG, PT, LH, EP, FL, NAM, FT	OR
V2_0 Draft	22/2/2017	Update for v2.1 of EPN – Draft for ESC/ETI review	JG, LH, AB, FT	JG
V2_1	10/5/2017	Minor updates	JG, LH, AB, FT	JG

Copyright

Confidentiality and Limitation Statement

TABLE OF CONTENTS

1. Introduction

1.1. Background and context

In 2012, the Energy Technologies Institute LLP (ETI) launched the Smart Systems and Heat (SSH) programme, aimed at developing a cost effective smart energy system supporting reduction in CO_2 emissions associated with energy use in buildings in the UK. This is undertaken in the context of the UK's commitment under the 2008 Climate Change Act to reduce emissions of greenhouse gasses by 80% versus 1990 levels by 2050. SSH comprises a number of Work Areas (WAs), including WA2 – EnergyPath[TM] design tools.

The objective of WA2 was to deliver a tool "the EnergyPath[TM] Networks (EPN) Tool" that would enable the ETI to work with Local Authorities (LAs) and their key strategic stakeholders to answer two primary questions:

> ► For each area, what is the optimum contribution to the UK 2050 target?

> ► For a given area, what is the optimum plan for transitioning energy supply networks to meet local needs in a manner aligned with the 2050 target?

The ETI issued a Request for Proposals (RfP) for WA2 in June 2013, and subsequently selected a team led by Baringa Partners and including Element Energy, Hitachi Europe Limited and University College London (UCL), to undertake the work. The initial work was undertaken in two stages:

> ► Stage 1 – detailed functional specification and development plan for Stage 2, completed in April 2014; and

> ► Stage 2 – delivery of a functional prototype version "1.0" of EnergyPath[TM] in October 2014 and version "2.0" in February 2015.

In 2015, responsibility for delivery of the SSH programme transferred into the Energy Systems Catapult[1], including the use and ongoing development of the EPN tool. As of the start of the 2017 the EPN tool has been used to inform an Energy Transition plan for Newcastle County Council and is currently supporting the development of a similar plan for Bridgend County Council.

Following the completion of the original EPN development project a number of updates have also been made to the tool itself, culminating in version 2.1 of the tool at the start of 2017. This document has been updated to reflect key changes from the original functional specification created in April 2014.

1.1.1. Stage 1 deliverables

The Functional Specification (D1) is one of seven Stage 1 deliverables. It should be read in conjunction with the other deliverable documents:

---

[1] https://es.catapult.org.uk/faq/what-is-the-smart-systems-and-heat-ssh-programme/

- ▶ (D2) Design Architecture

- ▶ (D3) Data Architecture

- ▶ (D4) Data Acquisition Plan

- ▶ (D5) Stage 2 Proposal

- ▶ (D6) IP Statement

- ▶ (D7) QA processes and test plan

It should be noted that these documents have not been updated to reflect version 2.1 of the EPN tool, however, the accompanying EPN tool "User Guide" has been updated.

1.2.   High-level objectives

1.2.1. Tool objectives

The key objectives with regard to the EnergyPath™ tool are as follows:

- ▶ To create a strategic planning tool to support decision-making on future local area energy systems to 2050, principally to help support Local Authorities (LAs) and their strategic stakeholders, such as Distribution Network Operators (DNOs)

- ▶ Specifically, for the tool to help prioritise and plan interventions in the local area energy system, including generation, network, storage and buildings projects, aligned with a national 2050 decarbonisation pathway

- ▶ For the tool to do this by focussing on the development and application of a cost-effective local plan, which is subject to 'real world' constraints and uncertainty; to help inform, more subjective, strategic decision making on a transition pathway as part of wider business planning processes;

- ▶ Be able to account, at least indirectly, for the potential impact of consumers on local area energy system pathway design; and

- ▶ To ensure that the tool and associated databases are based around a scalable architecture, in order to set the ETI/ESC on the trajectory towards building the full level of functionality and capability.

1.2.2. Key end-user requirements

The ETI (and now the ESC) is assumed to be the main 'model user', however, the outputs and insights generated by the EnergyPath™ Networks Tool are likely to be used primarily by LAs. Although they currently have no explicit requirement to set emissions targets, a large proportion of emissions are within their influence (most particularly from residential buildings), and according to the Committee on Climate Change they have a vital role to play in decarbonising the economy. Their main means of intervention in local energy provision are summarised below.

- **Local development plans:** used to identify sites for specific land uses and set out the criteria for approving planning applications.

- **Economic development strategy:** planning for the development of employment and skills in a local area.

- **Planning permission for buildings:** for individual new buildings or extensions, LAs can stipulate planning conditions.

- **Planning permission for power plant:** LAs have responsibility for authorising plant less than 50 MW.

- **LA led initiatives:** LAs may be best placed to initiate multi-party schemes such as district heating, possibly using Public Private Partnerships (PPP).

- **LA building stock management**: including both social housing and public buildings such as libraries, council offices, etc.

- **ECO / Green Deal:** LAs can be Green Deal providers or delivery partners.

- **LA incentive levers:** LAs have limited control over tax and spending policy (e.g., business rates) that could be used to accelerate and direct investment.

- **Community engagement:** advice from LAs, being responsible for residents' wellbeing, may be more accepted by the public as objective and impartial.

- **Accessing funding:** LAs play a key role in accessing funding for local energy schemes from devolved authorities, UK and EU level.

The diagram below identifies ways in which EnergyPath™ could assist LAs in performing their role in local energy provision.

Figure 1-1   The role of EnergyPath™ Networks Tool in assisting LAs plan local energy provision
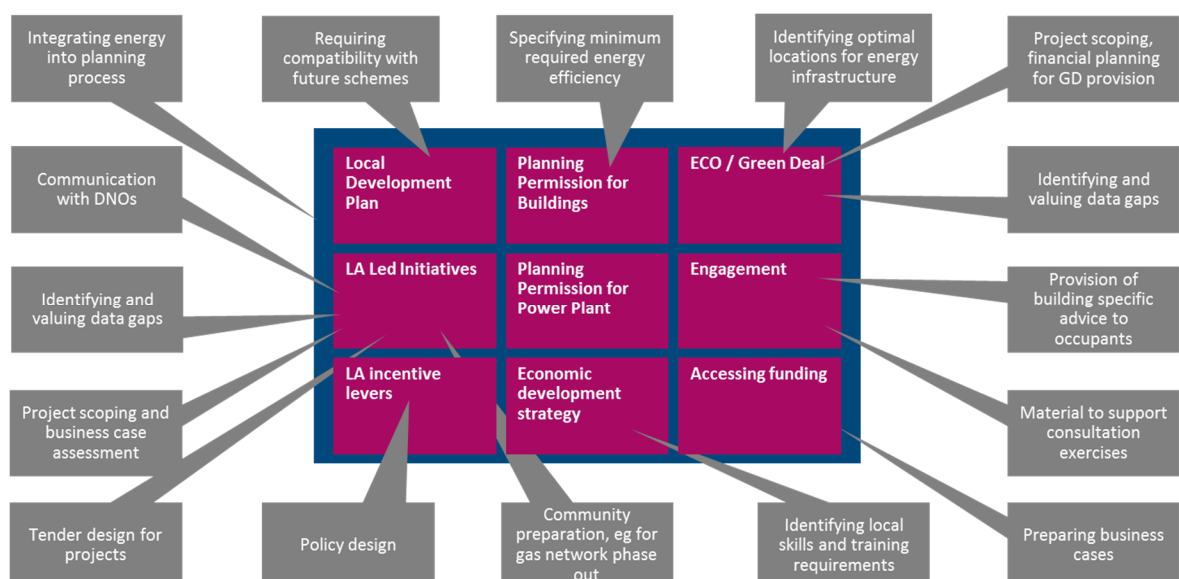
LAs are likely to use the outputs of the tool to engage with other parties including network operators, property developers and landlords, energy companies, financiers, local companies, community energy groups and final consumers. This engagement will include investment planning in different elements of the local energy system, including buildings of different types, energy sources and networks. The actors in local energy and their roles in different elements of the local energy system are summarised in the table below.

1.2.3. Key end-uses

Across all applications and users, we have categorised the main contexts in which the tool's 'outputs' will be used under five headings, listed below with example uses.

▶ **Local planning:** developing local master plans; setting local energy objectives; or specifying local interventions

▶ **Network planning:** helping to support the strategic, as opposed to detailed, planning of the extension, reinforcement or retrenchment of energy networks; or promoting new generation in constrained areas to avoid network investment

▶ **Financial planning:** assessing project costs, revenues and risks for investment decisions; estimating investment requirements for regulated assets; or understanding the impact of transition on existing business models

▶ **Engagement:** informing decision makers on optimal choices; engaging with LA executives on energy plans; or engaging with local communities on energy plans

▶ **National planning:** establishing local targets consistent with national targets

It is envisaged that the direct outputs of the tool are tailored (e.g. different graphical outputs or post-processing of results) for each of these uses, given the different stakeholders involved. In all these contexts, the tool must support and inform what is, ultimately, subjective decision strategic making. As such, it:

▶ Allows the user to develop and analyse the timing and quantum of potential deployment options and associated costs that together create a viable pathway towards a desired outcome, and to be able to compare between alternatives.

▶ Enables comparison to outcomes with no interventions ("Business as Usual") to provide a counterfactual case without intervention.

▶ Indicates the uncertainty associated with the analysis, and allow for testing of different scenarios, to enable the robustness of a decision to be tested against different outcomes. This will be supported by providing transparency in outputs, such that they can be flexibly disaggregated and re-characterised so as to provide a more nuanced understanding.

▶ Once particular deployment options have been identified, the tool supports the design of projects to implement these. As such it outputs sufficient detail on the actions and measures employed for a particular option to allow scoping of projects.

1.2.4. Key end-user expertise

As outlined in section 3, the tool requires a complex mix of energy system modelling and subject matter areas. Whilst the user interface has been designed to minimise the interaction with some of the more complex components (such as network load flow modelling) they still need a basic understanding of the underlying components and their concepts. For example, to be able to configure and apply the tool, undertake diagnostics of problems caused by input data (such as optimisation infeasibilities) and interpret the results.

As such, it is assumed that the model is designed only for an expert user, who has the following competencies:

▶ Strong quantitative skills, e.g. via a first degree in mathematics, engineering, economics, physics, etc.

▶ Basic understanding of key mathematical modelling techniques including optimisation and Monte Carlo sampling

▶ Basic understanding of key energy system modelling techniques such as

  – Electricity and heat load flow modelling

  – Whole energy systems modelling

  – Building energy simulation

▶ Basic understanding of key software packages including

  – GIS, to be able to interact with, add new or interpret key spatial input data and results

  – Databases, to be able to install and manage overarching sets of input data and results files across the tool

▶ Detailed energy systems understanding of

  – Domestic and non-domestic buildings

  – Energy networks - electricity, heat, gas and hydrogen

  – Other local area features – e.g. embedded generation

  – Other national level energy system features that impact on the solution – e.g. transmission connected electricity generation

In addition to the above, development skills will also require a core understanding of some of the key software languages proposed for the tool. These are outlined in more detail in the Design Architecture Deliverable (D2), but include SQL database programming, the AIMMS optimization language and Python.

1.2.5. Outputs and their use

Ultimately, the tool allows results to be interrogated and analysed in whatever manner and through whatever metrics are of value to the final users. Section 4 explores in more detail some means of presenting results along with the overarching use of the tool to deliver insights.

A typical interrogative process a user might be expected to go through, drilling down into the results from a high level overview down to specific projects, and some indication of outputs to facilitate each stage, are outlined below:

▶ **What are the high-level features of the overall area pathways?** For example, how total costs (capital and operational) vary for different pathways versus BaU over the horizon to 2050; how household energy costs are impacted[2]; the technologies which are key in implementing the pathway.

▶ **What are the key geographical and underlying features of different pathways?** For example, how would a new heat network will be laid out; what would be the location of plant supplying this network; and how would a map of cost of energy provided overlap with the map of fuel poverty?

▶ **What are the key points of uncertainty in the pathways and the value of reducing this?** For example, how results are distributed around the average and the risk of high cost outcomes; the reduction in cost uncertainty achieved from obtaining better data, for example on the installation cost of a particular measure in an area; the key external factors to which a pathway is sensitive.

▶ **What does the pathway imply for specific projects?** For example, the roll out of solid wall insulation in postcode XY12; the summary financials for a new CHP plant, determining the economic case for distributed energy storage, or the viability of transitioning a small geographic location to electric heating and potentially decommissioning the gas grid.

Key end-users of the tool, such as the ETI/ESC are likely to be the most demanding, requiring the highest level of interaction with the assumptions and operation of the model. The ETI/ESC, or potential subcontractors, will work closely to support LAs as well as other *indirect* users. The latter will be focused primarily on the details of the results, as opposed to necessarily using the tool themselves. Indirect users will also have varying degrees of interest in the outputs.

The diagram below summarises how different types of user may attach different levels of importance to the level of their interaction with the model.

Table 1-2   Required level of model interaction by user

High level of interaction	Low level of interaction

---

[2] As opposed to energy bills directly, as this is ultimately affected by the distributional considerations of various regulatory and market structures, for example whether network costs are incurred directly by the end-user (on a capacity or utilisation basis) or socialised across a wider set of end-users.

User	Value of usage mode				
	Full interactive model use	Full local area model outputs	Results for local area consultation	Final local area pathway design(s)	Aggregated insights for the UK
ETI/ESC	High	High	High	High	High
ETI members	Low	Medium	Low	Low	High
Local Authorities	Medium	High	High	High	Medium
Electricity and gas Network operators	Medium	Medium	High	High	High
Heat network developers / operators	Medium	High	High	High	Medium
Academic / research users	High	High	Low	Low	High
Energy companies	Low	Low	High	High	High
Central Government	Low	Medium	Medium	Medium	High

1.2.6. Informing the national pathway

A key objective for the EnergyPath™ Networks Tool is to examine local area pathways that are consistent with national decarbonisation pathways to 2050. However, by its nature the tool will provide a more realistic assessment of the costs and effective deployment options in a local area, by more accurately reflecting real-world spatial implications.

These insights may lead to a different aggregate solution at the national level, for example, the costs of large-scale district heat deployment may be higher than anticipated. This may focus national level solutions more towards electrification, but with commensurate impact on transmission connected electricity costs, which in turn affects the local area solution. Hence the interaction between the local area and national level pathway is ultimately a *2-way* iterative process.

It is clearly not pragmatic to wait for local area studies to be undertaken with every LA in the UK before updating the national pathway and hence there are two main routes for feeding back the insights gained at the local level:

▶ As early studies are undertaken with small numbers of LAs the key features of the pathway designs and the local area can be extrapolated, based on how representative they are as a proxy for other parts of the UK.

▶ 'Representative', hypothetical local areas (e.g. heavily urban, suburban, rural or mixed) could be generated for use in EnergyPath™, with insights from the resulting pathway

designs used to inform the national level – i.e. under what key local area conditions are certain designs favoured over others[3].

1.3. The EnergyPath^TM Networks Tool within the modelling landscape

The EnergyPath^TM Networks Tool is unique in the problem that it is trying to tackle, which combines the assessment of what a 'good' pathway design for a local area energy system is whilst considering in parallel:

1. A granular representation of local area building and network level options, that can easily accommodate more real-world data as it becomes available (e.g. from building level surveys)

2. Comparing trade-offs across multiple energy vectors and technology choices, across buildings, networks and other energy system features

3. Consistency with relevant features of the 'National Pathway decarbonisation blueprint'

4. Evolution of the design over a time horizon to 2050

5. Consideration of real-world constraints (e.g. planning restrictions) and the indirect impact of constraints imposed by consumer behaviour (on energy demand behaviour or technology uptake) on the pathway designs

Existing tools target one or more of these elements, but not in combination. For example,

▶ National-level energy system optimisation models like ESME or UK MARKAL explore 2, 3 and 4, but have very limited or no representation of individual geographic areas

▶ A wide range of planning tools or stock models for buildings, or electricity and heat networks, exist individually to tackle 1. E.g. WPD's spatially detailed, full-load flow modelling Falcon SIM[4] project or the more aggregated Transform model[5] which only consider electricity. Or alternatively, CitySim[6] which provides a very detailed spatial representation of building energy simulation – at individual building level - across an area. However, all of these do not tackle 2 as they only consider a small set of energy system options or energy vectors; and as result do not attempt to trade-off the

---

[3] For the long-term development of the EnergyPath^TM Networks Tool it would be possible to design an area generator to quickly create a random set of local area conditions based on key input parameters such as the total number of buildings and % urbanity.

[4] http://www.westernpowerinnovation.co.uk/Falcon.aspx

[5] Created under the Auspices of the DECC / Ofgem Smart Grid Forum All GB DNOs, Ofgem and DECC have a Royalty free licence to use the software. Other users may access the model on a commercial basis. The model only assess example, rather than spatially specific networks, and does not undertake network flow modelling.

[6] http://citysim.epfl.ch/

suitability of developing a heat network against longer-term electricity reinforcement, subject to wide-spread heat-electrification and insulation measures.

► In addition, many detailed planning tools focus on the relatively near or medium term rather than a longer term pathway (4) and are not understood within the context of wider national decarbonisation objectives (3)

The closest to the proposed EnergyPath[TM] Networks Tool is UrbEn[7] (integrated modelling framework for urban energy systems – also known as SynCity) was produced by Imperial as part of a BP funded research project.   This is comprised of four inter-linked layers looking at i) the optimisation of overall city land use layout (primarily to minimise travel requirements), ii) agent based modelling of transport and energy demand behaviour, iii) optimisation of energy use and transport of resources, iv) detailed network operation. UrbEn passes information through layers i) to iv) and then iterates to explore solutions for urban city planning over time which help to minimise carbon emissions. The question UrbEn is trying to tackle is, however, much broader than the EnergyPath[TM] Networks Tool and extends to transport planning and agent behaviour, which means there is correspondingly less detail in the pathway optimisation for the energy system aspects (only partial optimisation of some elements in layer iii) and more limited resolution of building energy use. For example, the more detailed engineering network configuration in iv) is estimated after the more aggregated flows have been established iii). This potentially runs the risk of specifying less appropriate upgrades by starting with the simplified optimisation earlier in the process. Although it would then be possible to iterate back to the simplified trade-offs and more detailed design, this loop makes it more difficult to explore wider uncertainty in the pathway itself.

The crux modelling issue for the EnergyPath[TM] Networks Tool is to create a framework that can bridge appropriately the various dimensions of spatial and temporal granularity and multiple trade-offs across different parts of the energy system (in particular network build and building reinforcement) over the full pathway to 2050. This is particularly important given the long-lead times for larger-scale infrastructure upgrade and development. As it is likely to be intractable to hold significant detail on all of these aspects simultaneously, it is important that the end-user can easily flex the level of detail in different parts of the tool and understand the impact on the pathway design.

Although the inputs to the pathway optimisation process are likely to be simplified across one or more of the dimensions it is important that the base input data (particularly for the understanding of building energy demand and networks) are as detailed as possible. Hence in contrast to UrbEn and some other models the problem is inverted and a detailed assessment of possible options (e.g. via network flow modelling) is undertaken *first* and this is used to inform the options in the pathway optimisation. The complexity in this approach is clearly in assessing a wide spectrum of appropriate network / building options in detail first.

To enable the above, the tool has drawn on a number of well-established modelling practices in key areas such as network flow modelling, dynamic building energy simulation, and pathway

---

[7] http://www3.imperial.ac.uk/urbanenergysystems

optimisation. These have been used to increase both the sophistication of the tool and make its development as cost-effective as possible.

Finally, once the more simplified representations have been explored within the pathway under a wide range of conditions and their resilience explored, it is important to be able to loop back and explore some of the key features of the proposed pathway (such as the network development) in more detail; via the same tools and processes that created the detailed representation initially.

1.4.   Purpose and structure of this document

The purpose of this functional specification is to document the properties and behaviour of the EnergyPath[TM] Networks Tool in sufficient detail that a software developer could develop the requisite code, which has now occurred up to version 2.1 of the model, and a user can understand what tasks the tool performs. As such it details the data, calculation processes or logic, controls and interfaces the tool employs to support the expected outcomes. It is structured as follows:

- ► Sections 2 and 3 provide an overview of the design requirements and the tool itself

- ► Section 4 discusses how the tool would be used as part of a wider set of business processes to support the objectives of the end-users

- ► Sections 5 to 8 outline in more detail the specification of the requirements, logic steps and implementation of the individual component modules of the overarching tool:

    - –   The Household Options Module (HOM)

    - – The Spatial Analysis Module (SAM)

    - – The Networks Analysis Module (NAM); and

    - – The Pathway Optimiser Module (POM)

- ► Sections 9 consolidates the discussion with respect to the key areas of uncertainty across the tool (both data and modelling approach)

- ► Section 10 provides a set of concluding remarks

1.5.   Glossary of key terms

Table 1-3   Glossary of key terms

Term	Description
Area	The overall spatial area considered by the tool within its analysis, such as a Local Authority
Capital costs	The investment costs for new technology options, the lump sum investment is converted into a stream of annualized costs (at a given discount rate) over the economic lifetime of the investment
Cluster	The spatial building blocks seen by the POM, an individual cluster is comprised of a number of unique zones. All clusters together represent the whole area under consideration.

Constraint	A restriction on the values variables can take within the optimisation process
Correlation	The degree of dependence in the values of one factor due to another. Within the tool this focused on the degree of correlation in input factors simulated from different distributions.
Database	A comprehensive collection of related data organized for convenient access
Dynamic building simulation	A model of the operation an individual building across the day or multiple days to understand the dynamic evolution of heating supply and thermal loss / gain
Energy services	The requirement for space heating, hot water, lighting, transport, etc as opposed to the energy used to deliver these service
Energy vector	The energy type used to deliver the above services including coal, oil, gas, electricity, biomass, hydrogen, etc.
Load flow modelling	Modelling of the steady-state operation of energy networks to understand the need for control and/or expansion of the system. For example, for electricity it is used to calculate the voltage drop on each feeder, the voltage at each bus, and the power flow in all branch and feeder circuits.
Operating costs	The costs associated with the use of a technology over the year after it has been built. These can be divided into fixed operation costs (FOM), which must be incurred such as annual maintenance versus Variable Operating Costs (VOM) which are affected by the scale of operation, for example due to additional 'wear and tear' under very high operation.
Optimisation	This process aims to maximise or minimise some objective (e.g. profit or cost, respectively) given available variables.
Pathway	The combined timeperiods considered from now until the final modelled timeperiod (e.g. 2015, 2020 … 2050)
Probabilistic simulation	Repeatedly undertaking the same core modelling process (i.e. the optimisation) with different sets of input data that have been generated as part of a Monte Carlo simulation process
Timeperiod	A block of years considered within the tool, within which the timeperiod represents the mid-point. For example, if the pathway is represented as 5-yearly steps 2020 may represent the period from 2018-2022.
Timeslice	Within year disaggregation of time by characteristic days and within day period as part of supply/demand balancing and representation of peak demand requirements. The combination of the two allows for the representation of e.g. a typical winter overnight period or an extreme winter early evening period.
Vintage	The year within which a technology/network/building option has been constructed, which is separate from the year in which it is operating. For example, the timeperiod for 2030 may contain a mix of plant which has been built in 2030 itself, as well as those built in 2020 and 2025 which are still within their technical lifetime. The earlier vintages of the same plant type may have slightly different characteristics such as lower efficiency.
Variable	A choice within the optimisation process as part of maximising or minimising the objective function. The freedom to change the value of the variables can be restricted by constraints
Zone	The smallest set contiguous spatial building blocks within the tool. Zones represent collections of buildings aggregated to a single road node. Groups of contiguous zones form clusters and groups of contiguous clusters represent the area as a whole.

1.6. Acronyms

Table 1-4   Acronyms and associated elaborations

Acronym	Elaboration
API	Application Programming Interface
ASHP	Air Source Heat Pump
BaU	Business as Usual

CAPEX	Capital Expenditure
CHP	Combined Heat and Power
DCLG	Department of Communities and Local Government
DEC	Display Energy Certificate
DECC	Department of Energy and Climate Change
DH	District Heating
DHN	District Heating Network
DNO	Distribution Network Operator
EHS	English Housing Survey
EPC	Energy Performance Certificate
ESC	Energy Systems Catapult
ESME	Energy System Modelling Environment
EST	Energy Savings Trust
EV	Electric Vehicle
FEED	Front End Engineering Design
FOM	Fixed Operating and Maintenance (costs)
GIS	Geographical Information System
GOR	Government Office Region
GSHP	Ground Source Heat Pump
GUI	Graphical User Interface
HLR	Heat Loss Rate
HOM	Household Options Module
HV	High Voltage (network)
LA	Local Authority
LiW	Living in Wales (survey)
LLSOA	Lower Layer Super Output Area
LP	Linear Program
LV	Low Voltage (network)
MIP	Mixed Integer Program
MLSOA	Middle Layer Super Output Area
NAM	Network Analysis Module
NPV	Net Present Value
OS	Ordnance Survey
OTEoEH	Optimising Thermal Efficiency of Existing Housing (ETI project)
POM	Pathway Optimisation Module
PV	PhotoVoltaics
SAM	Spatial Analysis Module
SAP	Standard Assessment Procedure
SBEM	Simplified Building Energy Model

SHCS	Scottish House Condition Survey
SWI	Solid Wall Insulation
TM	Thermal Mass
UPRN	Unique Property Reference Number
VOA	Valuation Office Agency
VOM	Variable Operating and Maintenance (costs)

2. Overarching design requirements

2.1. Key design principles

Key design principles for the EnergyPath^{TM} Networks Tool were tested and refined with the ETI and key stakeholders through a series of workshops and discussions. They have guided the definition of the structure and development of the tool, and are listed below.

- **The nature of the EnergyPath^{TM} Networks Tool:** the tool's overall objective when helping to construct a local area energy system pathway is to minimise energy system 'resource costs[8]', not to maximise social welfare or profit.

- **Co-optimisation of energy system choices:** building, technology and network choices are be co-optimised. However, it is important that the user has full flexibility over the choices that are controllable versus uncontrollable such that tool be applied to tackle a wide variety of possible questions.

- **Analysis of uncertainty:** the model captures uncertainty in inputs probabilistically, and facilitate exploration of their impact, but not attempt to produce optimal hedged positions via formal stochastic optimization. The tool is designed to help the user understand the resilience of particular pathway or deployment options.

- **Required level of resolution:** the tool is designed around an overarching vision of "data-driven" flexibility with respect to spatial and temporal resolution, subject to specific maximum limits in complexity within the networks and spatial modules. The user has the ability to aggregate up time periods or geographical coverage, or drill down to greater granularity, to extent the data allows it.

- **Accounting for consumer behaviour:** consumers are not modelled as entities within the model itself. Rather, the flexibility to test the impact of consumer preferences and behaviours through exogenous input assumptions or constraints is included. This can be divided into two main areas

  – *Consumer technology preferences* in the selection of building options such as insulation retrofits or new heating systems. Whilst the default approach assumes rationale economic decision making by consumers the tool can still test the implications of different consumer preferences on pathway designs. For example, this could be via adjusting the costs of discount rates applied to different technologies to monetise the 'hassle costs' or force in a given expected deployment of certain options.

  – *Energy using behaviour* as above whilst the default assumption is economically rational, for example the use of storage within in a building to minimse overall

---

[8] With the ability to include other monetised 'social costs' where appropriate such as air pollutant damages

ETI – SSH - EnergyPath^{TM} Networks Tool – Functional Specification   20/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

energy use/costs whilst maintaining comfort levels, it is possible for the tool to test the impact on different demand profiles on pathway design.

▶ **Linking to national level boundary conditions:** the primary links to the national level are through price of carbon and other nationally delivered resources, such as centrally generated electricity and natural gas. This retains flexibility in choices at the local level, but takes into account the associated cost. The pathway optimiser design enables the user to impose additional constraints (such as a local area carbon emissions target) if required.

▶ **Flexibility in optimisation problem granularity:** the tool supports optimisation at different temporal and locational resolutions to provide control over the computational effort required for solving.

▶ **Flexibility in defining cluster boundaries:** the tool supports manual intervention to override clustering decisions or set thresholds or parameters used in the clustering algorithms.

▶ **Flexibility in accommodating data:** the tool maximises the use of existing data and a number of elements of anticipated future data, which are likely to be available in some, but not all LAs.

▶ **Post processing integration of other data sets:** the tool should support the layering on to the results of other data (potentially in a GIS) to support end-user requirements, for example combining energy system cost with a socioeconomic datasets to explore the implications of a blueprint design on fuel poverty

Strategic and 'free market' investments

As mentioned above the overall tool framework is designed to be highly flexible, so that the user can apply it to explore a wide variety of questions. This means that some automated parts of the tool, such as the optimisation processes, are easy to configure such that the user can fix or control these aspects or leave them 'uncontrolled', i.e. via defined constraints.

For example, as an extension of accounting for consumer behaviour in the previous section, it is important within the analysis undertaken by the EnergyPath^{TM} Networks Tool to be able to understand the impact of wider 'free market' investments on more strategic investments, such as network infrastructure development. For example, consumer preferences may mean that the uptake of new heating systems does not follow an economically rational deployment pathway. A much longer continuation of gas boiler use may be followed by a rapid uptake of alternatives, or conversely the alternatives may take market share more rapidly than expected.

Given that key local area energy system stakeholders such as LA and DNOs have more limited control over these wider 'free market' investments it is important to understand how sensitive or resilient their strategic investments are to unexpected investments. For example, how is the decision to build a heat network affected if its cost-effective deployment requires a large number of buildings to connect outside of the scope of those managed or owned by the LA themselves.

From a practical user perspective within the pathway analysis functionality (see section 8 for further details), this means that instead of letting the tool cost-optimise *all* available choices this should be limited to network decisions only. As part of the analysis process the user could define a number of fixed levels of building upgrades for different heating systems (i.e. pre-specifying the consumer uptake preferences to be tested rather than letting them be cost-optimised) and explore to what extent the network upgrades change under these sensitivities.

2.2.   Other general requirements

The design and development of the tool have also taken into account a range of broader requirements, noting that in some cases a balance was struck between competing objectives.

- ▶ **Retention of IP by ETI/ESC:** in general, it would be preferable for all IP to be retained by ETI/ESC so that it has full capability to maintain and modify the tool

- ▶ **Minimising software or data licensing costs for end users:** this broadens the appeal of the tool to potential customers

- ▶ **Effective user interface:** this enables expert users to configure and apply the tool in an efficient manner and broadens the tool's appeal as part of longer term development, however, the UI design has to be balanced against the need to deliver end-users access to the full functionality of the tool.

- ▶ **Minimised overall development costs:** clearly this was in the ETI/ESC's interests, but competes with the desire to retain IP to the extent that it requires in house development of functionality that might otherwise be provided by third party solutions

The ETI counselled a pragmatic approach in balancing these agendas, setting cost effectiveness of the overall tool development as the guiding principle for design decisions, whilst bearing in mind the requirements other potential direct and indirect users. As a result this led, for example, to less focus on the user interface under the expectation that only expert users within the ETI/ESC are likely to use the tool in the near-medium term.

3. Overview of tool

The EnergyPath[TM] Networks Tool can be divided conceptually into four main **modules** as shown in Figure 3-1 below, which include:

▶ Household Options Module (HOM)

▶ Spatial Analysis Module (SAM)

▶ Networks Analysis Module (NAM)

▶ Pathway Optimisation Module (POM)

A high-level overview of each module and their interactions across the tool is provided below, with a more detailed description in sections 5 to 8. The full logic / process flow diagram is shown in the Appendix (section 11).

Figure 3-1   Overview of the EnergyPath[TM] Networks Tool
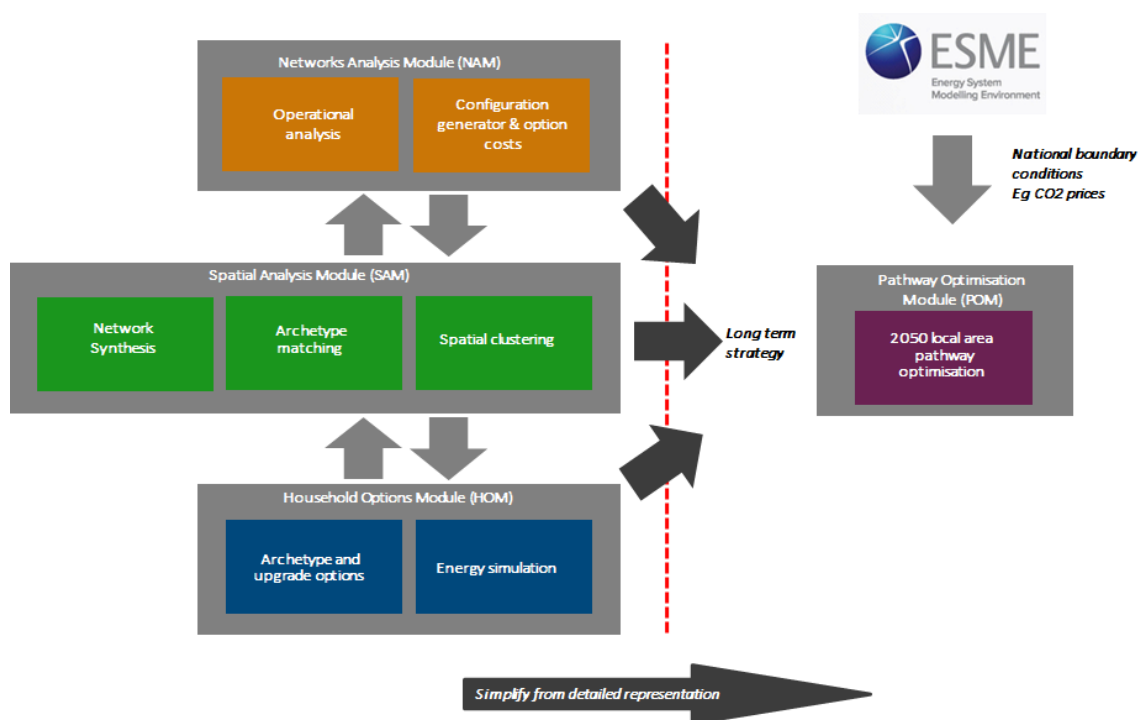
The primary role of the **HOM** is to generate a database of domestic building archetypes that could be used to represent the existing building stock in an area and the options for evolving this stock over time. The archetypes will be specified by, amongst others, the physical characteristics of the buildings and the heating systems they use (including the resulting profile of final energy demand from the operation of the heating devices), the climate in which the building is located (as identified by the Government Office Region) and the costs of converting/upgrading one archetype into another. However, the framework allows for better primary data to be entered as it becomes available. Consumers will not be modelled as entities within the tool itself. Rather, the flexibility to test the impact of consumer preferences for new technology options

and energy using behaviours through exogenous input assumptions or constraints will be included.

The primary role of the **SAM** is to create a detailed spatial (GIS-based) representation of the existing local area including key topographical features (roads, nodes, etc), existing energy networks, building locations and other energy system features such as embedded generation. It will automatically assign an archetype (from HOM) to each domestic building from a detailed pre-defined list by matching based on available statistical datasets at different levels of spatial aggregation, which are provided as an input to EPN. The user can manually enter known information (e.g. which buildings have already received insulation upgrades) and this is taken into account in the matching process. A similar, but simpler matching process is applied to non-domestic buildings, where typical non-domestic activity classes with corresponding benchmark energy and cost values[9] are mapped to individual buildings. The SAM also identifies future spatial considerations for the energy system such as the locations where new embedded generation or heat network supply sources could be sited or allowing users to specify assumptions on the location of new housing developments. The SAM passes spatial information to the NAM to support its analysis of network reinforcement and new build options. The SAM also passes back the results of the domestic building matching process to the HOM, so that a more refined set of future pathways for evolving the stock can be created (i.e. only considering the types of buildings in the area being analysed).

The primary role of the **NAM** is to use data on energy use in buildings, existing networks, and geographic topology to design and cost different configurations of energy networks, for each energy vector (electricity, heat, gas and hydrogen[10]). It outputs spatially varying cost functions for different configurations of each network, which can be used by the POM to compare network build costs and peak load capacity, as part of seeking a cost-optimised pathway.

The SAM and the NAM interact with each other to define a more aggregated spatial representation of the local area, by dividing it into a contiguous set of interconnected *clusters*, chosen to provide a reasonable representation of the costs of network options across the different vectors (i.e. by aiming to not bias the data towards one network type or another). This aggregation is necessary to make the final optimization problem in the POM computationally tractable, but means that a cluster may, for example, represent a group of streets rather than individual streets. Spatial detail is lost at this stage, but the EnergyPath™ Networks Tool is designed to be flexible to allow the definition of the clusters to be changed easily (e.g. adding more clusters to improve the granularity).

Within the tool, the **POM** is the engine that constructs a specific pathway. It receives data from the NAM, SAM and HOM on the representation of the local area energy system and potential future options for evolving this. The POM then takes a more aggregated representation of these options and calculates the choices which produce the lowest cost combination to create a pathway that satisfies a range of constraints and design standards (such as satisfying all householders energy service demands under a range of edge cases).

---

[9] In contrast to domestic buildings in the HOM, the associated non-domestic energy consumption data, costs and possible future pathways for conversion must all be specified exogenously by the user rather than being generated automatically.
[10] Via repurposing of the gas network as opposed to dedicated new build distribution networks

The national pathway produced by ESME is used to inform some of the **boundary conditions** within the POM, such as carbon or fossil fuel prices, and the availability, price and carbon intensity of transmission-connected electricity. The POM's outputs enable the user to investigate the features of the pathway and key areas of uncertainty, via the possibility to undertake a Monte Carlo simulation of key inputs at the point they are used within the POM.

*It is important to **separate** the way the tool is conceptually constructed, relying on techniques such as mathematical optimisation and Monte Carlo analysis, from the process of developing insights to inform more subjective strategic decision making. The following sections describe the separate conceptual elements of the tool in more detail, whereas the wider process of applying the tool is discussed in 4.*

4. Use of the EnergyPath™ Networks Tool

4.1. Business process for using tool

4.1.1. Overview

The tool will be sufficiently flexible to accommodate multiple modes of use, which may organically emerge through practical application. For the purpose of illustration however, this section outlines one such process by which an energy system transition plan within a LA area might be developed and implemented. The role of the EnergyPath™ Networks Tool within this process is to provide analysis and insight to decision makers to allow them to make strategic choices on the content and execution of the transition plan. The example process is considered at three levels of resolution.

▶ The **top level** end-to-end process, from the earliest consideration of a transition plan through to execution of projects, and involving the interaction and cooperation of multiple local stakeholders, most particularly the LA and energy distribution network operators.

▶ An **intermediate level** concerning specifically the core activities surrounding deployment of the EnergyPath™ Networks Tool, occurring within the LA and with activities likely led by a team tasked with managing the development of the transition plan.

▶ The **process of analysis, iteration, and refinement** through which specialist individuals tasked with operating the tool would step in order to produce an initial transition plan (including near term projects) for validation and consultation at escalating levels of authority. This would be the most intense phase of tool operation, though its ongoing use would in practice be likely to continue in a manner intermingled with the broader processes. These tasks are unlikely to occur in a completely linear and isolated fashion.

These levels are illustrated in the diagram below.

Figure 4-1   **Overview of activities in the planning and implementation of a local energy transition pathway**
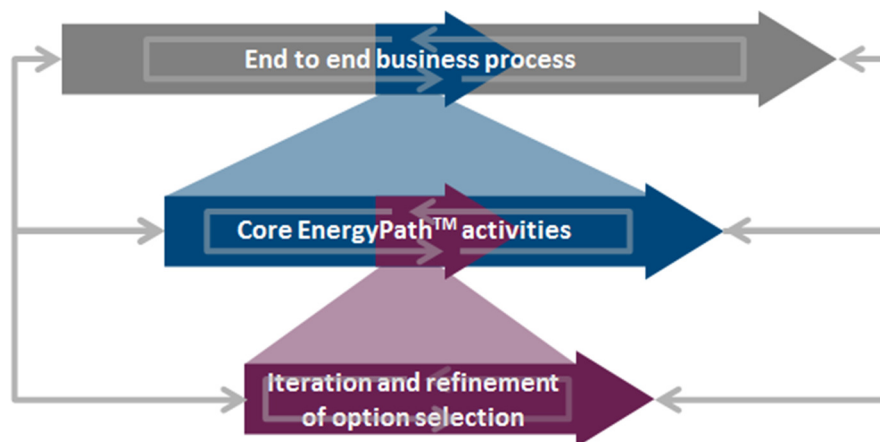
In this section we will refer to the transition plan, describing the plan for the low carbon transition of a local energy system, as being the final product of the end to end business process. This plan would comprise a number of components.

▶ A defined programme of projects and measures selected for implementation in the near term

▶ An indication of the main features and alternatives of pathways for progressing transition in the period beyond the implementation of the near term project programme, where uncertainty is too great to allow confident selection and definition of low regret projects; and

▶ A strategy for monitoring, maintaining and improving the transition plan over time, as time elapses and new data becomes available or is acquired, enabling selection and definition of further projects to facilitate the transition.

4.1.2. End to end process

The full process through which an LA would step, from setting the objectives and targets for a local energy transition plan, through to implementing or initiating the projects which will start to deliver it, is set out in the diagram below.

Figure 4-2   **End to end business process for design and implementation of a local energy transition pathway**
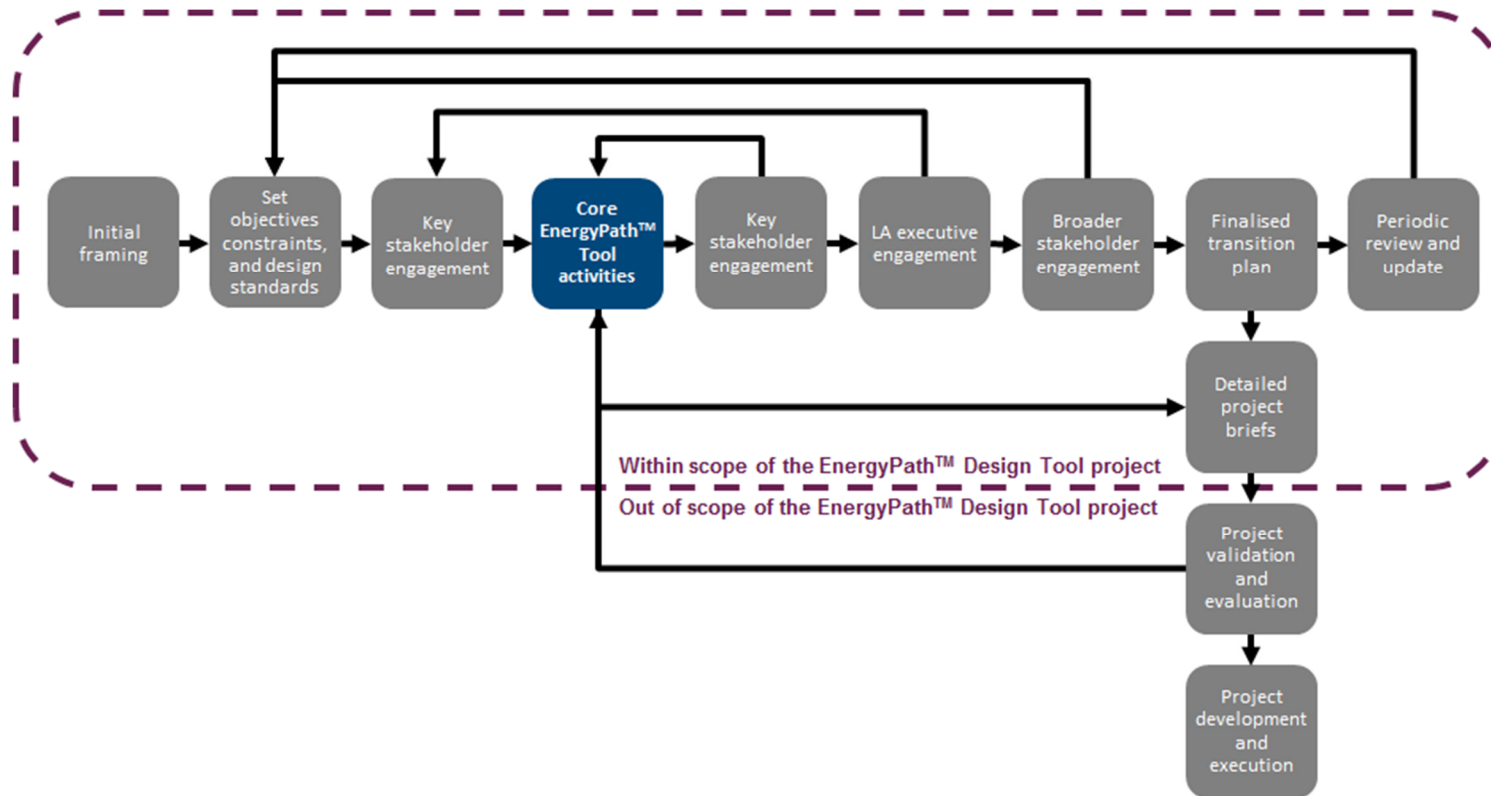
- ▶ **Initial framing:** this step describes activities to provide policy and decision makers with an understanding at a high level of the possible pathways that exist for an area, what particular technologies or options are likely to be viable, and at what range of costs. This will inform the objective and target setting process, and could potentially be informed by a version of the EnergyPath™ Networks Tool based on generic local areas (e.g. typical rural, suburban, urban).

- ▶ **Set objectives, targets and design standards:** an objective for reduction in carbon emissions may be determined based on an optimal allocation of national level targets between Local Authorities, or via imposition of a carbon price. This might itself be informed through use of the EnergyPath™ Networks Tool by a central agency. At the local level, the tool will determine cost effective pathways in this context, and support development of the LA's transition plan. Some LAs may wish to set further targets for the transition plan reflective of their own priorities, most probably concerning social factors. In particular, metrics describing fuel poverty or the level and stability of energy bills might be targeted. Additionally, design standards which viable pathways must accommodate will need to be set, as discussed in section 8.5.3. This in particular will regard inputs to the tool which cannot be forecast or modelled in a reliable manner, but have a direct and significant impact on local energy systems. Some standards may be externally specified by industry norms or regulations, such as the 1-in-20 winter standard already discussed. Other design standards to be incorporated might need to be determined internally by the LA, such as assumptions on demographic change, and how it affects energy usage profiles and building occupancy patterns.

- ▶ **Initial key stakeholder engagement:** it will be necessary to engage at an early stage with key stakeholders in local energy provision, which will play a pivotal role in implementing the final transition plan. Most obviously, this will include the owners and operators of the distribution networks for electricity, gas and if applicable, heat. It may also entail large housing providers or the owners or occupiers of significant campuses or installations in the local area. The engagement would involve securing their cooperation in the process, and agreement on the high-level objectives and constraints set.

- ▶ **Core EnergyPath™ Networks Tool activities:** this step entails preparing the tool, using it to identify viable pathways, discerning a set of projects for implementation in the near term consistent with those long term pathways, and securing support at a certain level of authority within the LA. This is described in greater detail in section 4.1.3.

- ▶ **Further key stakeholder engagement:** given the essential role of certain stakeholders in implementing any project or long term transition plan, as discussed, seeking their input on the feasibility and merit of a transition plan and specific projects would be an essential first step after securing internal LA non-executive approval. In practice continuous interaction with such parties is likely to be desirable. Feedback received may necessitate revision and iteration of the tool to accommodate new information or constraints.

- ▶ **LA executive engagement:** once a transition plan, including the specific near term projects, has been determined to the satisfaction of the key stakeholders, it is anticipated that executive level agreement within the LA would be sought. This would allow the main decision and policy makers within the LA to satisfy themselves that the

plan is consistent with the high level objectives and targets established. Objections at this stage may necessitate a greater degree of revision to the configuration of the tool, such that it would be necessary to re-engage with the key stakeholders to confirm the feasibility of proposed changes.

▶ **Broader stakeholder engagement:** executive approval of a transition plan is assumed to be the precursor to broader stakeholder engagement. This would involve consultation with the full range of local stakeholders (as identified in section 1.2.2) involved with or impacted by implementation, including local enterprises, community groups and individuals. Communication would encompass the specific projects for implementation in the nearer term, and the longer term element of the transition plan. Evidence of a substantial divergence of public or civic opinion from the parameters assumed in the initial establishment of the tool could entail recasting of the objectives or targets determining the pathway optimisation calculation in the tool, and a significant iteration.

▶ **Finalised transition plan:** approval at each level of consultation and engagement would result in the finalised transition plan. This will include the set of projects for near term implementation, and the longer term plan for monitoring and managing progress and planning further interventions to deliver the long term transition to decarbonised local energy provision. Some projects may be implemented directly by the LA and some by key stakeholders with which the LA has an established relationship (such as network operators). A large proportion will depend on decisions of private and independent individuals or organisations, for which the transition plan would lay out a strategy of education, engagement and incentivisation.

▶ **Periodic review and update:** once the finalised transition plan is obtained and implementation of near term projects has begun, long term management of the transition plan will begin. This will incorporate a range of activities, monitoring availability or planning acquisition of new relevant data regarding, for example, local infrastructure, changing local demography, the evolution of technology costs and fuel prices, developments in national level policy and so on. There will also likely be a role for regular review of progress against the plan and redetermination of viable pathways in the light of updated information and completion of projects. This will enable identification and specification of subsequent projects for implementation.

The finalised transition plan will identify a set of discrete projects which are to be implemented in the near term.

▶ **Detailed project briefs:** the tool will be used at its highest level of temporal and spatial resolution to determine detailed technical specifications and outputs for financial projections. These will provide the basis of project briefs which will launch the process of implementation. In the case of network projects, this will include detailed outputs from the NAM module.

▶ **Project validation and evaluation:** with a detailed specification of the project in place, an LA would likely perform an initial validation exercise, before committing large investment to implementation. This would include commissioning qualified technical expertise to validate cost assumptions and technical specifications. Financial information provided by the project brief would provide inputs to a financial and

business modelling package, being prepared by the ETI/ESC as part of a separate work stream. This would facilitate determination of the economic feasibility of the project and inform choices on the sources of capital and commercial strategy employed.

▶ **Project development and execution:** having scoped a project in detail, and investigated the potential financial and commercial models, the development and execution of the projects would commence. It is recognised that in a large number, or majority of cases, this will not fall within the direct remit of the LA itself, but in that of other stakeholders in the local energy system, as identified in section 1.2.2. This phase will then necessarily involve deep cooperation with these parties. For projects requiring physical intervention, this phase will commence with the commission of Front End Engineering Design (FEED) studies to provide the basis for final design, procurement and construction. For many projects, final implementation might be at the level of private individuals or businesses. In this case LA intervention might be through establishing or reshaping policy incentives or providing information to guide private investment for example.

4.1.3. Core EnergyPath™ Networks Tool activities

The diagram below illustrates the steps that would occur within the LA to prepare the tool for use, and prepare its outputs for external consultation. The precise mode of operation and allocation of responsibilities will depend on the commercial model chosen by the ETI/ESC.

Figure 4-3   **Overview of core EnergyPath™ Networks Tool activities**
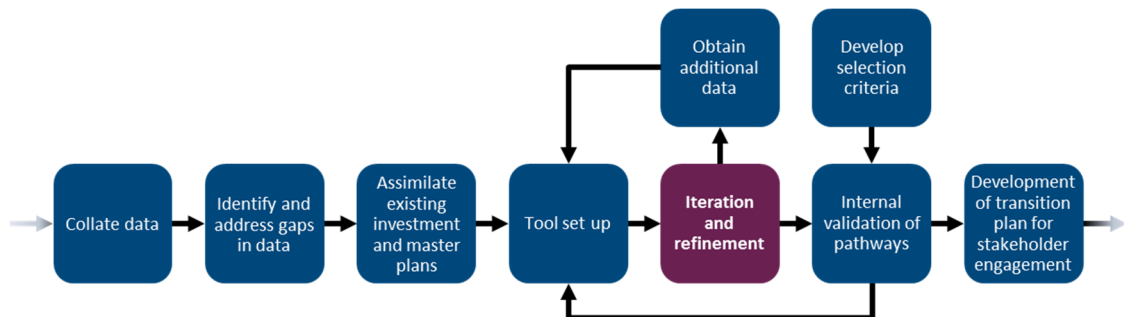
▶ **Collate data:** it will be necessary to collect and prepare relevant data from various sources to allow the tool to undertake analysis of the local area. In part this will be taken from standard national datasets, and in part from databases on local infrastructure held by LAs or other bodies (for example, details of listed buildings). Close cooperation from network operators or owners of other key assets will be necessary to construct an accurate picture of the local infrastructure.

▶ **Identify and address gaps in data:** there may be gaps in the available data that would impair the ability of the tool to identify viable and practicable pathways. Additional data may also be required to enable the tool to consider targets set by the LA regarding social or other local environmental factors such as air quality. To address these needs, the LA may choose to commission data specifically, to be gathered through local surveys or questionnaires.

- **Assimilate existing investment and master plans:** LAs may already have in place projects or schemes to develop or influence energy provision in a local area. For example certain areas may be earmarked for development of energy facilities. Development master plans with a broader scope regarding housing and economic development are also likely to be of relevance for example in identifying areas of future development, or existing areas scheduled for redevelopment or repurposing. Distribution network owners may also have associated plans for upgrade of their infrastructure.

- **Tool set up:** having assembled the relevant input data, it will be necessary to compile and format this for entry into the tool, together with representations of the objectives and constraints that have been set to bound the optimisation calculation.

- **Iteration and refinement:** this is the process of identifying, analysing and filtering options within the EnergyPath™ Networks Tool, as performed by the user, so as to arrive at a set of projects for implementation in the near term, and indication of the high level features of pathways that deliver long term transition consistent with these projects. This activity is discussed in more detail in section 4.1.4.

- **Obtain additional data:** in the process of operating the EnergyPath™ Networks Tool, areas may be identified where gathering additional data would be valuable in refining and reducing uncertainty in the tool's outputs, and aiding the decision making process it supports. This process is described in further detail in section 4.1.4. Where the need and justification for additional data is identified, there would be a process to plan for and undertake its acquisition, and its preparation in a form suitable for the tool's use. It would then be integrated into the tool's configuration.

- **Develop selection criteria:** externally to the operation of the tool, it would be good practice to identify the metrics by which pathways will be judged. This could encompass a broad range of conditions, for example uncertainty on the cost or level of CO_2 savings of the pathway, or the degree of sensitivity to national electricity and fuel prices of the annual cost of energy provision in LA owned housing stock, or the proportion of investment targeted in lower income wards. These may be employed through direct comparison of different pathways, or if quantified limits to the metrics are determined, tests may be formulated that would determine the acceptability or otherwise of a pathway.

- **Internal validation of pathways:** based on the predetermined selection criteria, approval at an intermediate level of authority within the LA for initial projects and consistent long term pathways would be sought. To the extent that these are not found to meet requirements, it would be necessary to iterate once more, recalibrating the optimisation constraints as necessary.

- **Development of transition plan for stakeholder engagement:** at this stage, a set of initial projects, and the higher level options for progressing transition thereafter, would have been approved. These would require development and documentation in to a more formal transition plan ready for use in consultation with key external stakeholders or presentation to the LA executive. This would contain plans for the ongoing

maintenance, management and improvement of the transition plan, and a strategy for the identification and definition of projects for implementation beyond the initial set.

4.1.4. Iteration and refinement of option selection

Due to the computational effort that would be required, it is unlikely to be feasible in one step to obtain a full set of results at the greatest level of spatial and temporal detail and exploring the full range of uncertainty with regards the options. Instead, it is envisaged that temporal resolution, spatial resolution and option uncertainty will be traded off to a certain extent in each operation of the tool, to enable iterative refinement of the solution. The diagram below describes the process the team or individual assigned with operating the tool would follow to identify options, and then specify projects, which will progress the local area transition in a manner consistent with the priorities and constraints set up front. The selected options should also be resilient, such that they would continue to facilitate pathways consistent with the overarching objectives under a range of outcomes regarding uncertainties in the projects themselves (for example, the cost of preparing trenches for the installation of heat network piping), and external factors (for example, electricity prices).

Figure 4-4 Overview of the EnergyPath^{TM} Networks Tool modelling steps as part of the refinement of options for a viable local energy system pathway
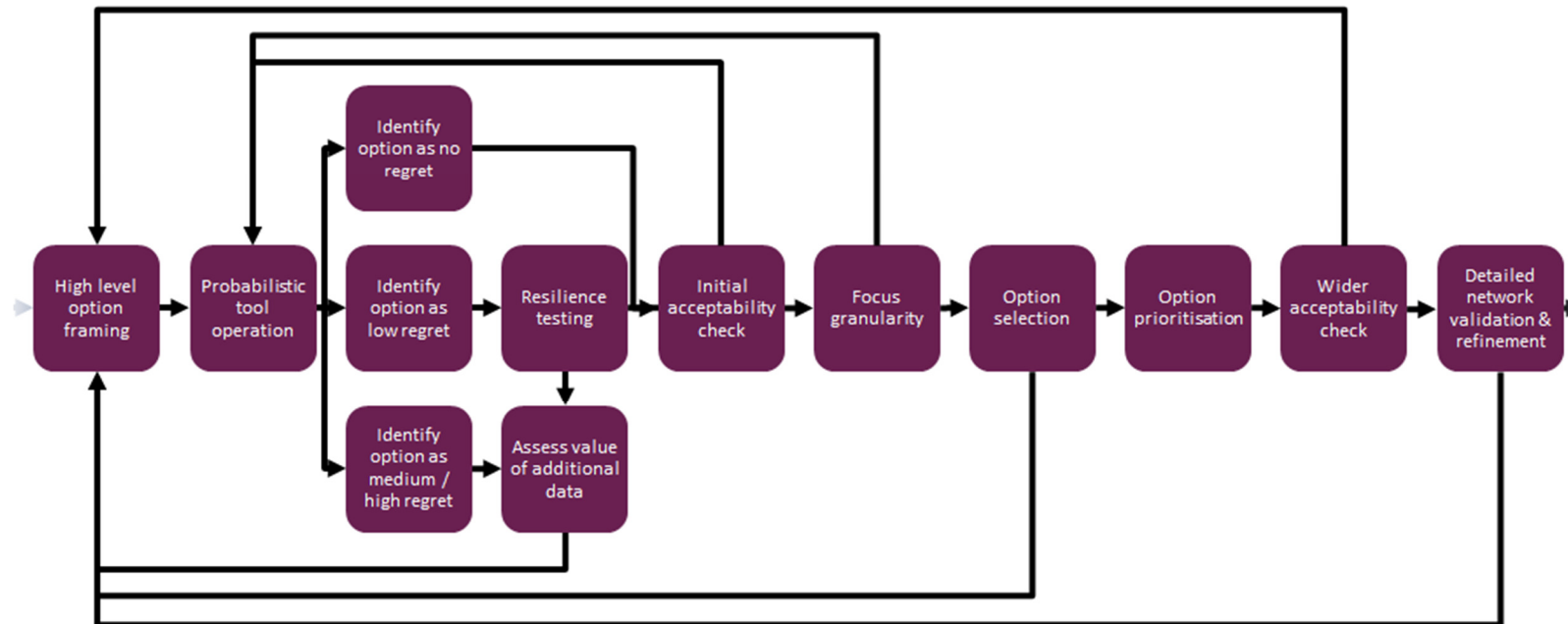
► **High level option framing:** after configuration of the tool, the first step would likely be operation in deterministic mode, under a number of scenarios. The scenarios would be constructed so as to explore a range of internally consistent outcomes, regarding such factors as national electricity, gas and carbon prices, and potentially others such as technology capital costs. Where the tool selects options in a number of scenarios, they are likely to be strong candidates to be carried forward for further investigation. Determination of how exactly to define an option will require judgement, given the scope for different modes of transition in different locations of an LA area, for variation over time, and for interdependence between options. By considering only a relatively small number of scenarios which still consider a wide range of outcomes however, this problem should remain tractable.

► **Probabilistic tool operation:** having identified options meriting further investigation, the next step will be to consider these more specifically under probabilistic operation of the tool, allowing fuller assessment of the uncertainty around the optimal outcomes of the deterministic runs, and the factors driving selection of the option.

► **Identify option as no regret:** some options may represent sound decisions, in that they reduce CO_2 emissions and / or achieve other objectives in a cost effective manner, under all reasonable circumstances. In this case the option would be identified as no regret, and will be selected for future implementation. Examples may include low cost high effectiveness insulation measures.

► **Identify option as low regret:** some options may be optimal under a wide range of circumstances, that is to say they are selected in a large proportion of the simulations generated by the EnergyPath™ Networks Tool in probabilistic operation. These options would be identified as low regret, and would represent strong candidate projects for implementation.

   – **Resilience testing:** the tool user would undertake analysis of the low regret options to determine their resilience. Resilience can be characterised as the "*propensity for an option to remain cost-effective when the tool input assumptions and constraints, in combination or alone, are set so as to produce circumstances that would tend to diminish the cost-effectiveness of the option being considered, or enhance the effectiveness of alternative options*".

   – This may be expressed as a tipping point, especially where an option is sensitive to an input parameter subject to considerable uncertainty. The tipping points are identified through posing questions in the form, for example, of "*what value must occupants of a particular building type attach to visual amenity in order that an air source heat pump is no longer the optimal source of heat provision*?" Having identified this value it would be compared to expectations for the value of visual amenity to inform opinion of the option's resilience. If an option is determined to be resilient it would be selected for more detailed definition and future implementation. If its resilience cannot be determined from the data available, it may be decided that gathering of further specific information may assist decision making (discussed in more detail below).

▶ **Identify medium/high regret options:** some options may appear to be optimal in a smaller range of circumstances. That is to say that in a significant proportion of the simulations generated by the EnergyPath™ Networks Tool in probabilistic operation, the option is not selected. Such options would probably be rejected for implementation in the near term, unless they are shown to be highly sensitive to costs which are known only to a low degree of certainty.

  – **Assess the value of additional information:** in the case that a medium/high regret option is sensitive to costs which are known only to a low degree of certainty, it may be decided that further specific data should be gathered to enable a more confident assessment. This data would be included in the tool's configuration and the process iterated. A similar course may be taken if the resilience of a low regret option cannot be adequately determined without further data.

▶ **Initial acceptability check:** an initial sense check at this stage would filter out options that, although identified as cost optimal by the tool, may in fact be considered untenable for other reasons, potentially revealing that important practical constraints have not been entered in the model's initial configuration, such as the acceptability of.

▶ **Focus granularity:** with promising options selected, it will be desirable to investigate them at a more granular level in the tool; that is to resolve their geographical and temporal properties to a greater detail by iterating the tool whilst focusing on particular periods of time or locations in order to enable more precise definition.

▶ **Option selection:** having been investigated in detail, those options which are considered to be no regret, or optimal under a large range of possible outcomes, have been defined as projects and are selected for implementation. They are fixed in subsequent iterations of the tool, which may alter the propensity of the tool to select other options, allowing them to be selected or rejected, and so on until all resilient low regret options have been identified.

▶ **Option prioritisation:** the iterative process described above may imply a certain ordering or grouping of projects, where their 'optimality' is co-dependent on the implementation of others. Having defined and selected projects to be carried forward, this step would consider how they should be assembled as a package. For example, some projects may only be implemented once others have been. Some projects may have greater uncertainty associated with them, or tend to lock an area in to a certain pathway. It may be desirable to defer implementation of such projects in order to mitigate against the risk that as circumstances evolve they transpire to be high cost. Careful consideration would also be given to the degree of direct control the LA has on the projects, that is whether they are undertaken directly by the LA, or for example are dependent on decisions of private householders.

▶ **Wider acceptability check:** at this stage, more detailed checking will be necessary to ensure that, whilst the projects progressed to internal validation are optimal from a cost perspective, they are compatible with wider LA policy or objectives.

▶ **Detailed network validation and refinement:** a further check will be to ensure that the projects selected are viable given the planned state of local energy networks. In addition, the costs of the potential network options for a given part of the local area could be further refined before incorporation within detailed project briefs. This would occur using the NAM's more detailed analytical capability, as described in section 8.7. *Whilst this step would further enhance the robustness of the projects, it is not intended to be a substitute for the detailed planning and analytical work that individual DNOs would undertake as part of their direct planning and development activities.*

At this stage it is anticipated that a prioritised set of projects, and the high level features of a number of alternative pathways that include the identified near term projects and facilitate the long term transition, would be set and ready for internal validation.

4.2. Key outputs and reporting

As discussed in section 4.1.1, the tool is sufficiently flexible to accommodate different methods of use. This section however illustrates some of the outputs and reporting modes that the tool could support by stepping through a process similar to that envisaged in section 4.1.4, and imagining the questions that might be asked and analysis that might be undertaken at each stage. This is necessarily a simplified example, and in practice an LA area may require a much richer data set, making option selection and project definition more complicated.

High level option framing

The first step envisages deterministic operation of the model under a number of scenarios in order to acquire a high level understanding of the options that could comprise a transition pathway, and the areas and circumstances in which there may be competing pathways. The key at this stage would be the ability to summarise data at a level of detail that provides for ready interpretation and communication of the main distinguishing features of viable pathways. One way in which this may be provided is through a waterfall plot illustrating the contribution of different measures to overall emissions reduction over the pathway. If a particular measure makes a large contribution in each scenario, it may be assumed that it is reasonably resilient to different outcomes and is worth investigating.

In the example below, the user would observe from the waterfall plot that the building efficiency measure makes a strong contribution to the transition pathway in all of the scenarios considered, and therefore merits further investigation. District heating networks are also selected for large scale deployment in most of the scenarios, but in some cases appear to be displaced to some extent by GSHPs. This option also merits further investigation to understand better the circumstances under which it is preferred, and the circumstances under which GSHPs are the lower cost option.

Figure 4-5 **Example output: contribution to emissions reduction, 2014 to 2050, by measure**
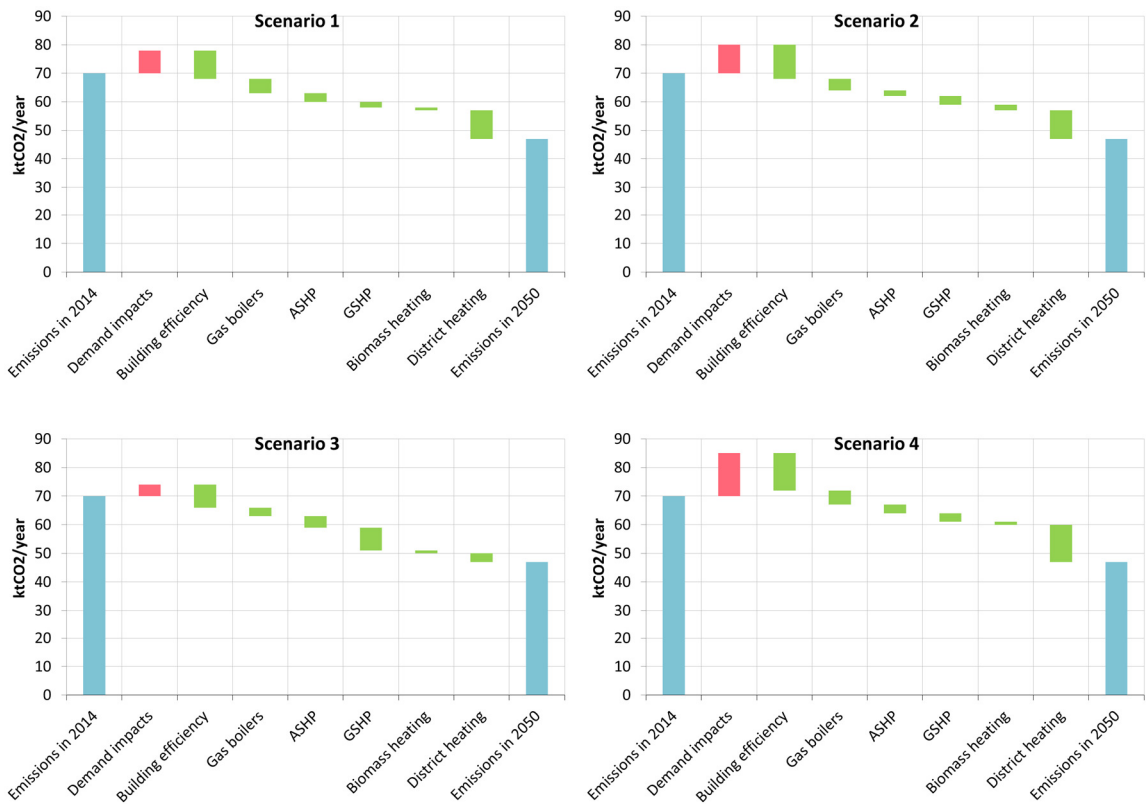
This analysis could be focused on different zones, or aggregations of zones, within an LA area to identify how the pathway may vary geographically. The development of these results over time may also be examined, as in Figure 4-6. These show that deployment of building efficiency measures is made early in all scenarios, reinforcing this option as one of interest. More efficient gas boilers appear to be a useful early option, although longer term their role is reduced. In most scenarios, district heating tends to be deployed at a steady rate over the first decade of the modelled period, whereas GSHPs tend not deployed at scale until the late 2030s.

Figure 4-6 **Example output: deployment by measure over time**
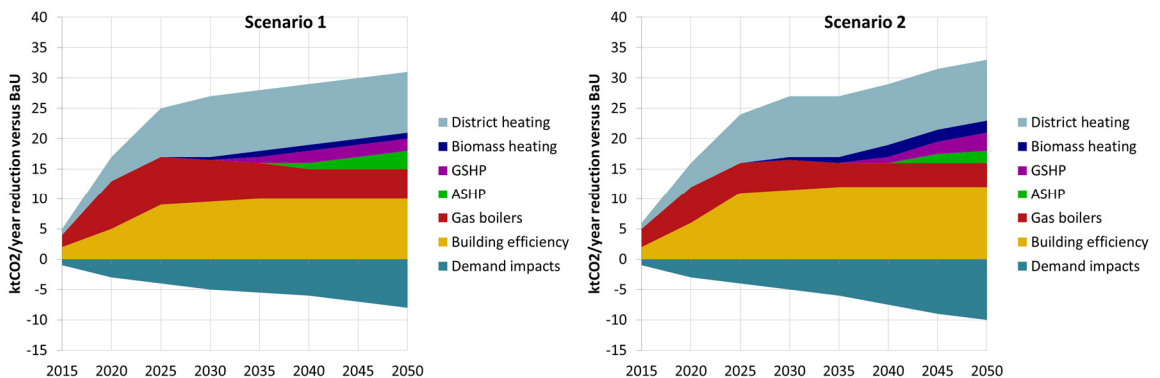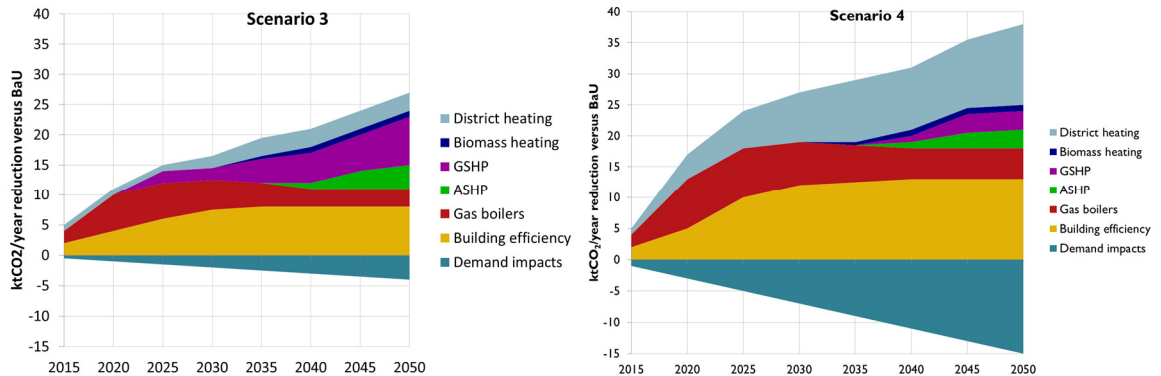
Probabilistic tool operation

Having established that district heating networks and building efficiency appear to be strong options, and that GSHP is favoured under certain circumstances, a more detailed investigation of the uncertainty surrounding these options is appropriate. This is done by introducing a distribution to the inputs and running the tool in probabilistic mode, where the distribution is randomly sampled from over multiple simulations. This will allow the user to build a fuller picture of what factors the options are sensitive to, and how resilient they are to different future outcomes.

Figure 4-7   Example output: CHP capital cost distribution, 2020 and 2050
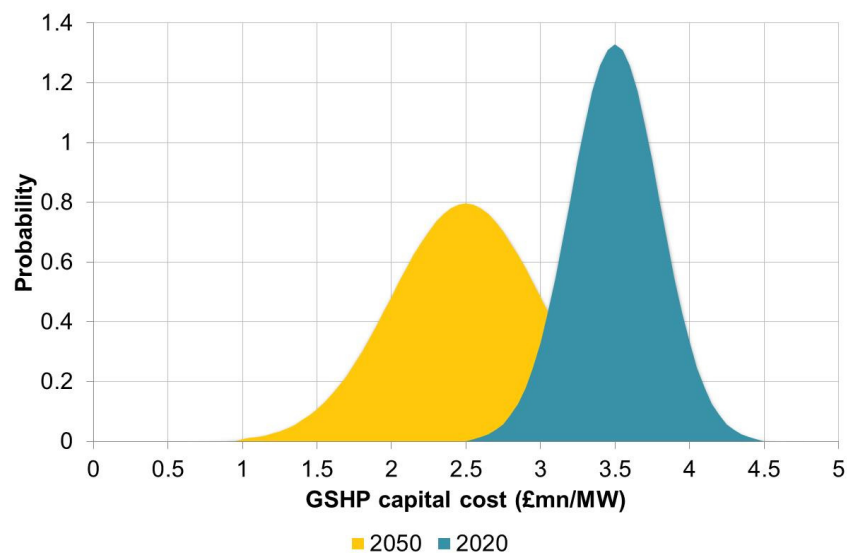
Option appraisal

Probabilistic operation brings a richer set of data to the user, allowing him or her to undertake a better appraisal of the options by understanding the impact of uncertainty on the cost-optimal deployment of these solutions, whilst also meeting the requisite constraints and design standards.

The process in section 4.1.1 envisages categorising options as no regret, low regret, or medium/high regret, initially as part of a manual process of using the tool, though it may be possible to automate elements of this in the future. Options which are chosen by the tool in all simulations will be cost-effective under all reasonable circumstances and provide a no regret route to decarbonisation. That is to say that once implemented, no 'reasonable' future eventuality is likely where an alternative option would have delivered equivalent emissions reduction for lower cost. Options frequently chosen by the tool will be cost-effective under a wide range of future eventualities, and would rarely be regretted.

Probabilistic operation allows options to be identified in this way, starting the process of identifying a package of projects representing a course of action which is robust to changing circumstances. An output which could assist this process is a scatter plot of the absolute reduction in emissions achieved by each measure versus BaU, against the per-unit cost of that reduction, over all simulations. Examining where these points accumulate allows categorisation of measures in to those supporting large scale emissions reduction at low costs (likely to be low regret), those supporting limited emissions reduction at high cost (likely to be high regret), and those in between these poles.

In the example, higher colour density indicates a greater concentration of simulation results in this space. Insulation appears to be no or low regret. It is consistently selected by the tool to provide a large contribution to emissions reduction, and does so at a comparatively low cost. Biomass fired district heating is also consistently chosen by the tool to provide a large contribution to emissions reduction, however it does so at a higher cost relative to insulation. It may be low regret, and as such will require more detailed investigation. GSHPs are selected by the tool to provide a relatively low contribution to emissions reductions in most simulations, but the cost of doing so is subject to significant uncertainty. It may be a higher regret option, but could merit further investigation into costs.

Figure 4-8   Example output: cost effectiveness versus scale for various measures, 2050
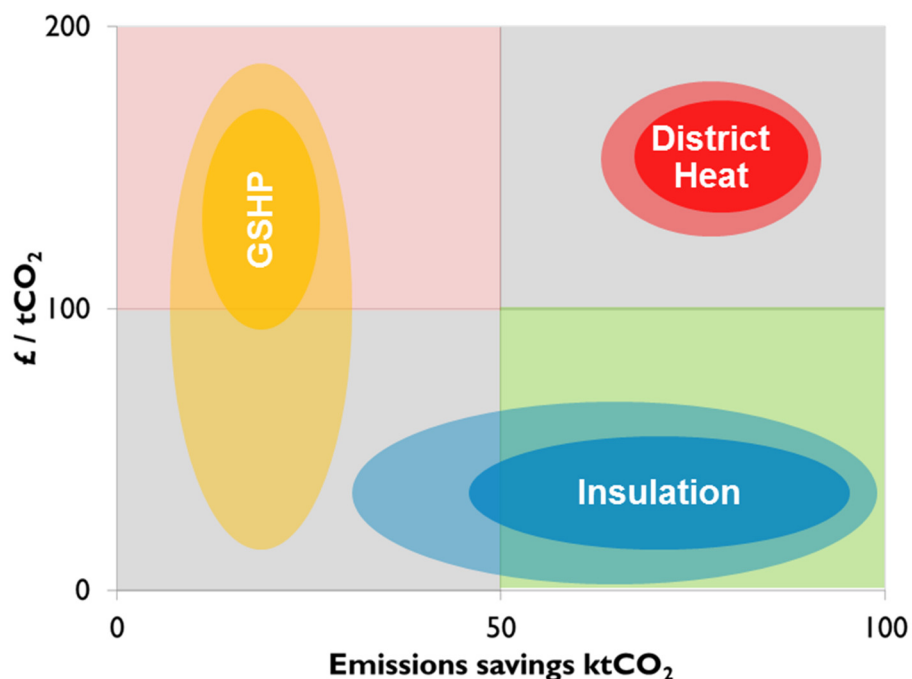
Further geographical analysis reveals that in simulations where GSHP deployment is selected, it is limited to a specific area within the LA, which is likely to have distinctive properties, for example, sufficient building outside space, low electricity network reinforcement costs, or the absence of potential anchor loads for heat networks. Further investigation of this option would be concentrated in this area.

Figure 4-9   Example output: geographical distribution of GSHP deployment
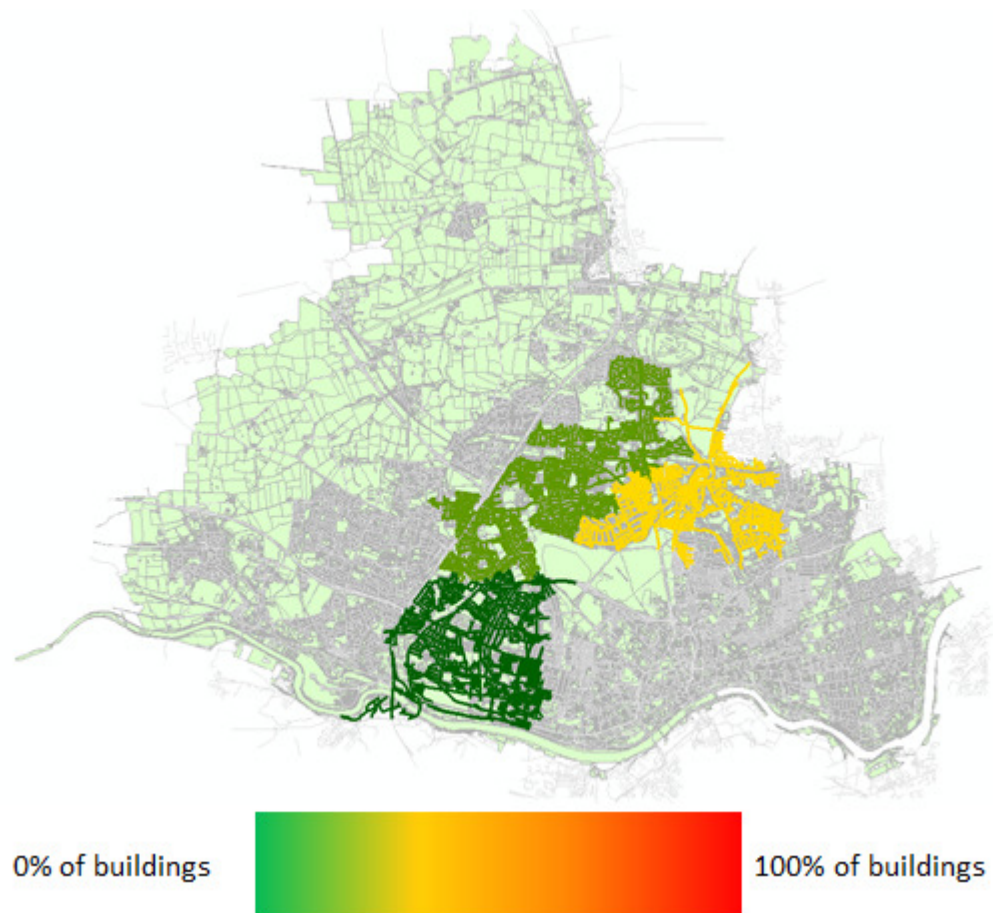
Resilience testing

Where an option is identified as low regret, the user may wish to test and identify more explicitly the circumstances under which it may cease to be the optimal choice, in order to inform decision makers on whether the risk of such a circumstance occurring is one they are prepared to accept. This could be accomplished in part through sensitivity testing, where one variable is flexed whilst others are held constant, in order to determine the prime drivers on the selection or otherwise of particular measures. It may also involve closer examination of results distributions to examine where correlations between factors or patterns exist.

Here, this approach identifies that capital costs is the chief driver of the level of deployment of district heat networks. It could be extended to identify the cost of capital at which a particular heat network ceases to be cost effective. LA decision makers may decide that they are confident that the cost of capital can be held below this level through access to government backed finance or guarantee schemes, and hence they are comfortable with accepting it.

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   41/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.
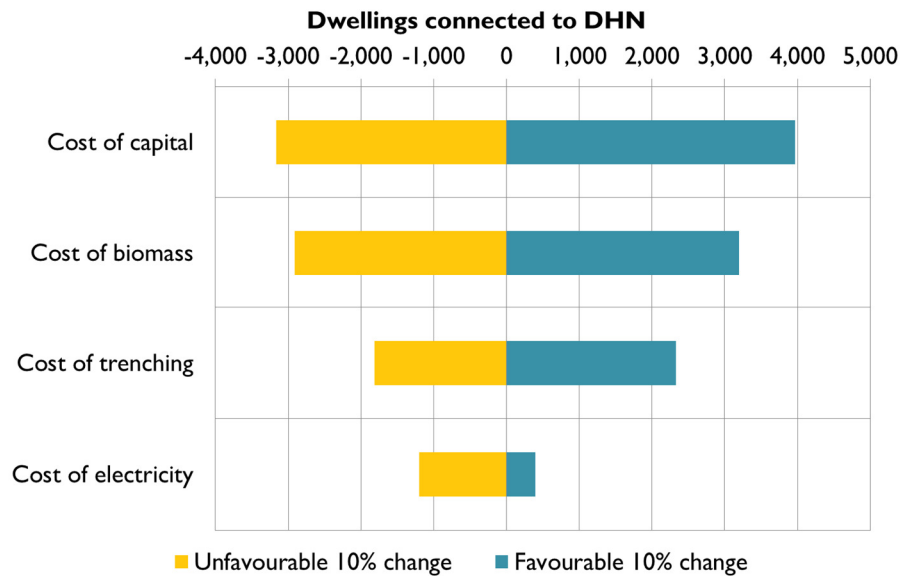
More detailed examination of the distribution of results may also reveal further insights. The scatter chart below represents a subsection of the data from Figure 4-8, and considers the area of the LA where there appears to be close competition for provision of heat between a district heat network based solution and a GSHP based solution. It can be seen that the GSHP solution offers a slightly lower cost per dwelling on average, but provides less certainty on emissions reductions and exposes householders to more risk of very high price outcomes. Further investigation may reveal that the most significant drivers are electricity prices and installation costs, and consumer behavior. The former two factors are beyond the control of the LA. The latter is also beyond the control of the LA and is subject to ongoing uncertainty that will continue through the life of the asset. It may therefore be decided that GSHPs represent a higher regret option.

Figure 4-11   Example output: per dwelling cost versus CO₂ savings

Legend:
- **District heat network solution from one simulation**
- **GSHP solution from one simulation**
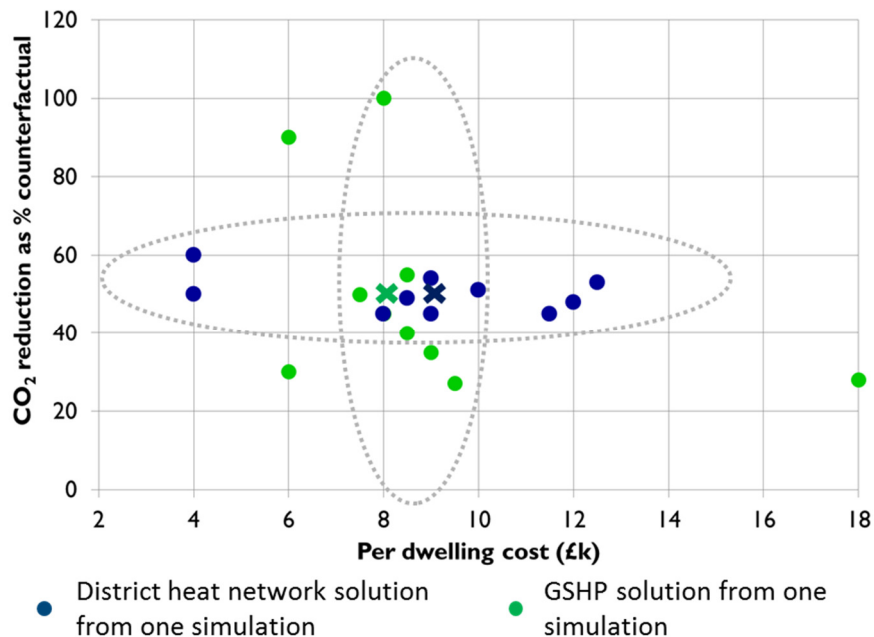
Assess the value of additional information

It may be that the data available to a user is limited initially in some aspects, particularly where it may be dependent on local conditions or where there has been little imperative to gather it prior to consideration of a transition plan. The tool could be used to identify where more data would be valuable in identifying low regret options or rejecting higher regret options, by illustrating how results change when the uncertainty on a particular assumption is reduced or it is assumed to have a slightly higher or lower value.

In the example, it has been determined that whilst GSHPs may provide a lower cost solution in a particular area, but it is subject to risks that the LA may be uncomfortable holding. Before it is rejected the user may wish to confirm that this is a reliable result. Sensitivity analysis showed electricity costs, behavioural factors and installation costs to be the main drivers. Uncertainty around the former two cannot be reduced, but could be for the latter through local ground surveys, for example. If it is assumed that this would reduce the uncertainty on installation costs by 30%, the tool could quantify the extent to which uncertainty on capital costs in this zone is reduced, focusing for example on LA owned properties. This is illustrated in the figure below.

Figure 4-12   Example output: capex on GSHP in LA owned properties
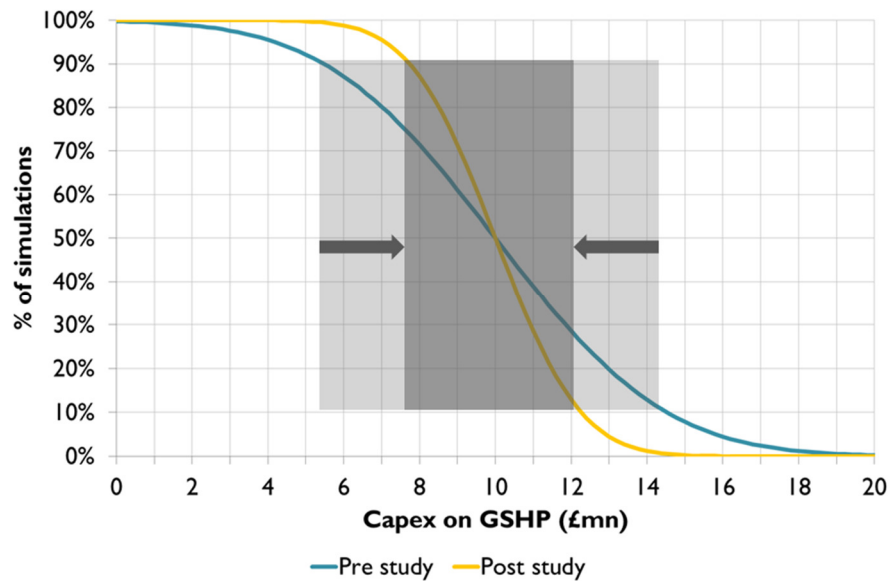
The result might also be calculated when reduced uncertainty and a slightly higher or lower mean value is assumed. It might be found that a slight higher mean installation cost has very little effect on the deployment of GSHPs, but that a slightly lower mean has a positive impact, as illustrated in the figure below. Under these circumstances it might be decided to commission a study in the area identified in Figure 4-9 to acquire a better data set on which to make the decision.

Figure 4-13   Illustrative impact of study on input parameter uncertainty
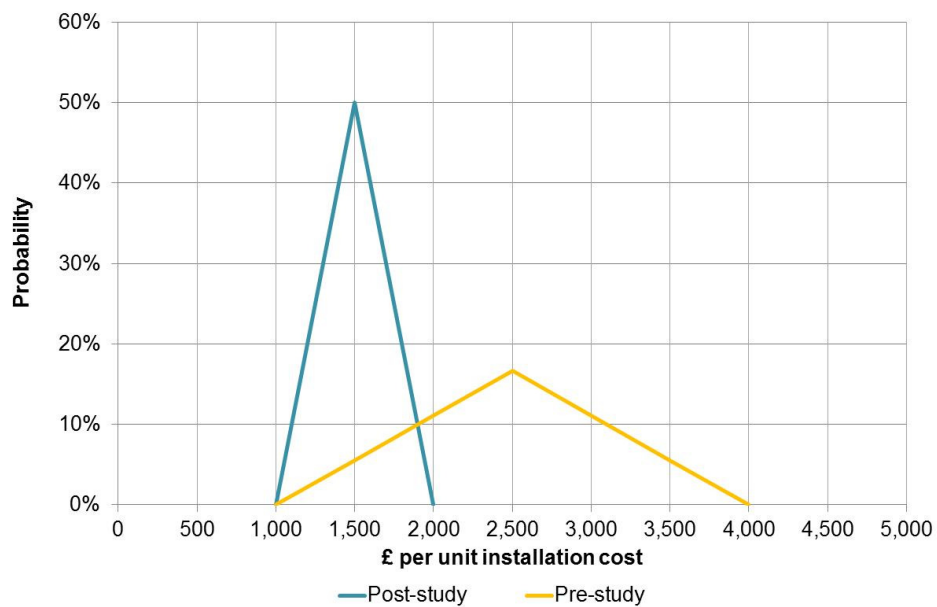
Figure 4-14   Example output: deployment of GSHP pre- and post-study
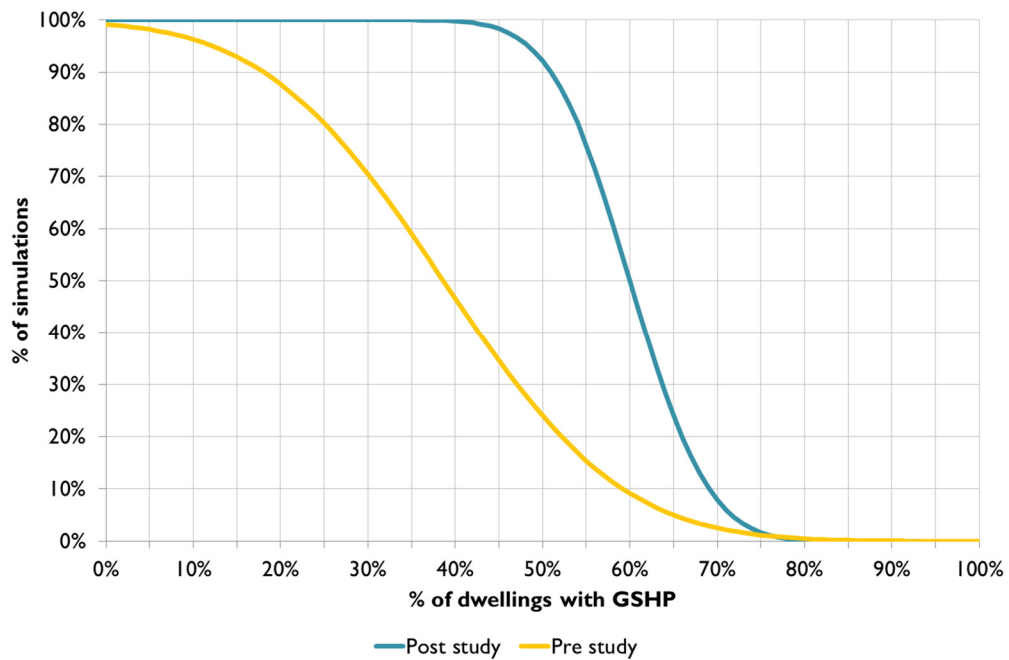
Initial acceptability check

At this point, options that have been identified as warranting further investigation would be subject to an initial sense check to confirm that they are viable, that they obey known constraints, and are aligned to high level policy objectives.

Focus granularity and option selection

As low regret options are identified at a high level and proven to be resilient, an iterative process of "zooming in", in spatial and temporal terms, allows the more exact determination of boundaries and timings of the projects over the pathway to 2050 that would realise these options. These options would then be fixed in a further iteration of the model, which may in turn narrow the breadth of outcomes regarding other options, until the full extent of low regret options that can be identified with a given state of knowledge has been determined.

In the example, as the specifications of an insulation programme are progressively fixed, the details of the district heating network is also resolved in more detail. Whilst the extent of insulation was more uncertain, a CHP plant at location Y was selected by the tool for deployment in 2025. As this uncertainty is reduced and the insulation programme fixed, the optimal deployment date for this CHP recedes to 2030.

Figure 4-15 **Example output: heat demand density and CHP plant details**

Figure 4-15 Example output: heat demand density and CHP plant details
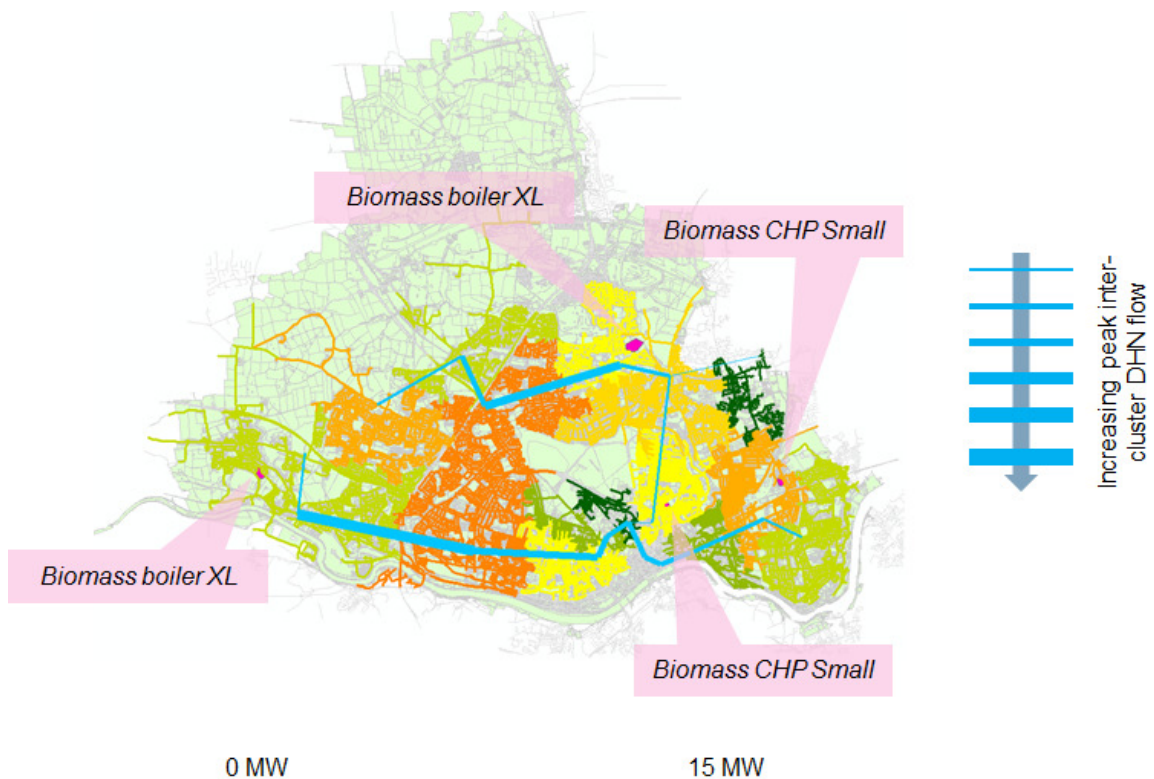
Option prioritisation

Once the projects are selected and defined, projects are prioritised and programmed to form part of the broader transition plan. The projects that are lowest regret, or those that are prerequisite enablers to other projects (such as insulation or grid upgrades), are likely to be selected for implementation first. Later deployment will be desirable for options that tend to lock in a pathway, or exclude options subsequently.

LAs may also wish to implement first those projects over which they have direct control or at least strong influence, such as those within LA owned buildings, to secure early progress. Other projects requiring implementation by private and independent decision makers are likely to involve a more complex strategy of long term engagement, influence and incentivisation, and be subject to significant uncertainty regarding timing.

Projects where the CO_2 emissions reduction effectiveness is subject to a significant degree of ongoing uncertainty, driven for example by consumer behavior, may need to be managed carefully over the duration of the transition plan to avoid over reliance on their effectiveness to deliver overall targets.

Wider acceptability check

At this stage a programme of projects, and high level options for subsequent pathways, has been identified. The EnergyPath™ Networks Tool is focused on the building, network and other technology choices necessary for the development of a least-cost local energy system transition plan, subject to the impact of carbon pricing or carbon constraints. Social factors may be

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification 46/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

included to the extent that they can be parameterised as costs or constraints, but this may not always be feasible. Therefore, at this stage it is appropriate to check performance of the pathways against wider targets. In particular, this is likely to include the cost of energy provision.

The majority of the tool's outputs can be presented in a GIS format, and potentially combined with other existing GIS datasets. For example the cost of household energy could be combined with a socioeconomic GIS dataset of income levels to understand the potential implications for fuel poverty. The tool could also enable estimates of other factors, such as noise, job creation or air pollution, by relating back to proxy drivers in the pathways. High level summary of the most important parameters may also be useful to assist communication.

In the example, a particular zone within the LA with a high level of social deprivation has been selected. The annual cost of energy provision for all dwellings and single occupancy flats has been plotted in 2025, under BaU circumstances and after implementation of the initial programme of projects. Costs are reduced in both cases.

Figure 4-16   **Example output: annual cost of energy provision, by dwelling type**
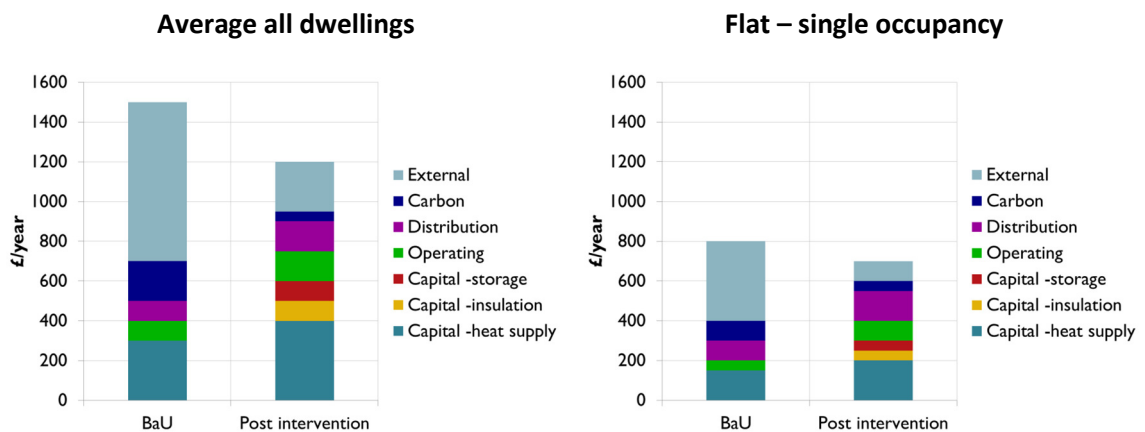
Summaries of headline figures have also been prepared for internal communication purposes, such as that below, illustrating cumulative and annualised cost under BaU and one of the high potential pathways.

Figure 4-17   Example output: total cost of energy provision, annualised and cumulative
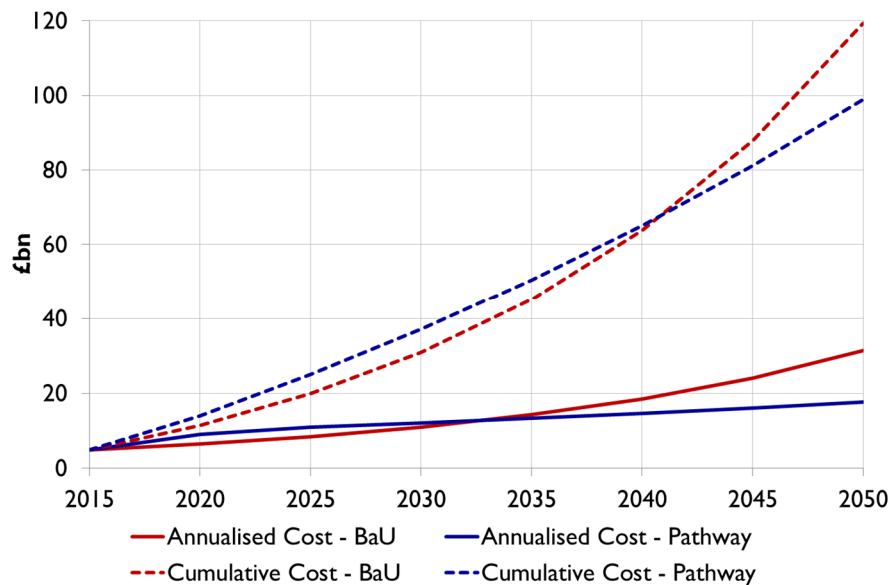
Analysis has also been undertaken of the locations of the investment to be undertaken in a high potential pathway, as an indicator of the proportion of the pathway decisions lying in the jurisdiction of different types of actor. This will be an important consideration in developing the final transition plan. Figure 4-18 illustrates that a relatively small part of the pathway would be directly enacted by the LA, limited to that occurring within LA owned housing stock and municipal buildings and facilities. A somewhat larger portion would be concentrated within the remit of a small number of network operators and large scale investors, with whom the LA may have an established relationship and a degree of influence. A large portion, approximately 60%, would however be disaggregated across the domain of numerous private individuals, businesses and other organisations free to make independent decisions. The effectiveness in reducing CO_2 emissions of investments in these locations once made is also likely to be dependent to a large degree on uncertain consumer behavior. The LA may well identify this as a point of risk in a transition pathway based on such a pathway, but consider it unavoidable given the nature of the local area. It may therefore embark on a process of engagement locally, and potentially with national government to ensure it is equipped with the necessary powers to incentivise decision making in the required manner.

Figure 4-18   Example output: investment by actor type

Legend:
- Other buildings
- Business buildings
- Private rental dwellings
- Owner occupied dwellings
- LA owned dwellings
- LA buildings
- Generation infrastructure
- Network infrastructure
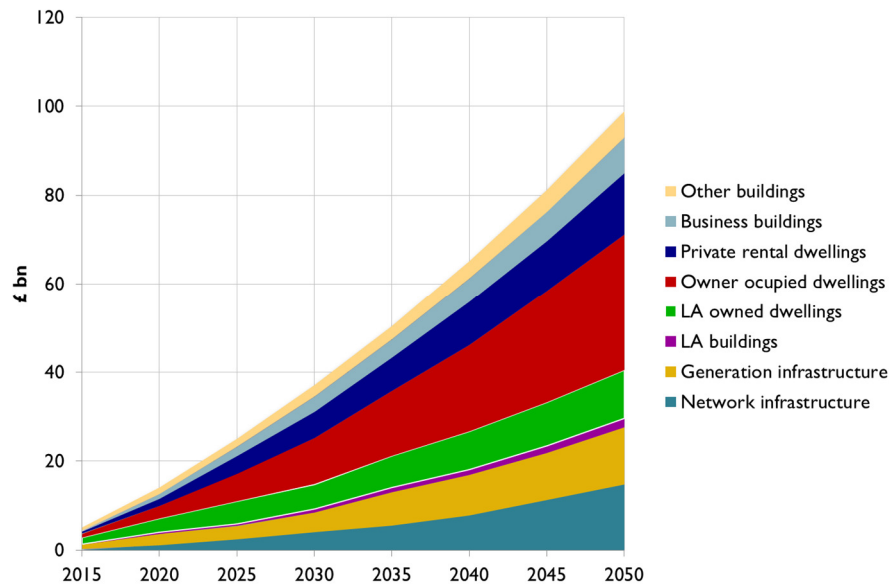
Project details

At the end of the consultation phase, when projects are approved for implementation, the tool will be used to generate technical specifications for projects and data that will enable financial modelling. This will include figures for such factors as the quantum and timing of capex and fixed opex, generation or energy savings (as appropriate) and their value based on a shadow price and output calculated for each time slice and period.

The example below sets out summary financial details for the initial CHP plant linked to the district heat network (plant X in Figure 4-15) and identified for commissioning in 2020.

Table 4-1   Example output: CHP plant X summary financials

	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
Capacity operational (MW$_e$)				80	80	80	80	80	80	80	80	80
Capex (£ mn)	5.0	35.0	25.0	15.0								15.0
Generation electrical (GW$_e$)				341	340	339	362	370	359	358	383	391
Generation thermal (GW$_{th}$)				316	322	312	318	325	301	322	345	351
Electricity price (£/MWh)				41.00	42.00	43.00	41.00	41.00	43.00	44.00	42.00	42.00
Heat price (£/MWh)				19.00	19.00	20.00	20.00	20.00	22.00	21.00	20.00	20.00
Electricity revenues (£ mn)				14.0	14.3	14.6	14.9	15.2	15.5	15.8	16.1	16.4

	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
Heat revenues (£ mn)				6.0	6.1	6.2	6.4	6.5	6.6	6.8	6.9	7.0
Opex (£mn)				12.0	12.2	12.5	12.7	13.0	13.2	13.5	13.8	14.1

Figure 4-19 provides illustrative financials for this plant over its assumed 25 year life, including periodic refurbishment capex, and the cumulative discounted free cash flow to the firm. At a 10% discount rate (pre-tax real), indicative of private sector return requirements for such a project, it is apparent that the NPV for the project is below zero. This may indicate that the project may need to be undertaken within the public sector, or more likely indicate the level of subsidy required for the project to be financed using private capital.

Figure 4-19   Example output: CHP plant financial summary
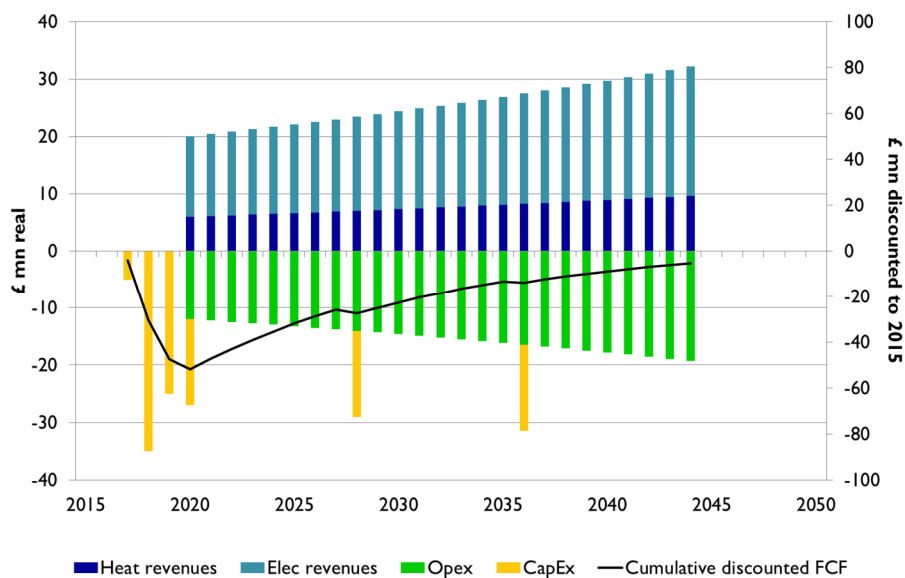
Figure 4-20 illustrates summary financial data over an assumed 25 year life time for Solid Wall Insulation, which is selected for installation later in the 2020s in a relatively remote area of the LA dominated by detached and privately owned properties. An 8% discount rate has been used for consumers in this case. At this rate, the measure provides a small positive NPV for the householder. However the extent to which this calculation holds true is likely to be highly dependent on the behavior, lifestyle and circumstances of that household. Such factors can only be approximated by the tool based on larger populations, and are subject to great uncertainty and variation over time.

Furthermore personal discount rates are also highly uncertain, and so the extent to which individual households make an investment decision in the case of a given financial projection is also subject to uncertainty. Therefore LAs are likely to approach such "projects" with caution, seeking carefully to understand the interaction of consumer behavior and incentive schemes (such as grants or Feed in Tariffs for heat under the Renewable Heat Incentive, RHI) with take up of particular measures. To some extent this can be captured in the configuration of the

EnergyPath^{TM} Networks Tool, for example by representing availability of the RHI for a technology as a reduced cost. Detailed representation of consumer investment decision making is not however component to the tool's cost optimisation approach, and some "post processing" of results may be necessary for an LA to determine a level of private take up of a given measure on which it is content to rely for its transition plan.

Figure 4-20   Example output: SWI summary financials
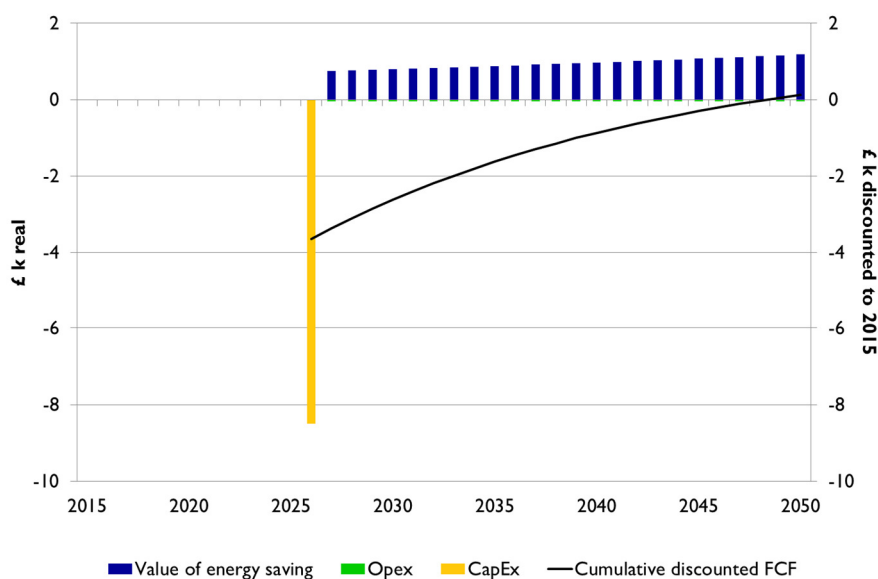
Accommodating changes arising from consultation or validation exercises

In the process of consultation, it is probable that new constraints or objectives may be identified which will affect the optimisation calculation, and lead to changes in the pathways determined. Some examples, and how these might be accommodated and impact on the tool's outputs are briefly discussed here.

▶ Work by the LA's environmental protection unit has identified a protected species on a piece of land selected for development of a CHP plant in a draft transition plan, and has begun a process to protect it from development.

  – In the configuration of the tool, the zone containing the proposed plant is now identified as being unavailable, constraining it from being selected for development. When the tool is re-optimised with this constraint, it is found that district heating remains the favoured solution, but now two smaller CHP plant are selected for development in alternative locations. There is a relatively small increase in costs associated with reduced efficiency of the overall pathway.

▶ Internal consultation within the LA reveals that the updated economic development master plan proposes a new office park on the outskirts of a settlement.

- – Accommodating demand from the new office park will require reinforcement of the electricity network at the 33 kV level from 2018, when the development is assumed to be completed. This is included in the known network developments, and the tool is re-optimised on this basis. The upgrade provides headroom for a group of dwellings which had previously remained on gas based heating until 2045 to implement electric heating solutions from 2025, enabling decommissioning of the gas network in this area to be brought forward.

▶ Consultation with the local water utility reveals that they are planning investment in an anaerobic digestion facility at a sewage treatment plant.

- – When this new heat source is included in the known development of local infrastructure, district heating becomes the optimal solution for heating provision in nearby housing, and investment in electricity network reinforcement is postponed.

▶ The LA has decided that it can no longer afford to subsidise operation of a swimming pool, which will be forced to close.

- – When this anchor load is removed from the tool, a district heat network in this area is no longer the lowest cost solution. The tool tends to select a more complex range of measures in the area, including selected investment in more marginal forms of insulation (such as SWI), earlier electricity grid reinforcement and transition to electric heating, and some use of biomass boilers. The tool user undertakes more detailed analysis of these options to confirm their resilience and to resolve project details to a higher level.

4.3. Practical illustration of key modes of use

This section is designed to illustrate how the user of the tool would practically address particular issues that may arise, or pertinent questions that may be posed by decision makers, as part of the process of using the tool to assist in design a local energy transition plan. Such issues include:

▶ How engagement with the community and technical engineering expertise may impact and be integrated within the business and transition plan design process

▶ How the tool can inform the degree of headroom included in designs, so as to build infrastructure which is resilient to different outcomes

▶ How the existing plans of LA or network owners can be accounted for

▶ How the tool can assist consideration of energy system choices over which an LA can exercise a degree of influence or control versus those it has more limited control over (e.g. consumer uptake behaviour in owner occupied or rented properties)

▶ How the tool can assist consideration of investment decisions made in centralised monopoly service providers (such as DNOs) versus those by individual private households exposed to free market competition; and

▶   How the tool can assist consideration of the impact of consumer behaviour on developing a robust transition plan.

As a hypothetical scenario, it is imagined that during the phase of assimilating existing investment and master plans in the example in section 4.2, it was determined that the LA has already developed plans to build a limited district heat network in a certain area, based around anchor loads provided by LA owned facilities.

This option would be fixed upfront by entering data in 3 areas of the tool:

▶   Its locational details would be entered via the SAM – e.g. the zoning of the heat sources, pipe network and connection points to heat loads, etc

▶   The integrated database interface to define the associated characteristics of the heat network – e.g. the size, operating lifetime, costs (investment and operating), etc

▶   User-defined constraints with the POM to force the deployment of this network at the right time within the pathway analysis (i.e. so it is user-controlled rather than an open economic decision). This would cover both the network itself and the upgrade of the LA owned buildings providing the anchor load

Figure 4-21   Business process excerpts: including existing development plans
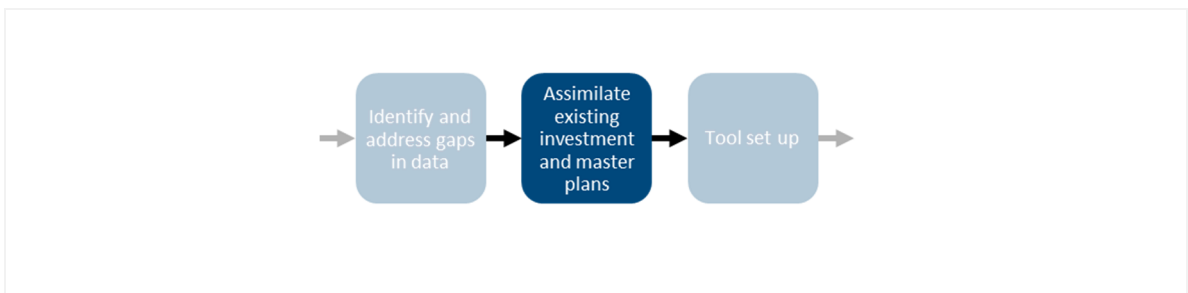
When the model is developed and the POM run, the results indicate that the cost optimal solution is for this network to be expanded to include an adjacent, mainly privately owned residential, district. When this plan is presented for internal approval, the steering committee express concern over what level of take up can be expected in this district, given householders are free to choose whatever heating solution they prefer. They ask the team responsible for use of the EnergyPath™ Networks Tool to further investigate the sensitivity of this option to different uptake outcomes.

Figure 4-22   Business process excerpt: iteration after internal validation
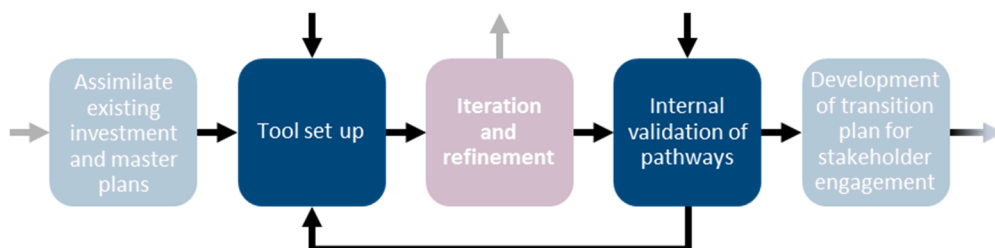
The EnergyPath™ Networks Tool team investigates by forcing in the POM different levels of uptake of heat network connections in the district. They identify that below 40% uptake, all

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   53/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

other things being equal, the POM stops selecting this option in favour of further electricity grid reinforcement.

The team also investigates under what conditions this uptake rate may transpire, assuming that if DHN connection presents a positive investment case at a private consumer level discount rate, the barrier to uptake is an uncosted "hassle factor". In the POM, this is explored by assigning different levels of cost to consumer hassle, converted from an equivalent cost per hour the consumer spends in handling the administration of a heat network connection or enduring degraded domestic conditions during installation. It is found that at a cost of hassle of £140/hr, uptake in the district falls below the critical 40% level.

The team identifies this factor as a possible subject for future investigation. By examining the total system annual costs the team is also able to identify the value of increasing uptake of heat network connections. At the minimum uptake of 40%, total costs are at parity with the competing electric solution, at 60% take up annual costs are reduced by £15 mn and at 80% take up annual costs are reduced by £40 mn.

Figure 4-23   Example output: annual system costs with one standard deviation range around expected cost of consumer hassle
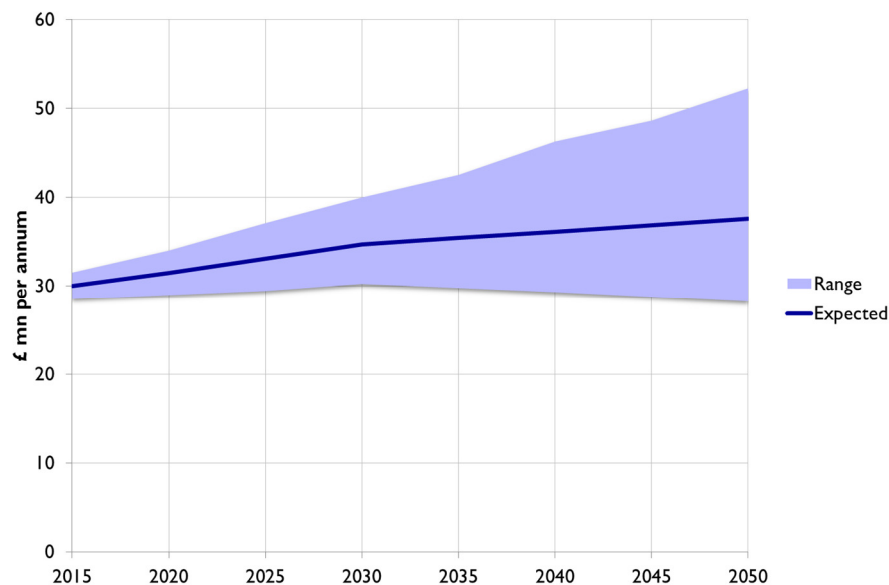
The team presents their findings for internal approval. The steering committee are now satisfied that there is sufficient understanding of how consumer take up and other factors such as energy prices impact this particular project. Whilst energy prices are beyond the LA's control, it is considered that DHN connection uptake is to some extent controllable. The LA has powers to vary council taxes within a certain range, and can support financing of home energy measures as a Green Deal provider. As part of the development of the transition plan, a policy unit within the LA is instructed to consider how these levers can be combined to incentivize, in an equitable manner, uptake of DHN connections and insulation in the area.

Figure 4-24   Business process excerpt: transition plan development
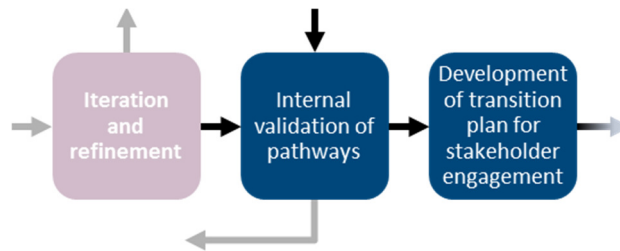
The DHN project is approved for inclusion in the plan for further consultation with key stakeholders including the regional electricity DNO, who use the analysis to consider their plans for grid reinforcement in the area. By studying the detailed outputs of the NAM when DHN uptake is fixed at different levels in the POM, they gain insight into what level of resilience against outturn DHN uptakes would be provided by different levels of electricity network reinforcement – i.e. if fewer customers connect than expected and peak electricity demand is higher than anticipated.

To further inform this analysis, the DNO also considers the impact on reinforcement requirements in the area of the network when different diversity scalars are assumed when calculating peak demand. In dialogue with the LA, the DNO determines a sensible level of headroom to include in its future reinforcement and asset replacement plan in the area, which is *now consistent* with the boundaries and residual uncertainty associated with the DHN plan.

The projects are now reviewed and approved by the LA executive and put to the public at large for consultation, and eventually included in the final transition plan. The DHN project (comprising the district heat network and to service it two energy centres, one based on biomass fired CHP and one on GSHP), now begins the process of implementation. The first step is to acquire expert review and validation of the project design and costs, before committing more substantial funds to a full FEED.

Figure 4-25   Business process excerpt: consultation to project validation
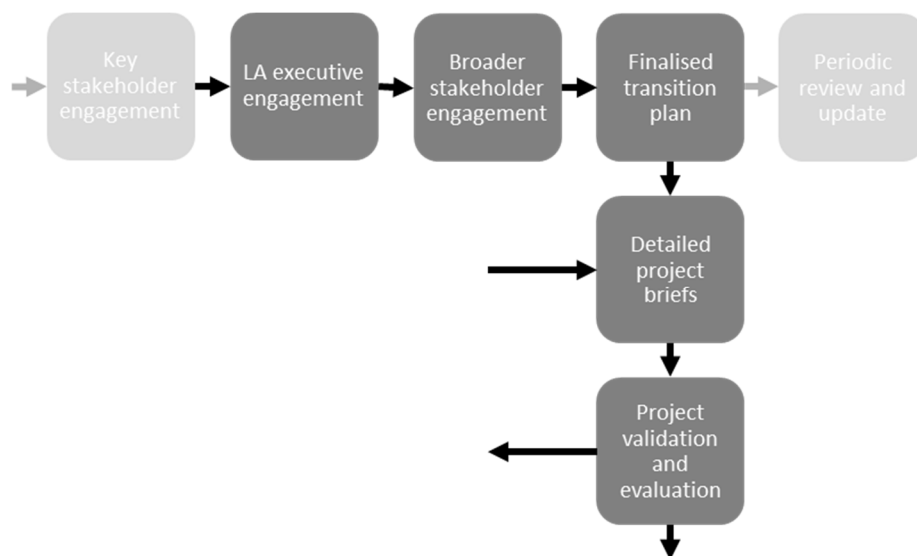
The LA commissions an engineering and planning consultancy to review the expected demand for the system, the technical specifications, and the cost estimates. The review undertaken determines that a GSHP based energy centre is not feasible at the site selected as the density of settlement thereabouts is too great to enable a sufficient length of ground collector to be laid. Additionally, in surveying the housing stock in the area to assess demand, a building type has been identified using materials and construction techniques peculiar to the area. The properties of this stock are such that energy efficiency can be improved at a lower cost than other housing of a similar age and size. As a result, the EnergyPath^{TM} Networks Tool team makes a number of amendments to the tool set up.

▶ In the HOM, details of the new domestic building archetype are entered, in the existing state and in various states of upgrade.

▶ In the SAM, locations of this building type are identified where known. The area initially identified for GSHP deployment is now flagged as being unsuitable for this technology, and hence this information will automatically flow through and the option will be excluded in subsequent analysis in the POM.

The POM is rerun with these new conditions. The new solution identifies a greater degree of insulation in the newly identified building archetype as being cost optimal. This in turn enables the heat demand of the network to be satisfied by a single enlarged biomass fired CHP plant. Plans for the larger CHP plant are validated by the LA's advisors, and expected costs found to be within the range projected by the tool. The plan then moves into the development phase with commissioning of a full FEED study.

4.4. Interfaces

The focus of development to date has been to produce a functional tool, suitable for an expert level user. As such, development of a fully integrated Graphical User Interface (GUI) is not within the scope of this functional specification, however, a simple Excel GUI has been developed to run the tool and streamline management of input/output data. In addition, the tool makes limited use of existing 3rd Party interfaces (such as those in the ArcGIS package to feed in spatial inputs). In the longer term, further development of a GUI would make the use of the tool by expert users more 'efficient' replacing the existing Excel GUI, as this has been developed in a manner that is separate from the underlying tool logic and data.

4.4.1. Scenario and sensitivity management

In applying the tool to a local area, a range of processes will need to be followed, as discussed in section 4, including data input and infrastructure analysis. The user will typically configure a range of scenarios and sensitivities, and be running the model in multiple modes (including deterministic, probabilistic for analysing uncertainty, and constrained runs for resilience testing). This will be managed through a control interface, providing the capability to manage the process flow, set up specific runs and sequences, and manage results sets.

The model will ensure that configuration control files are generated as part of the tool use, to ensure an audit trail and reproducibility. The resulting data, from deterministic use, bespoke scenario testing and probabilistic operation will be archived and managed so as to make it

readily accessible to multiple users. The tool has the ability to automate the execution of a standard package of uncertainty analysis.

Data storage

The EnergyPath^{TM} Networks Tool has the ability to store *all* data generated in simulations, labelling and archiving it in a fashion that allows easy access and management. In practice, the extent to which this is desired may be limited by the availability of data storage and the performance of host hardware. Accordingly, the tool facilitates straightforward adjustment of settings regarding data storage, allowing operation in a mode suitable for the task being undertaken and the performance of the machine used.

For example, the full energy supply/demand balancing results for every within day time period for every simulations will constitute a significant quantity of data. Whilst this data might be used as part of post-processing to create final summary results, it may be unnecessary to store the full set of disaggregated results as these could always be reproduced at a later date if required.

4.4.2.   Audit

The successful application of the EnergyPath^{TM} Networks Tool depends on a very large amount of underlying data. This comes from a wide range of sources, including nationwide data sets held by both private and government sponsored entities, locally focused datasets and plans, asset databases held by network companies, and assessments of technology components. These datasets will over time evolve and be updated, or replaced or supplemented with new sources.

A data catalogue has been produced, managed by the ESC and consistent with their internal processes, to identify and map these different data elements, and be structured so as to ensure ease of ongoing maintenance.   When the tool is operated, the specific versions of each dataset used are be logged, to ensure reproducibility and provide an ability to track data provenance and produce audit trails.

5. Household Options Module (HOM)

5.1. Overview

The Household Options Module (HOM) is one of the four key components of the EnergyPath[TM] Networks Tool. The HOM provides the domestic archetypes required for other software components to characterise the existing building stock within Local Authority areas under assessment, and the options for upgrading this stock in future over the pathway from 2015 to 2050. Specifically, the HOM provides other modules with data on energy consumption, peak demand and the costs of retrofit interventions (including insulation, heating systems, etc) in different building types.

The HOM also captures important parameters regarding the behavioral interactions of building occupants with their energy using equipment and systems, which is passed to the Pathway Optimisation Module (POM) to enable the uncertainty surrounding different technology choices to be estimated.
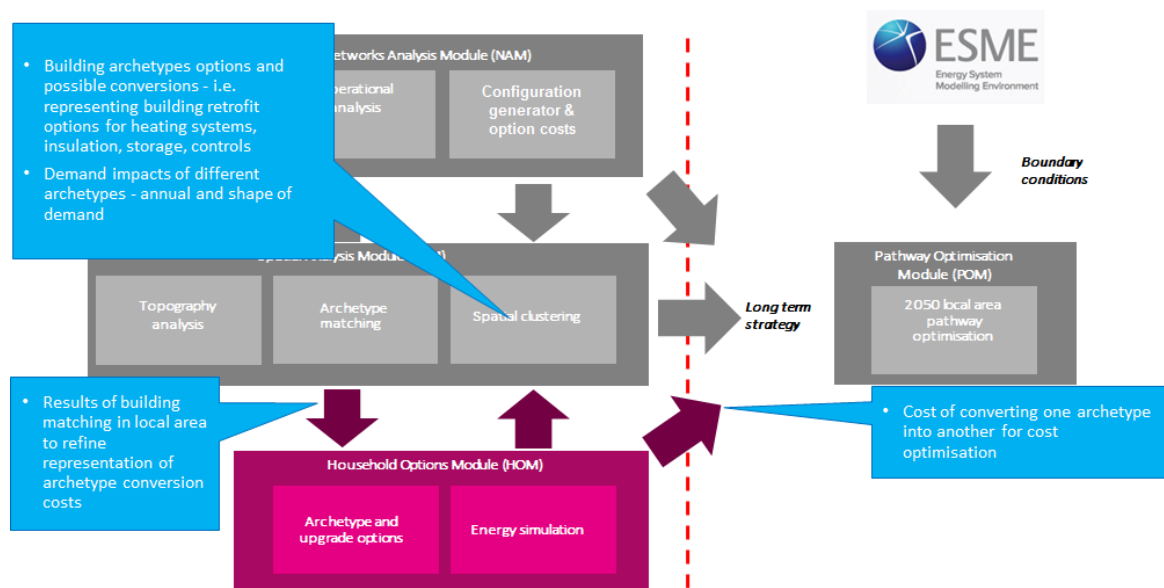
The characteristics of buildings are represented in the HOM using a so-called "archetype" approach. This is analogous to the creation of a large catalogue or a reference library of pre-defined building types against which data on real world buildings can be matched. This library is refined by information from the SAM reflecting which buildings have been matched to which archetypes across the different parts of the local area (i.e. if certain types of buildings do not exist they do not need to be included in the final catalogue). This is a deliberate data architecture design decision to help simplify the final pathway optimization undertaken within the POM (see section 8). Estimates of energy demand from domestic building archetypes are derived from bottom-up engineering simulations.

A challenge for the HOM is collapsing the vast matrix of possible characteristics used to identify and represent real world domestic buildings to only those which are most material for understanding the key outputs of energy consumption, peak demand and retrofit costs in a

particular LA. This is achieved through the use of mathematical problem decomposition (k-means clustering[11]) techniques to reduce the problem dimensions down to those building performance metrics which most strongly affect the desired outputs. In other words, where the characteristics of a number of archetypes are very similar, it is possible to aggregate them into one archetype with limited loss of information. To support this decomposition process the HOM uses a fast parameterised SAP (Standard Assessment Procedure) model to simulate the typical annual levels of electricity and heat provision required to maintain comfort conditions in each dwelling type. The approach used is aligned with the method used for assessing building regulations compliance in the UK[12].

Once the 'short list' of the most material domestic building archetype has been created, bottom-up dynamic thermal modelling of space heating and hot water, using the EnergyPlus[13] software, is then used to generate possible demand profiles for a number of 'characteristics days[14]' at 30-minute granularity (electricity demand profiles for EVs, lighting and other appliances would be added from exogenous assumptions).   This is undertaken for a range of different heating technology options for the archetypes including gas boilers, heat pumps, hybrids, etc and consider the optionality introduced by features such as storage.

The number and definition of characteristic days that can be considered is flexible and can cover typical average seasonal days (winter, spring, summer, autumn) as well as more extreme cases to test the resilience of the pathway or to design to a particular standard, for example, to meet a 1-in-20 or a 1-in-50 cold winter day.

5.1.1. Module diagram

The diagram below shows the key logic / process steps for the HOM, the equivalent for the full tool covering all modules is shown in section 11.

---

[11] Note that clustering here refers to a specific set of mathematical techniques and is separate to the broader, model specific concept of spatial clustering in EnergyPath[TM].
[12] For these calculations, EnergyPath[TM] employs an ISO 13790 model based on the UK Standard Assessment Procedure (SAP) methodology
[13] http://apps1.eere.energy.gov/buildings/energyplus/
[14] As part of this assessment it is necessary to simulate a period of time around the day of interest to account for e.g. the slower variation in thermal mass

Figure 5-2 Key HOM logic / process steps
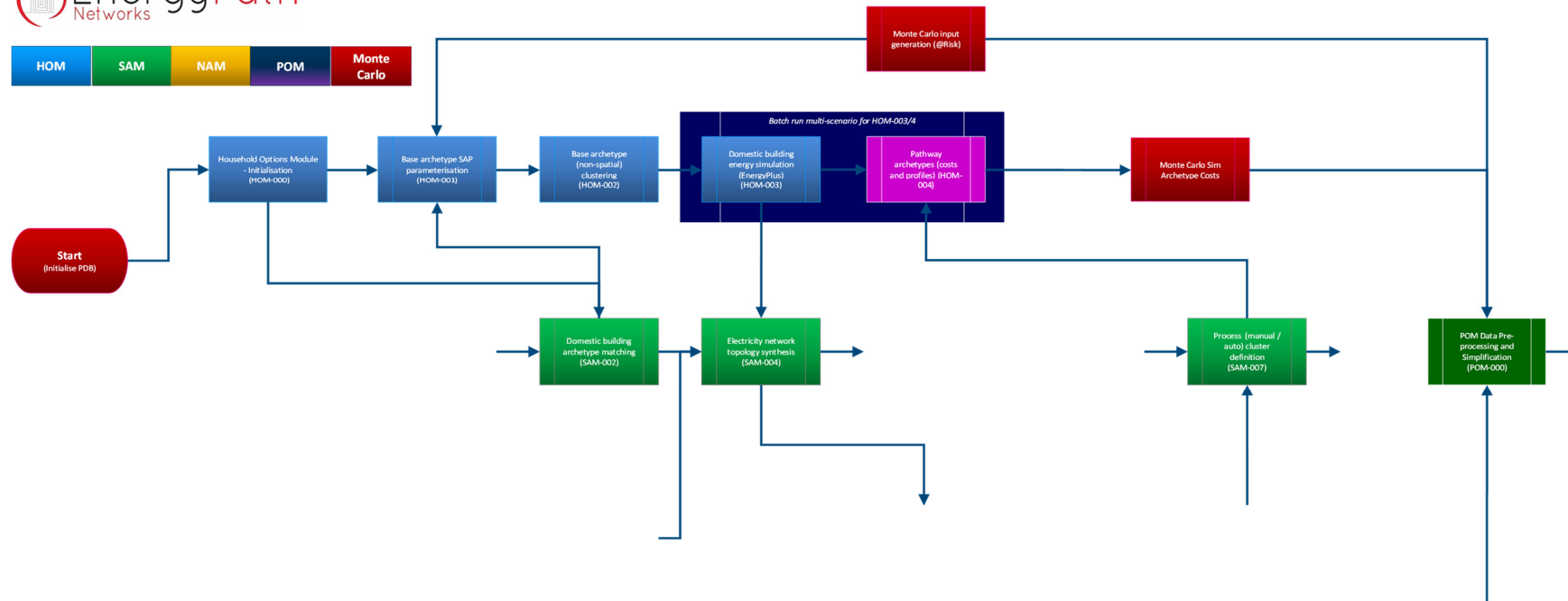
5.2. Key outputs

Key outputs from the Household Options Module include:

▶ The final building catalogue used to represent a local area is given by a set of archetype *pathways.* These represent the starting condition for different types of buildings within the stock and the spectrum of a set of possible choices about how that building could evolve (both in terms of efficiency measures and heating systems) from 2015 to 2050. For example, an improvement in efficiency in 2020 and further improvement in 2030 along with a shift from gas boilers to heat pumps at this point. The number of possible pathways for a given starting building is a function of the number of possible insulation retrofit options and heating system options defined, adjusted for implicit or user defined restrictions (e.g. no GSHPs in flats or no assumed possible conversion from low carbon technologies back to high carbon technologies).

▶ Parameterised final energy demand profiles for different building archetypes (which vary by user-defined characteristic days and can vary over timeperiods[15]) for different energy vectors on individual characteristic days, which are provided to the NAM to help establish the peak demand requirements for zonal load flow modelling and against which the POM can perform demand/supply balancing, when deciding which archetype pathways the different parts of the building stock should follow.

▶ The base timeperiod annual energy consumption of individual building archetypes for different energy vectors, which are provided to the SAM to support the validation of the archetype matching process

▶ The combined costs of retrofitting building fabric, alternative heating technologies, storage and controls which are passed to the POM for the purposes of numerical cost optimisation, when deciding which archetype pathways the different parts of the building stock should follow.

---

[15] For example, to reflect rising mean internal temperatures

Table 5-1 **Output data summary**

ID	Data field	Destination	Purpose	Data granularity	Uncertainty parameters
1	Base and Final archetype pathway catalogues	SAM / POM	SAM – support building archetype matching process POM - cost optimisation	Potentially around <10^3 final archetypes referenced against full set of around 10^6 valid archetype characteristic combinations Matrix of valid combinations to convert one archetype into another	N/a
2	Final energy demand profiles by archetype	POM / NAM	POM - supply / demand balancing NAM – to help understand maximum possible peak demand to bound load flow analysis	• Electricity (heat-related), electricity (EV), electricity (other), gas, network hot water, and other final energy vectors (biomass, H2) • By each archetype (some archetypes have >1 profile to reflect flexibility of operation of some devices/storage), • By characteristic days • By half-hour within day granularity • By timeperiod (e.g. if want to reflect changing mean internal temperatures over time)	Uncertainty in profile shapes for electricity (heat), gas, network hot water mostly influenced by: • External temperature on characteristic days • all related factors listed for "ID 2: Energy consumption by archetype" Uncertainty in profile shapes for electricity (other) mostly influenced by • Occupancy [by default occupancy will be inferred from floor size, but the user will be able to override this as part of the simulation process to test different sensitivities] • Power rating and number of electrical appliances • Occupant operation of electrical appliances
3	Annual final energy consumption by archetype	SAM	SAM – support validation of building archetype matching process	• Electricity (heat), electricity (other), gas, network hot water, and other final energy vectors (biomass, H2) • by each archetype, • for the base timeperiod	Uncertainty in consumption for electricity (heat), gas, and network hot water is mostly influenced by: • Building thermal conductivity • Building inner thermal mass • Occupancy • Occupant heating hours • Occupant heating temperature Uncertainty in consumption for electricity (other) influenced by the same factors as listed for "ID 1: Final energy demand profiles by archetype"
4	Archetype retrofit / Conversion Costs	POM	POM cost optimisation	• Costs of converting one archetype into another considering separate retrofit options for building fabric and heating systems for each archetype,	Uncertainty in retrofit costs influenced by: • Complexity in required alterations to building fabric (labour/materials) • Complexity in required changes to heating system (labour/materials)

				• Costs will vary by vintage of retrofit to account for changing costs (e.g. of heating systems in different timeperiods)	• Uncertainty inherent in matching starting condition of a dwelling to its archetype representation

5.3. Key functional requirements

5.3.1. Characterising building archetypes

The EnergyPath™ Networks Tool is a strategic Networks Tool for investigating technology transitions at Local Authority scale. A key functional requirement for the HOM is therefore to represent energy demand from buildings to a sufficient level of detail to derive insights on possible technological change.

Approaches to modelling residential sector energy demand can be characterised broadly as either top-down or bottom-up methods. Tools such as the EnergyPath™ Networks Tool seek to represent the outcomes of endogenous technological change at a detailed level and so a bottom-up approach is required to allow for explicit representation of the housing stock.

Figure 5-3   **Broad Approaches for Characterising Residential Energy Demand (Based on Swan & Ugursal, 2009[16])**
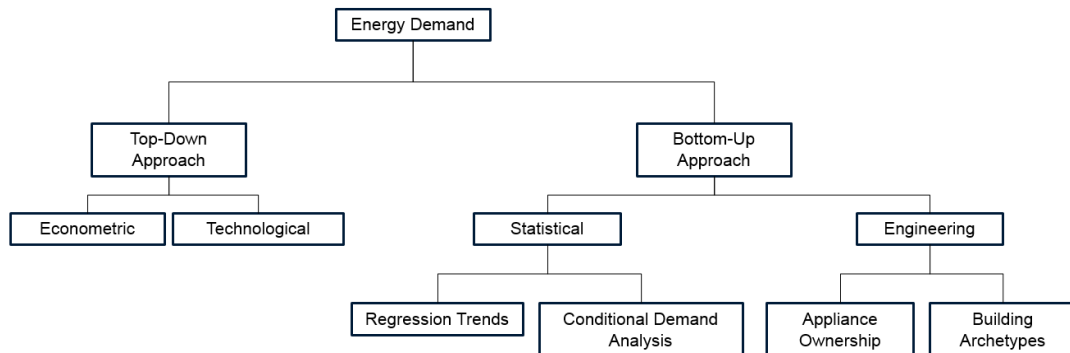
All complex models require some abstraction from reality in order to ensure that the problems considered remain computationally tractable. The approach taken for the Household Options Module (HOM) is to create a large dataset containing pre-defined housing archetypes and information on their estimated energy demands, against which data on real-world addresses can be matched. The housing stock of the United Kingdom is diverse and spans a wide range of construction methods, architectural styles, form factors, and vintages amongst its 26 million unique homes.

A challenge for defining archetypes in the HOM is to reduce the number of archetypes so that only the most material characteristics for simulation of the required outputs are being modelled. The approach taken is described in detail in Section 5.5 below.

5.3.2. Estimates of demand and within day profiles

A key functional requirement for the HOM is to estimate the final energy demand profiles of individual building archetypes for a set of characteristic days, such as average seasonal days across the year or may extreme cases such as a 1-in-20 cold winter day, which reflect a design standard that the viable pathway must be able to accommodate. The profiles are used by the

---

[16] Swan, L. G., & Ugursal, V.I. (200(. Modeling of end-use energy consumption in the residential sector: A review of modelling techniques. Renewable and Sustainable Energy Reviews, 13(8), 1819-1835

NAM to help understand the maximum likely demand seen on a section of a network and within the POM as part of the overall supply / demand balancing process.

Bottom-up approaches to modelling residential energy demand can be broadly characterised as either engineering or statistical methods:

▶ Statistical approaches broadly use empirical data on real-world energy use, such as meter readings, to derive observed relationships between phenomena like internal temperatures and the use of energy in homes over time.

▶ Engineering approaches seek to represent physical processes that occur in buildings such as fuel combustion from heating appliances and thermodynamic interactions between the heating system and the structure.

The science of modern building energy modelling is dependent on a mixture of both statistical and engineering-type approaches due to challenges associated with capturing real-world data on how people inhabit and use energy in their homes.   The approach taken for the HOM therefore combines statistical and engineering approaches for estimating different end-user energy demands:

▶ Space heating demand is estimated using an engineering approach that uses a detailed technical simulation of thermodynamic processes in different building archetypes.

▶ Non-heating electricity and hot water demand are derived from parametric methods that use statistical estimates of energy using behaviour linked to occupancy. More detail is described in Section 5.5.4 below.

Temporal granularity is a key consideration in dynamic simulation models, and covers two key dimensions.

▶ ***The number and type of characteristic days.*** This is intended to be a flexible element of the HOM, subject to computational constraints within the dynamic simulation process discussed in 5.5.4). As a minimum however, the peak design conditions for the end-user building systems need to be captured. This varies for different sectors and for different end-use demands. For domestic heating plant in the UK, for example, the peak design condition is likely to be a winter day, while for commercial buildings with electrical air conditioning systems, the peak could fall on a summer day. This suggests that at least 2 characteristic days are required for dynamic simulation of all building archetypes. End-users may want to evaluate design conditions that are more onerous than typical cold and hot periods, such as heat waves or exceptionally cold 1 in 20 year winters.

  – A distinction is often made between profiled energy demand on weekday and weekend periods so design conditions may also need to take into account the day of the week as well as the season.

  – Finally, dynamic simulation of characteristic conditions typically requires an extended period of between 4-8 days in advance of the target day to be represented in order for transient effects such as thermal mass response and storage to be captured adequately.

▶ ***The temporal granularity within day.*** These can vary from simple on- or off-peak blocks across the day down to very fine resolution (e.g. <1 minute) intervals. Very high time resolution steps are more appropriate for detailed engineering system design tools rather than software used for evaluation of strategic technology deployment potential, like the EnergyPath™ Networks Tool. Greater time resolution also has implications for computational performance i.e. program run times.

The combination of characteristic day and within day period is referred to within the EnergyPath™ Networks Tool parlance as a *timeslice.* The level of granularity in the timeslicing ultimately depends on the resolution needed to understand potential trade-offs in pathway design within the model.

For example to understand how desirable a hybrid boiler is, it is important to consider the combination of peak shaving (e.g. to avoid network reinforcement) within day versus overall emissions and direct running costs, which depend on operation across the year. As heat demand swings substantially across the year (along with other factors such as the cost of electricity) sufficient temporal detail (e.g. seasonal / month) is necessary for the model to understand the trade-offs in running the gas versus heat pump component. Without this detail the user must instead make an exogenous assumption about the balance of operation of these components across the year. Similarly, within day there must be sufficient granularity to understand where the electricity peak is likely to arise given overlapping demand profiles and the ability to shift load across the system, either from the operation of hybrid boiler or from other options such as within building or network connected storage.

As a result the model uses the following typical timeslices

▶ Four typical seasons by working and non-working days (in addition to this the user can define a number of additional non-typical characteristic days to represent design standards)

▶ 30-minute time resolution within day

It is not tractable to represent this level of detail within the POM optimisation problem, alongside significant spatial and pathway timeperiod detail. However, having the HOM produce the final energy demand profiles at this level of resolution gives the user (via the POM - see section 8.6.1.2) flexibility over how the final timeslices are defined for pathway analysis. The user can easily collapse or exclude certain timeslices and understand to what extent the solution changes – e.g. if you retain 4 seasons and 1 peak day versus collapsing this to 1 peak day and 1 typical day to represent the entire year, or aggregating the 30 minute periods into 3 blocks within-day.

It is important to note that the HOM produces a 'standalone catalogue' of building archetype pathways and their associated final energy demand profiles in *isolation* for each building. Once buildings have been allocated spatially and different pathway upgrade options (e.g. in terms of electrifying heating) it is then important to understand demand diversity impacts between buildings in a localised area, to avoid overestimating peak demand levels and by extension network/generation reinforcement requirements. *This cluster-level diversity is discussed further in section 8.6.1.1.*

5.3.3. Costs of retrofitting / converting building archetypes

Domestic

A key output from the HOM is the cost of different retrofit options for individual building archetypes, as making changes to the existing building stock is one of the many options available to the POM. The POM needs this information in order to make solution choices based around the costs of the simulated housing options.

Retrofit options include making improvements to the thermal efficiency of the buildings by improving the building fabric with measures such as insulation, and deploying alternative heating technologies such as different types of electric heat pump in lieu of conventional gas boilers[17]. Not all building retrofit options will be applicable to all building archetypes. For example, buildings which have solid walls by definition cannot be selected for cavity wall insulation, and ground source heat pumps are not likely to be realistic options for individual above-ground flats.

While changes to the heating system are effectively binary choices (the system is changed or it is not), in reality, even for a single housing archetype, there are many degrees of fabric retrofit which are possible. For example, loft insulation could be added as a stand-alone measure, or in combination with wall-insulation. This is an important optimisation problem to consider, as the costs of implementing measures in concert may be lower than making multiple discrete interventions.

The approach taken in the EnergyPath[TM] Networks Tool is to create a flexible framework to define archetype retrofit combinations – i.e. one archetype is converted into another at a given cost with resulting changes to its final energy demand profiles (e.g. due to different heating systems or improved insulation). What the conversion actually represents could be a minor incremental improvement or a whole house retrofit. The number of retrofit permutations that can be accommodated will be determined as part of performance testing requirements in Stage 2.

5.3.4. Capturing 'householder behaviour'

User behaviour is a strong determinant of energy use in and ultimately accounts for much of the observed variation in levels of energy demand and patterns of energy use between otherwise similar buildings. A key functional requirement of the HOM is to capture an estimate of this variation from individual building archetypes. Consumer preferences also directly relate to the acceptability of different technologies and implicitly affect their installation costs.

The approach taken for representing householder behaviour in the HOM is:

▶ To capture uncertainty from energy-using behavior in buildings by estimating the upper and lower boundaries of demand from a combination of endogenous probabilistic simulation and exogenous ranges determined from literature

---

[17] Where a particular vector is no longer available e.g. the gas grid is decommissioned the cost of converting appliances (e.g. gas to electric cooking) would also need to be accounted for.

▶ To handle consumer preferences as indirect exogenous inputs, as noted in section 2.1

5.3.4.1. Energy-using Behaviour

Building occupants operate the energy using appliances in their homes to meet their daily needs and comfort requirements, which vary between individuals and households. Variation in how people operate their power and heating systems is driven by factors such as individual preferences and social and cultural norms (washing clothes every day, expectation of thermal comfort levels etc). While it is difficult for any single building to estimate the exact pattern of energy use that will arise on a given day, it is possible to estimate a measure of the spread around the typical or average day. *The approach to uncertainty across the tool is discussed more in section 9.*

5.3.4.2. Consumer Purchase Behaviour

Different households in otherwise identical buildings with the same heating system and electrical appliances may exhibit different patterns and levels of energy use, as described above. Another important factor to consider in demand modelling is that different households demonstrate variation in terms of their appliance ownership and how this changes over time. Consumer preferences may result in different households choosing to buy greater or fewer electrical appliances, to invest in microgeneration systems, or to choose to drive an electric vehicle (EV). This means that over time, two initially identical buildings could eventually arrive at a future time period where they possess different levels of insulation and completely different heating systems.

Carrying out a detailed modelling exercise to estimate the pattern of future investments in energy using technologies over time between different building and household combinations is complex. This requires discrete choice econometric methods or an explicit agent-based approach to understanding the influence of technology availability and cost, market structure, supply chains, and end-user demand on what may be bought and when. This level of complexity is best approached by using tools that have been specifically designed to investigate these issues, and this is not the focus of a strategic techno-economic pathway tool like the EnergyPath™ Networks Tool.

This does not mean consumer preferences cannot be captured in the EnergyPath™ Networks Tool, but it does mean that the onus is on the end user to understand and represent the effects of consumer choices by varying the inputs to the model. The effects of different consumer preferences can be compared and contrasted through establishing different sets of scenario inputs, informed by data and insights derived from other tools and studies. For example:

▶ A scenario where consumers exhibit a strong preference for a certain technology can be examined by constraining the model to force a minimum level of uptake.

▶ Another example is to change the modelled costs of different technologies to reflect their implicit consumer value, such as representing the "hassle" costs associated with the installation of a technology that is hard to integrate and involves disruption to daily life i.e. representing the negative value associated with unpopular technologies.

5.4. Key inputs

The key inputs for the HOM are summarised below. All inputs are ultimately required for creating the building archetypes, performing energy demand simulations for domestic and non-domestic buildings, and determining retrofit costs. Data inputs to the HOM can be separated into:

▶ Information that is required for matching building archetypes with address-level data from the SAM. These are the *class or type identifiers* that can be used for characterising buildings when detailed local survey information is not available to provide a *complete* set of data for an individual building.

    − It should be noted that there is a two-way interaction between the HOM and the SAM modules. The HOM initially provides the set of data used to define the master set of base building archetypes to inform the SAM building matching (i.e. what is the closest match for individual building in the area to the master set of archetypes). The results of this matching process are then passed back to the HOM to help refine and make the costing of the forward looking archetype pathways options more efficient, i.e. by excluding combinations of archetypes which do not exist in specific parts of the local area under investigation.

▶ Information that cannot be provided by the SAM, but which is nevertheless required for building archetype energy simulation. This information often needs to be inferred statistically from best available third party datasets. For example, building construction details (wall, loft and window u-values) are rarely available from primary spatial data, and are typically inferred based on characteristics such as the building age.

▶ Information that is purely required for the purposes of building energy calculations. In a few cases this information is exogenously user-defined, which implies that the end user must make specific assumptions to frame their investigation from a series of choices that are defined in POM, for example the desired internal temperature.

The tables below describe what is technically required within the tool and how it is used. The relative importance of the primary data inputs and other factors such as costs and licensing restrictions, particularly where they must be purchased from an external provider, are discussed in the original deliverable (D4) Data Acquisition Plan. It should be noted that the ESC currently maintain the master set of input data for the current v2.1 EPN tool and this is not documented here.

Table 5-2   Input data – archetype identification characteristics for SAM matching – information from primary data

Data field(s)	Input type	Purpose	Granularity
Property Type	Primary	Key for determining morphology of building, in particular external facing wall area and volume, which are material for energy use	Per domestic archetype ~5 classes. Examples: Flat, terraced, bungalow, semi-detached, detached
Floor Area Banding	Primary		Per domestic archetype ~5 classes. Examples: <50m², 50-69m², 70-89m², 90-109m², 110m²>
Property Age	Primary	Age on its own does not give energy use, but important for matching many other morphological characteristics	Per domestic archetype ~7 classes Examples: pre-1870, 1871 - 1919, 1920 - 1945, 1946 – 1954, 1955 - 1979, post-1980, post-1990
Government Office Region	Primary	Location determines factors such as external temperatures in summer/winter, and solar insolation, so this is material for space heating energy use and possible microgeneration from photovoltaics	Per domestic archetype ~11 classes. North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East of England, London, South East, South West; also Wales, Scotland

Table 5-3   Input data – archetype identification characteristics for SAM matching - information that can be inferred from aggregated sources (postcode, LSOA, MSOA, LA, GOR, national scale etc.)

Data field(s)	Input type	Purpose	Granularity	Uncertainty parameters	Coverage
No. of Storeys	Primary	Key for determining morphology of building, in particular external facing wall area and volume, which are material for energy use	Per domestic archetype ~3 classes. Single Storey, Two Storey, Two Storey	EHS reporting provide no data on the accuracy of this metric or how it was collected, analysis of raw data needed to quantify uncertainty	England
Wall Type	Primary	Key for determining thermal efficiency and retrofit options, material for energy use	Per domestic archetype ~4 classes. Examples: Cavity Filled, Cavity Unfilled, Solid Wall Uninsulated, Solid Wall Insulated Externally, Solid Wall Insulated Internally	"	"

Window Type/Glazing Type	Primary	Key for determining thermal efficiency and retrofit options, material for energy use	Per domestic archetype ~8 classes. Examples: Mixed Types, Single Glazed Wood Casement, Single Glazed Wood Sash, Single Glazed UPVC, Single Glazed Metal, Double Glazed Wood, Double Glazed UPVC, Double Glazed Metal	"	"
Loft Insulation Thickness Band	Primary		Per domestic archetype ~5 classes. Examples: 0-50mm, 51-100mm, 101-150mm, 151-200mm, 200>	"	"
Existing Primary Heating System Type	Primary	Key for determining space heating and hot water use, demand profiling	Per domestic archetype ~20 classes. Examples: Gas Boiler with Storage, Gas Combination Boiler, Electric Resistive Heating, Oil Heating, Solid Fuel + various new heating system types	"	"
Secondary Heating System Type	Primary	Key for determining space heating and hot water use, demand profiling	Per domestic archetype Examples: Electric Resistive Heating, Solid Fuel	"	"
Heating System Controls	Primary	Key for determining space heating and hot water use, demand profiling	Per domestic archetype Examples: Time Clock Control, Cylinder Thermostat	"	"
Hot Water Storage Type	Primary	Key for determining space heating and hot water use, demand profiling	Per domestic archetype Examples: Tank Size / Type combinations	"	"
Hot Water Tank Insulation Type	Primary	Key for determining space heating and hot water use, demand profiling	Per domestic archetype Examples: Foam, mineral wool jacket	"	"
Hot Water Tank Insulation Thickness Band	Primary	Key for determining space heating and hot water use, demand profiling	Per domestic archetype Examples: None, 0-160mm, 160mm>	"	"
Type of microgeneration installation	Primary	Affects final energy vectors	Per domestic archetype Examples: None, Solar PV, Solar Thermal, Micro-Wind	"	"

Table 5-4   Input data – variables for building energy and retrofit cost simulation that cannot generally be matched to spatial datasets

Data field(s)	Input type	Purpose	Granularity	Uncertainty parameters	Coverage

Component cost and performance of archetype upgrades – building fabric	Primary	Cost overall conversion	Based on ETI optimising thermal efficiency study – considers • Insulation options Decline in costs over time	High, medium and low costs	UK
Component cost and performance of archetype upgrades – heating systems	Primary	Cost overall conversion	Based on ETI Smart Systems and Heat Work Area 1 data – considers • Heating systems • Storage • Controls Decline in costs over time	Likely that performance and cost data will fall within upper and lower bounds	UK
Component cost and performance of archetype upgrades - microgeneration	Primary	Affects final energy vectors	Per archetype. Examples: Solar PV, Solar Thermal, Micro-Wind	Source-dependent	N/A
Target Mean Internal Temperature	Primary	Domestic energy simulation	Single Point	User-defined input	Per Domestic Archetype
Occupancy levels	Secondary	Domestic energy simulation	Number of occupants and pattern of occupancy	Inferred from floor area data	Per Domestic Archetype
Electrical Appliance Consumption Band	Primary	Key for determining electrical energy demand	Per domestic archetype. Examples: <3,000 kWh, 3,000-7,500 kWh, 7,500 kWh>	Elexon dataset does not provide information on uncertainty, may be possible to infer statistically from Home Electricity Survey	Per Domestic Archetype
Ventilation System	Primary	Most UK properties are naturally ventilated, possible to infer others from heating system/type/age/size – generally newer properties	Per domestic archetype Examples: Naturally ventilated, mechanically ventilated	User-defined input	N/A
External Temperature on Characteristic Days	Primary	Domestic energy simulation	Hourly or better for up to 5 characteristic periods of at least 5 days each (to allow for thermal mass effects)	Uncertainty inherent in future projections of weather conditions	Per Domestic Archetype
Domestic Electrical Appliance (Non-Heating) Demand Profile - covers - EVs	Primary	Domestic energy simulation	Half hourly per archetype for up to 5 characteristic periods of 1 day each	Uncertainty inherent in source information will filter into the HOM	Per Domestic Archetype

- Appliances and lighting				
Domestic Hot Water Demand Profile	Primary	Domestic energy simulation	Hourly per archetype for up to 5 characteristic periods of 1 day each	Uncertainty in time and scale of hot water demand	Per Domestic Archetype

5.5. Key HOM logic / process steps

5.5.1. Household Options Module Initialisation [HOM-000]

*This sub-module creates the most detailed set of all possible building options in EPN - **base archetypes** - given their physical characteristics. It then assesses the underlying Heat Loss Rate (HLR) of each base archetype via SAP model calculations.*

As indicated in Section 4.4, a wide range of physical housing characteristics are important for determining estimates of energy use in dwellings. The first step in the process for creating domestic building archetypes is to determine the various combinations of characteristics that will be modelled:

▶ Building characteristics that are both material to energy consumption and available in SAM datasets (e.g. floor area) and cannot generally be inferred (see Table 4-2) provide the high level matrix of possible archetype permutations. I.e. these are key to enabling each individual building in the area to be matched to its closest base archetype as part of the SAM-002 sub-module (see section 6.5.3).

▶ Other SAM matching characteristics[18] can be inferred from aggregate datasets, such as national housing surveys (see Table 4-3). Statistical analysis of this data can be used to determine whether there are any combinations of defining characteristics that do not exist in the housing stock or which are statistically not significant at a user-defined level (for example <1% of the stock). For example, modern new-build flats in urban areas which have oil fired heating are technically possible to construct, but are arguably not useful to simulate in the EnergyPath^{TM} Networks Tool, as there may be few or no real-world examples, nor is it likely to be a potentially attractive technology for retrofit in a smart heat system. Cavity wall construction was introduced to the UK in the 19th century, so buildings constructed prior to this period should only have solid walls, etc. Combinations of retrofit options that appear improbable can also be filtered at this stage, such as:

– Heating systems that require external fuel storage (such as LPG or oil heating) in individual apartments

– Heating systems with thermal storage tanks in small properties (e.g. <50m^2) that have already been designed with or converted to use combination boilers

---

[18] For example, there are no existing national datasets that show precise "Wall Type" at an individual address level, but data at a national level does enable wall type to be inferred by property age, type and tenure. This is a limitation of current data availability. If additional address level data on building characteristics becomes available there is no barrier in principle to including this when generating the high level matrix of characteristics.

–   Heating systems that require ground-floor access on properties that may not have ground-floor access, such as ground-source heat pumps in individual apartments

–   Roof-mounted micro-generation systems on properties that may not have roof access, such as individual apartments

5.5.1.1.   Base variables and allowed values for domestic buildings

Table 5-5 and Table 5-6 describe the variables used to define '*Base Archetypes*' for existing and new dwellings, and the possible values those Base variables may take. The full list of Base archetypes is formed from all valid combinations of these Base values, subject to the 'validity rules' set out below. The variables and their values have been developed following careful consideration of materiality on energy use and availability of existing data (through sources such as Geoinformation and the EHS) and future data availability.

Heating system combinations, comprising a primary heating system, an optional secondary heating system, an optional choice of storage options and an optional heating controls category, will be formed from the individual elements shown in Table 5-7[19]. From the large number of total possible heating system combinations, the user can manually form a set of ≈10-50 heating system combinations for each model run. The user can also be able to define the fraction of the total heating system capacity met by the primary and secondary heating systems. We note that hybrid systems will be treated as a combination of primary and secondary heating systems. Different sizes of hot water storage tank (e.g. 200l or 500l) can also be set as part of the possible heating system combinations[20].

Table 5-5   **Existing dwellings: Summary of Base variables and allowed values**

Variables for Base archetype formation	Number of values	Values
Property Type	10	1. End terrace 2. Mid terrace 3. Semi-detached 4. Detached 5. Converted flat – top floor, end terrace 6. Converted flat – top floor, mid terrace 7. Converted flat – mid floor, end terrace 8. Converted flat – mid floor, mid terrace 9. Purpose-built flat – top floor 10. Purpose-built flat – mid floor
Floor Area Band	8	1. Floor area band 1 (less than 50 sqm) 2. Floor area band 2 (50 to 69 sqm) 3. Floor area band 3 (70 to 89 sqm)

[19] These relate to the pre-configured set of EnergyPlus heating system objects used in the HOM-003 sub-module.

[20] It may in some cases be able to set a storage size which is too small for the dwelling and heating system size, leading to an ability to provide the target internal temperature level. This is assessed through the EnergyPlus simulation in the HOM-003 sub-module, combinations which cannot meet the target level (subject to any user-defined tolerance bounds) are excluded.

		4. Floor area band 4 (90 to 109 sqm) 5. Floor area band 5 (110 to 199 sqm) 6. Floor area band 6 (200 to 299 sqm) 7. Floor area band 7 (300 to 399 sqm) 8. Floor area band 8 (400 sqm or more)
Property Age	5	1. Pre-1914 2. 1914-1944 3. 1945-1964 4. 1965-1979 5. 1980-present
Number of Storeys	3	1. 1 2. 2 3. 3
Wall Type	5	1. Cavity (filled) 2. Cavity (unfilled) 3. Solid (uninsulated) 4. Solid (insulated, internal) 5. Solid (insulated, external)
Glazing/Window Type	4	1. Single glazing 2. Double glazing 3. Triple glazing 4. More than triple glazing
Loft Insulation Thickness Band	5	1. No loft 2. None 3. less than 100 mm 4. 100 up to 199 mm 5. 200 mm or more
Ventilation type	2	1. Natural 2. Mechanical
Primary heating system	≈10-50	Set of manually defined heating system combinations constructed from individual heating system elements
Secondary heating system		
Storage option		
Heating controls	3	1. None 2. Basic 3. Advanced

Table 5-6   New dwellings: Summary of Base variables and allowed values

Variables for Base archetype formation	Number of values	Values
Property Type	10	As above
Floor Area Band	8	As above
Property Age	3	1. New build HLR 1 2. New build HLR 2 3. New build HLR 3
Number of Storeys	3	As above
Wall Type	1	1. New build wall type
Glazing/Window Type	1	1. New build glazing type
Loft Insulation Thickness Band	1	1. New build loft insulation
Ventilation type	2	As above
Primary heating system	≈10-50	As above
Secondary heating system		
Storage option		
Heating controls	3	As above

Table 5-7   Individual heating system elements included in heating system combinations

Heating system category	Heating system element
Boiler	Individual Gas Boiler
	Individual Oil/LPG Boiler
	Individual Biomass Boiler
	Individual H2 Boiler
Fuel cell CHP	H2 Fuel Cell CHP
District heating network*	District heat enabled building – Shared Heat Pump – Large Scale Heat Pump DHN – Large Scale Natural Gas CHP DHN – Large Scale H2 DHN – Waste Heat DHN – Low Temperature DHN – Biomass DHN
Individual heat pump	Individual HT ASHP
	Gas Ab/Adsorption Heat Pump
	Individual GSHP
	Individual LT ASHP
Electric resistive	Individual electric resistive

Solar hot water	Solar hot water ('typical')
Hot water storage	Hot water storage
Heating controls	Basic
	Advanced

* Note that for DHN technologies the building component reflects the heat exchanger and internal building upgrades. The supply source for the DHN is a separate choice in the POM.

5.5.1.2. *Invalid combinations of Base values*

Certain combinations (pairs) of Base values are, for various reasons, invalid within the definition of the Base archetypes. The invalid pairs are listed in Table 5-8, however the user can further configure these. These rules are applied to generate the pre-defined master list of Base archetypes.

5.5.2. Base archetype SAP parameterization [HOM-001]

The Heat Loss Rate (HLR) and Thermal Mass (TM) for all of the base archetypes are simulated using a parameterised ISO 13790 compliant model based on the UK's Standard Assessment Procedure (SAP) developed for compliance with the European Energy Performance of Buildings Directive (EPRD).

A large number of additional variables are required in order to determine the Heat Loss Rate (HLR) and Effective Thermal Mass (ETM) of the Base archetypes through the SAP model. These variables are inferred from the set of Base variables using the English Housing Survey (EHS) dataset, SAP and Rd-SAP documentation and other literature sources. The full set of variables associated with each Base archetype required for SAP modelling can be configured by the user. Further, the user is provided with the list of additional required variables with the corresponding units and allowed values. Whilst it is not envisaged that the user will need to define additional Base archetypes, the user can therefore do this by completing the list of additional required data tables manually up to the point of use in HOM-003.

Table 5-8   Invalid combinations of Base values

1st Base variable		Invalid combination with 1st Base variable		Description
Base variable	**Value**	**Base variable**	**Value**	
Property type	All flats	Individual heating system element	Individual GSHP, Gas Ab/Adsorption Heat Pump, H2 Fuel cell CHP, Solar hot water, Individual Biomass Boiler	Assume flats have no suitable space for these heating systems or for biomass fuel storage
	All mid-floor flats	Loft insulation thickness band	None, less than 100 mm, 100 up to 199 mm, 200 mm or more	Assume mid-floor flats have no loft
	All flats	Number of storeys	2, 3	We define flat archetypes as a single storey

Floor area band	Floor area band 1 (less than 50 sqm)	Individual heating system element	Individual GSHP, Gas Ab/Adsorption Heat Pump, H2 Fuel cell CHP, Individual Biomass Boiler	Assume dwellings in this Floor area band have no suitable space for these heating systems or for biomass fuel storage

5.5.3. Base archetype (non-spatial clustering) [HOM-002]

A large number of possible base archetypes configurations exist (e.g. >10^4) even before differences in heating system are taken into account. As a result, it is not possible to run all of these in the more detailed EnergyPlus building energy simulation. HOM-002 module uses a k-means clustering algorithm to group similar base archetypes – i.e. those likely to have similar levels of overall space heat demand.

Mathematical 'cluster' analysis is used to group building types with similar key performance characteristics together, thus reducing the number of simulations that needs to be performed to a computationally tractable number *(note that mathematical clustering here is not related to the broader concept of spatial clustering used throughout the rest of the EnergyPath™ Networks Tool).*

'Cluster' analysis takes the full spectrum of possible user-defined building permutations and represents them by only the most material characteristics that affect their required outputs and the trade-off in choices seen by the POM. The most material outputs from HOM-001 which affect the energy consumption of the building and the diurnal load profile of the underlying space heat demand are:

- ▶ The thermal efficiency of the building fabric and its overall exposed surface area, which is expressed as a Heat Loss Rate (HLRs) in W/K as calculated by the SAP model. Note that the variation across the base archetypes covers the full spectrum of HLRs from existing through to those with whole house efficiency measures in place.

- ▶ The effective (short-term) Thermal Mass of the building which is expressed in J/K as

- ▶ The overall floor area of the property, as overall demand broadly scales with size all else being equal

Algorithmic decomposition is used to cluster simulated buildings into groups based on relative density and a user-specified number of target clusters. A variation of Lloyd's algorithm, also known as k-Means clustering is used. Each building archetype, expressed as a point within the HLR/TM problem space, can have multiple heating systems associated with it, giving a 3-dimensional grid.

Figure 5-4   Illustration of k-means clustering of base archetype HLR, thermal mass and floorspace
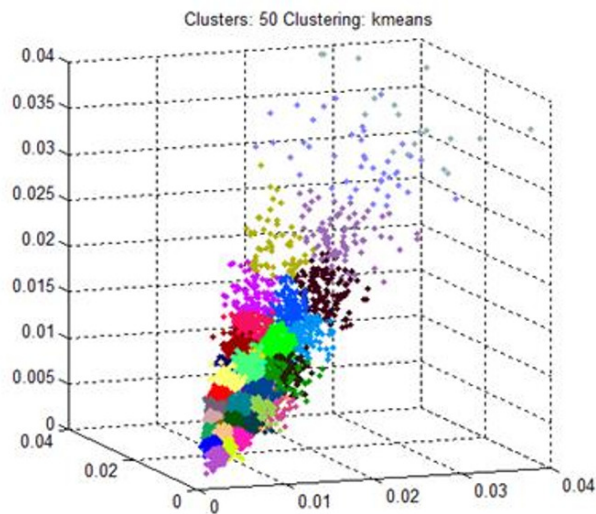
Section 12 - Appendix - Base archetype Simplification provides a more detailed description of this process and the testing undertaken, which led to the choice of k-means as the preferred algorithm.

5.5.4. Domestic building dynamic energy simulation (EnergyPlus) [HOM-003]

*The clustered groups of archetype from HOM-002 (reflecting only the physical properties of the building shell) are combined with the plausible set of heating systems to create **reduced archetypes** (~10^2). In HOM-003 each of these is run in EnergyPlus to create a detailed simulation of the operation of the building to create the final energy demand profiles for electricity, gas, etc.*

A dynamic building simulation tool (EnergyPlus) has been used for determining the possible operational profiles of individual heating systems and the consequent final energy demand for electricity, gas, or heat provided from a district heating network. The key outputs to the model beyond the HOM will be the sample curves for each characteristic day (typical and edge cases) and what the impact on peak power is.

The structure of the EnergyPath[TM] Networks Tool data architecture, which is designed to solve a large and complex optimisation problem, necessitates a trade-off between detail and complexity when modelling the "smart" aspects of the system, such as in-building energy storage and the choice of different operational profiles for load shifting at a network scale. These design choices are particularly relevant for electrical heating options such as heat pumps, as future grid-reinforcement costs are likely to be a function of the extent to which intense peaks can be mitigated or avoided through these types of option.

The ideal method of determining design choices of this nature would be to employ an optimised dispatch algorithm for all of the individual systems in every building on the network simultaneously, along with information about key constraints such as electricity pricing, and the electrical engineering constraints (voltage drop, thermal limits etc.) on the network. However,

the architecture for the EnergyPath^{TM} Networks Tool compartmentalizes all of these components separately for reasons of computational tractability.

Building operation is determined essentially off-model in HOM, network engineering constraints are modelled in the NAM, while the impacts of diversity and overall supply / demand balancing are determined in POM. This means that when carrying out dynamic simulation in HOM to determine the operating profile for a heat pump, the model has no ex-ante scenario information about what may be happening at a grid level to try and mitigate peaks, such as the use of price signals or active load control.

A number of key factors are material for the operational strategy employed by dwelling space heating systems and the resulting load profile shapes of the final demand vectors (gas, electricity, biomass etc.). Operational strategies are determined by considering a range of factors as outlined in the table below.

Table 5-9   Drivers of operational strategies

Key Optionality Drivers	Process Approach
External temperature	External temperatures can be determined for each characteristic day being modelled in the dynamic simulation. As part of determining the characteristic day information it will be necessary to model a longer period (e.g. a week before) to understand the implications for thermal mass.
Internal temperatures and hot water requirements	Target thermal comfort temperatures during occupied periods will be specified for individual archetypes and hot water demand profiles (in litres of hot water per half hour).
Heating system configuration	A limited number of configurations are possible for space heating and hot water delivery. These include: • Primary heating only • Primary heating with secondary hot water heating • Primary heating with secondary hot water and space heating Note that a particular configuration will be specified for each reduced archetype
Size of primary heating system	For each reduced archetype and each heating system type the system size can be matched to heat demand following relevant guidance from engineering best practice e.g. CIBSE design guides. In the majority of cases this will result in a limited range of system sizes that are capable of meeting the demand requirements for each archetype.
Size of secondary heating / hybrid systems	Secondary heating systems are typically sized to charge hot water storage when the primary system cannot reheat the cylinder quickly enough to meet demand requirements. Reheat times can be estimated for different archetypes based on the size of the property, the likely number of hot water using systems, and the size of the thermal store. Secondary heating systems are also used for "boosting" the space heating output when the primary system cannot meet demand. Hybrid-type systems where one of the heating devices has a low delivery temperature may be designed intentionally in this fashion. Note that the ratio of primary to secondary heating system size is a user determined input
Suitability of building structure for pre-heating	Pre-heating does not need to be considered for thermally lightweight buildings because there are no benefits to employing this strategy.
Size of storage tank	Domestic storage tanks come in a limited number of standard sizes. The size of the tank can be determined as a function of the primary and secondary heating systems and the size of the property, which are both effectively constraints on tank size. The operation of the storage is focused primarily on minimizing peak system demands, thereby reducing the size of the primary heating system.

The approach taken in the HOM for a given reduced archetype is to run a series of tests with an increasing size of primary heating system (where relevant for the archetype the secondary heating system size is defined as a ratio of the primary). The defined characteristic periods (several weeks) across the year are run reflecting all of the exogenous factors (external temperature, thermal mass and HLR of the reduced archetype, etc) and with the aim of meeting the user-defined target internal temperature profile and hot water demand profile across each day.

The number of hours where it is not possible to meet these targets is tracked and compared to a user-defined limit (e.g. 1% of hours). If the limit is exceeded the archetype is re-run with a larger heating system until this limit is no longer exceeded. Where storage is part of the reduced archetype being tested its operational strategy is to flatten the overall load on the primary heating system as much as is possible. As a result the iterative testing and operational strategy aims to try to minimize the size of the building heating system (and its associated capital costs) and the corresponding peak load on the wider system be it electricity, gas or from a district heat network.

Other modules in the EnergyPath™ Networks Tool will principally be using the final load profiles for the purposes of demand-supply matching in different time slices. EnergyPlus provides half-hourly diurnal profiles,

5.5.4.1. *Other non-heat related building electricity*

As indicated by the data inputs described in Section 5.4, the HOM addresses non-heat related electricity demand / supply largely through the application of exogenously determined inputs for consumption and profiling. These will be separated into:

- ▶ EV charging profiles, so that the assumptions around charging profile and implications for spatial cluster diversity in the POM can be more easily processed (see section 8.6.1.1), drawing on ETI/ESC data initially.

- ▶ All other non-heat related electricity (lighting and appliances), note that these are also fed into EnergyPlus for the purposes of calculating internal gains, which may reduce heating requirements.

- ▶ In addition, the HOM contains profile, cost and other relevant data for building-scale micro-generation technologies such as solar PV. These do not lead to additional archetype classifications, but are flags are applied to say whether the building is appropriate or not for the technology. The POM can then decide whether it is cost-effective to deploy these technologies (limited by data from the SAM on e.g. estimated roof area for each building in the area).

5.5.5. Pathway archetypes (costs and profiles) [HOM-004]

*The purpose of this sub-module is to combine the information on base archetypes and the simplified set of reduced archetypes from the previous sub-modules and convert this into a catalogue of costed **pathway archetypes** for use in the POM.*

- ▶ An individual **pathway archetype** reflects a set of simplified building characteristics in the starting year 2015 (e.g. a large inefficient building with a gas boiler[21]) and a set of defined choices for how that building will evolve to 2050. For example, in 2020 the building's efficiency is improved, whilst in 2030 the gas boiler is replaced with an ASHP and the efficiency is improved further. Beyond 2030 no further changes are made outside of like-for-like replacements at the end of each individual measure's life.

- ▶ For a single starting point there are multiple possible pathways to 2050 given different heating systems, efficiency options and the timing of such changes. The POM ultimately decides which pathway to use based on its costs, as calculated in this sub-module, and the impact of the pathways final energy demands as estimated through EnergyPlus (e.g. what is the determined profile of gas demand across the year).

- ▶ To manage the large number of possible pathways seen by the POM (ideally to ~$<10^4$) the user can define a series of simplifications or restrictions (in addition to those made at earlier stages of the HOM process). For example:

  - – The spectrum of possible transitions should not allow the move from low carbon back to high-carbon heating systems, or once the efficiency or a building has been improved it cannot be allowed to become less efficient.

  - – The maximum number of transition points (at which a major retrofit decision is made) is defined via global parameter "MaxPathwayDepth" in the table [mcf].[pi_GlobalParameters] in primary database.

  - – Individual efficiency options can be grouped into smaller number of retrofit 'packages'

The costing of a pathway archetype is a complicated process, however, the principles of this are consistent with the wider EPN tool in that the underlying calculation is built up from the most detailed level possible before being simplified. In this case it is managing the wide variation across the building stock – both in terms of what exists now and what this could change into – and convert these into a manageable set of the most material choices for use in the POM.

As part of this each pathway is inherently mapped to multiple base archetypes (via the reduced archetype simplification process in HOM-002), this mapping is itself further informed by what base archetypes actually exist across the local area via the building matching process in SAM-002. In addition, building and cluster level constraints are also considered to differentiate building stock that are similar in energy and cost terms (i.e. same final archetype) but have varying possible pathway transitions due to user defined constraints. Thus each pathway archetype, associated with a constraint (if applicable, otherwise NULL) is mapped to suitable base archetype stock.

---

[21] This is a stylised description, the specific simplification reflects the move from base to reduced archetypes described previously.

The costing process is first undertaken assuming that each base archetype follows the features of the individual pathway's evolution from 2015-2050 (i.e. the heating system and efficiency choices and their timings). The weighted average cost of retrofitting all base archetypes, which are mapped to the pathway, is used to create a single set of pathway archetype annualized costs on a £/building/year basis. This weighting is based on the stock of applicable base archetypes with the relevant constraint and for each cluster. This approach captures variations in underlying retrofit costs for base archetype buildings that may otherwise may look similar in energy terms. For example, retrofitting external solid wall insulation to a large terraced house versus solid wall insulation for a smaller semi-detached house.

In some cases the implied cost differentials of the retrofit choices are assumed to be so great that the reduced archetypes are purposefully split into an intermediate "final archetype" before being used to create the pathway archetypes. For example, the application of solid wall insulation in a pathway never occurs in conjunction with cavity wall insulation as although their affects may be similar in energy terms, averaging across base archetypes which have different proportions of each would give highly distorted pathway costs and impacting on the POM's decision making.

A more detailed description of how the pathway archetype costs are built up from the original base archetypes is outlined below.

5.5.5.1. Determining valid conversions between Base archetypes and the associated costs

Insulation retrofit measures

Retrofit events are described as conversions between Base archetypes. The key insulation-related measures that cause a conversion between Base archetypes are those that affect 'Wall Type', 'Glazing/Window Type', 'Loft Insulation Thickness Band' or 'Ventilation type' (Base archetype conversions based on a change of heating system combination are described separately below).

The set of key insulation-related measures that cause a conversion between Base archetypes, termed here 'primary measures', is shown in Table 5-10 along with the effect of the primary measure on the Base variables. A blank cell denotes no effect on the Base variable.

Other energy efficiency measures, applied in isolation, do not cause a conversion between Base archetypes. Examples of such measures are draught proofing and floor insulation. Even though those measures cannot be modelled here in isolation, it is desirable to capture the energy efficiency benefit of those measures by including them within retrofit packages. These measures, termed 'secondary measures', are shown in Table 5-11.

The table also shows which primary measure or measures each secondary measure is 'associated' with (as defined by the user). In this description, if a primary measure is installed in a dwelling, the associated secondary measure(s) will also be installed (though only if the secondary measure is suitable for the Base archetype in question, as described below).

Internal consistency is ensured by accounting for this association between primary and secondary measures during the imputation of the additional variables for SAP modelling (for

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   84/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

example, it is ensured that Base archetypes with solid wall insulation have an infiltration rate consistent with the 'Reduced infiltration 2 (whole dwelling)' measure having been applied).

Table 5-10   Effect of primary insulation measures on Base variables

Insulation measure	Wall Type	Glazing/Window Type	Loft Insulation Thickness Band	Ventilation type
Cavity wall insulation	Cavity (filled)			
Internal wall insulation	Solid (insulated, internal)			
External wall insulation	Solid (insulated, external)			
Double glazing		Double glazing		
Triple glazing		Triple glazing		
More than triple glazing		More than triple glazing		
Loft insulation			200 mm or more	
Mechanical ventilation				Mechanical

Table 5-11   Example of secondary insulation measures associated with Primary insulation measures

Primary insulation measure	Associated secondary insulation measure (1 = associated, 0 = not associated)			
	Floor insulation	Reduced infiltration 1 (draught proofing windows and doors)	Energy-efficient doors	Reduced infiltration 2 (whole dwelling)
Cavity wall insulation	0	0	0	0
Internal wall insulation	1	1	1	1
External wall insulation	1	1	1	1
Double glazing	0	1	1	0
Triple glazing	0	1	1	1
More than triple glazing	0	1	1	1
Loft insulation	0	0	0	0
Mechanical ventilation	0	1	1	1

Suitability of insulation retrofit measures

Individual insulation retrofit measures may not be suitable for all archetypes. For example, cavity wall insulation is not suitable for dwellings with solid walls, or for dwellings that already have cavity wall insulation.

Table 5-12 shows the default 'suitability table', which describes invalid combinations (pairs) of measures and Base values. A measure will be unsuitable for a Base archetype if the measure is unsuitable for one or more of the Base values defining the archetype. The user will be able to edit this suitability table within the model.

Packages of retrofit measures

For each model run, the user is able to define several (≈2-3) insulation retrofit packages made up of the individual primary measures above. Three example insulation retrofit packages are shown in Table 5-13. Secondary measures associated with the primary measures selected will be included in the package. It should be noted that there will always be defined a 'do nothing' package containing no insulation measures. This is necessary to allow archetype conversions consisting only of a heating system replacement (see below).

Packages are applicable for all archetypes, but whether a measure in the package is actually implemented depends on the suitability of the individual measure for the archetype. It could be that no measure in a package is suitable for a given archetype, and hence the package does not cause a conversion of the archetype. In such instance, no pathway exists for the given base archetype using such a package.

Table 5-12   Unsuitable Base values by insulation measure

Insulation measure	Property Type	Floor Area Band	Property Age	Number of Storeys	Wall Type	Glazing/Window Type	Loft Insulation Thickness Band	Ventilation type
Cavity wall insulation					Cavity (filled) Solid (uninsulated) Solid (insulated, internal) Solid (insulated, external)			
Internal wall insulation					Cavity (filled) Cavity (unfilled) Solid (insulated, internal) Solid (insulated, external)			
External wall insulation					Cavity (filled) Cavity (unfilled) Solid (insulated, internal) Solid (insulated, external)			
Double glazing						Double glazing Triple glazing More than triple glazing		
Triple glazing						Triple glazing More than triple glazing		
More than triple glazing						More than triple glazing		
Loft insulation							No loft 200 mm or more	
Mechanical ventilation								Mechanical
Floor insulation			1914-1944 1945-1964 1965-1979 1980-present					
Reduced infiltration 1 (draught proofing windows and doors)						Double glazing Triple glazing More than triple glazing		
Energy-efficient doors								
Reduced infiltration 2 (whole dwelling)								

Table 5-13   Example retrofit packages formed from Primary insulation measures (packages are user-defined)

Retrofit package	Primary insulation retrofit measures				
	Cavity wall retrofit option	**Solid wall retrofit option**	**Glazing retrofit option**	**Loft retrofit option**	**Mechanical ventilation retrofit option**
INS1	Cavity wall insulation	None	None	Loft insulation	None
INS2	Cavity wall insulation	External wall insulation	Double glazing	Loft insulation	None
INS3	Cavity wall insulation	Internal wall insulation	Triple glazing	Loft insulation	Mechanical ventilation

5.5.5.2. Heating system and heating controls replacement

A retrofit event may also include a replacement of the heating system (covering a combination of primary, secondary storage and controls). In that case, the resulting heating system combination is selected from the list of ≈10-50 defined in the step above. Whether a heating system replacement is suitable for the archetype is defined by the suitability of the resulting Primary heating system, Secondary heating system and Storage technology for the archetype (as described in Table 5-8). It is noted here that applying the suitability table will not preclude, for example, replacement of a low carbon heating system with a high carbon heating system. It is possible to introduce additional logic rules to allow the user to constrain the possible transitions between heating systems (for example, do not allow replacement of a low carbon heating system with a high carbon heating system) at a later stage in the modelling, during the construction of 'Pathway archetypes'.

The heating system replacement may also include an upgrade of the heating controls. The user is able to define heating controls in table [mcf].[si_HeatingControls] to choose whether e.g. No, Basic, Advanced or HEMS heating controls are available. Suitability rules are defined for heating controls, as shown in Table 5-14.

The list of Base archetype conversions resulting from the heating system replacement options and heating controls packages will be combined with the list of Base archetype conversions resulting from the insulation retrofit packages, to give the full list of possible Base archetype conversions. There will also always be defined a 'do nothing' heating system replacement option, in order to allow for archetype conversions which consist only of insulation measures.

Table 5-14   Unsuitable Base values by heating controls measure

Initial heating control status	Heating controls measure
Advanced or HEMS	Install Basic heating controls
HEMS	Install Advanced heating controls

5.5.5.3. Cost of archetype conversions

Unit cost database

The cost of each of the possible Base archetype conversions is then calculated. The user is able to input 'unit cost' data for each retrofit measure in terms of the following cost categories:

- ▶ Fixed capex (£ per dwelling)

- ▶ Variable capex (£ per unit)

- ▶ Fixed opex (£ per dwelling per year)

- ▶ Variable opex (£ per unit per year)

Table 5-15   Example of unit cost parameters

Measure	Fixed capex unit	Variable capex unit	Fixed opex unit	Variable opex unit
Cavity wall insulation	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
Internal wall insulation	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
External wall insulation	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
Double glazing	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
Triple glazing	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
More than triple glazing	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
Loft insulation	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
Mechanical ventilation	£/dwelling	£/unit	£/(dwelling.yr)	£/(unit.yr)
Floor insulation	£/dwelling	£/m2	£/(dwelling.yr)	£/(m2.yr)
Reduced infiltration 1 (draught proofing windows and doors)	£/dwelling	£/m	£/(dwelling.yr)	£/(m.yr)
Energy-efficient doors	£/dwelling	£/door	£/(dwelling.yr)	£/(door.yr)
Reduced infiltration 2 (whole dwelling)	£/dwelling	£/m	£/(dwelling.yr)	£/(m.yr)
Heating systems (All)	£/dwelling	£/kW	£/(dwelling.yr)	£/(kW.yr)
Hot water storage tank	£/dwelling	£/unit	£/(dwelling.yr)	£/(unit.yr)
Heating controls (Basic and Advanced)	£/dwelling	£/unit	£/(dwelling.yr)	£/(unit.yr)

Furthermore:

- ▶ Both the fixed and variable cost categories will be defined separately for each combination of Property Type and Property Age.

- ▶ The cost of each retrofit measure will be vintaged by a scalar 'inflator/deflator' for each year between the base year of the model and 2050

- ▶ A range of costs will be defined (Low, Central, High) in order for a Monte Carlo simulation to be performed in the POM (see section 9 on approach to uncertainty)

Calculating the cost of allowed Base archetype conversions

Using the unit cost data, the capex and opex associated with each possible base archetype conversion is calculated. To achieve this, the set of measures installed during each allowed Base

archetype conversion is inferred from the definition of retrofit packages and the measure suitability tables. The variables required to calculate the variable costs, such as the areas of fabric elements and the size of heating system required, are taken from the list of additional variables imputed for the SAP calculation.

In addition, the user has the option to apply a reduction to the fixed costs of a conversion. This option reflects the fact that the fixed costs for a package of measures might be lower than the sum of the fixed costs associated with the individual measures in the package (for example, to avoid double-counting the cost of re-decoration). This is achieved by defining a factor (between 0 and 1) for each possible combination of measures applied, by which the fixed cost of the package will be reduced.

After completion of this step, the full set of costs (capex and opex) of all possible Base archetype conversions, for all possible years of retrofit, is obtained.

5.5.5.4. *Formation of Final archetypes*

Rationale for forming Final archetypes

As described previously, processing limitations mean that the final set of options presented to the POM must be constrained in number. This is the purpose of the process of clustering ***Base archetypes*** according to Heat Loss Rate (HLR), Floor Area and Effective Thermal Mass (ETM) to create a much smaller number of **'*Reduced' archetypes energy states,*** which are combined with active heating system combinations to derive the **'*Reduced' archetypes***.

However, this process is likely to cluster *Base archetypes* which, although similar in energy consumption terms, entail very different retrofit costs. For example, an insulated solid wall archetype and an insulated cavity wall archetype may be similar in energy consumption terms, but the costs of converting to those archetypes from the corresponding uninsulated archetypes will be very different (as solid wall insulation is significantly more expensive than cavity wall insulation). It is important that these archetypes are separated in order to account for the fact that insulating cavity wall dwellings is more economic than insulating solid wall dwellings, information which should be carried through to influence the choice of optimum energy system pathway in POM.

Furthermore, it is necessary to keep certain *Base archetypes* separate on the basis of suitability of heating systems. For example, if a *Reduced archetype* were formed from a mixture of flats and detached houses, it would not be possible to define a binary suitability of the heating systems that are not suitable for flats.

Therefore, we apply a number of rules to disaggregate *Reduced archetypes* to form a new set of '*Final' archetypes*. It should be kept in mind that the purpose of this step is to account for the most important drivers in retrofit cost and suitability (i.e. both insulation measures *and* heating systems) whilst retaining a tractable number of Final archetypes with respect to the final set of options presented to the POM. Ultimately this is only an intermediate step as part of developing the resulting pathway archetypes seen by the POM.

Proposed methodology for forming Final archetypes

The proposed methodology for forming *Final archetypes* from *Reduced archetypes* is outlined in Table 5-16. The most important identified drivers of retrofit cost are the following:

▶ *Floor area band*

- The variable cost of all measures is dependent on the dwelling floor area, in most cases very strongly (such as wall insulation, glazing replacement and loft insulation).

- However, Reduced archetypes are already differentiated by Floor area band (along with HLR and ETM), so no additional step to separate Floor area bands is required.

▶ *Property type*

- Property type determines the number of external walls, and therefore the area of wall to be insulated and the area of glazing to be replaced, thus influencing the variable cost of those measures strongly. We note that as loft insulation is relatively low cost compared with solid wall insulation and glazing, and as a result it is not deemed necessary to separate mid floor and top floor flats when constructing final archetypes i.e. the same final archetype can have both mid floor and top floor flat as associated base archetypes. However the additional cost of loft insulation for top floor flats is considered when the final step of costing is performed at a base archetype level.

- Property type also, along with Floor area band, fully defines the suitability of heating systems.

▶ *Wall type*

- Wall type defines whether solid wall insulation or cavity wall insulation is required, and hence influences strongly the cost of retrofit.

▶ *Heating system type*

- The combination of primary, secondary, storage and heating control components. To make the final problem for the POM more tractable the user can also define the active final archetype heating system combinations, rather than assuming by default that all possible combinations of heating system components are available.

Table 5-16   Base variables used for disaggregating Reduced archetypes to form Final archetypes

Base variable	Number of groups	Group description	Base values contained in groups	Comment
Property Type	5	4 external walls	Detached	Separate by number of external walls as this is a significant driver of retrofit cost through wall insulation measures and glazing replacement. Separate flats and other Property types because certain heating systems are unsuitable for flats.
		3 external walls (not flats)	Semi-detached End terrace	
		3 external walls (flats)	Converted flat – top floor, end terrace Converted flat – mid floor, end terrace Purpose-built flat – top floor Purpose-built flat – mid floor	
		2 external walls (not flats)	Mid terrace	
		2 external walls (flats)	Converted flat – top floor, mid terrace Converted flat – mid floor, mid terrace	
Wall Type	2	Solid wall	Solid (uninsulated) Solid (insulated, internal) Solid (insulated, external)	Separated as Wall type (solid versus cavity) is a significant driver of retrofit cost (cost of wall insulation)
		Cavity wall	Cavity (unfilled) Cavity (filled)	

5.5.5.5. Formation of Pathway archetypes

Definition of Pathway archetypes

The last archetype class defined in the HOM, which is ultimately used in the POM is the '*Pathway archetype*'. A Pathway archetype describes a single set of retrofit choices made over the period from 2015 to 2050. As illustrated in Figure 7-1, Pathway archetypes are defined by:

- ▶ The initial Final archetype as well as the possible building or cluster level constraint class to apply across the modelling horizon.

- ▶ The Final archetype resulting from the first retrofit event, and the year of the first retrofit event

- ▶ The Final archetype resulting from the second retrofit event, and the year of the second retrofit event

- ▶ The Final archetype resulting from the third retrofit event, and the year of the third retrofit event

- ▶ etc.

Final archetypes F_n include the building fabric state and the heating system combination. Retrofit events could in general involve fabric insulation measures, a heating system replacement, or both.

The purpose of Pathway archetypes is to pass to the POM a tractable set of retrofit options accurately costed over the period 2015 to 2050. Given that, depending on the user choices made earlier in the model flow, there will be of the order of 10^4 Final archetypes, each of which can be converted to approximately 10-100 other Final archetypes, simple arithmetic shows that the total number of possible Pathway archetypes is extremely large. It is therefore necessary to specify a set of logic rules for the construction of Pathway archetypes.

Figure 5-5   Diagram showing a Pathway archetype defined by its retrofit events

F_0 :	Initial Final archetype
t_i :	Year of ith retrofit event
F_i :	Final archetype after ith retrofit event
τ_i :	Lifetime of heating system in F_i
N :	Number of years prior to end of life that replacement of heating system is allowed
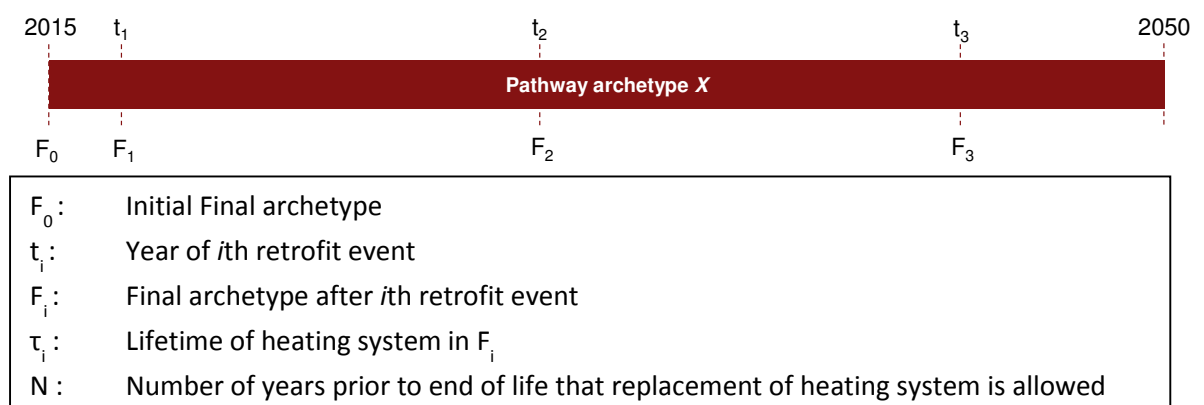
It is important to note that under this representation the POM is still making a cost optimised choice looking across the entire tool pathway against other time dependent variables, there is no notion of making a decision 'up-front' for the whole pathway. At the limit, choosing from all possible combinations of final archetype conversions along the pathway, or 1 archetype pathway from a list which represents the cross product of all possible transitions across the pathway

effectively gives the POM the same choices. The rationale for using the *pathway archetype* approach is that it allows the POM to see a more accurate representation of the costs of upgrading buildings and allows the user final control over the granularity passed to the POM, in a similar manner to the control over network options, spatial clusters, timeperiods and timeslices.

Logic rules for Pathway archetypes

We define the following logic rules in order to constrain the number of Pathway archetypes passed to the POM. It is envisaged that the user will have control over these logic rules to define the range of Pathway archetypes available.

1. Fabric insulation measures are only implemented when the heating system is replaced[22]

2. t_1 is in the range between 2015 and $(2015 + \tau_0)$ to reflect the initial age distribution of primary heating system (single user choice)

3. t_i is in the range between $(t_{i-1} + \tau_{i-1} - N)$ and $(t_{i-1} + \tau_{i-1})$ where N is the number of years prior to end of life early replacement of primary heating system is allowed (default 5 years, the lower the number the smaller the number of pathway archetypes)

4. Primary heating system in F_i is allowed as a replacement for the primary heating system in F_{i-1} (see Table 5-17) if:

   – Primary heating system in F_i is not in a higher 'carbon intensity' category than the primary heating system in F_{i-1}

   – The replacement does not involve a transition from a network-based heating system to an individual heating system, or from a heat pump-based individual heating system to a non-heat pump-based individual heating system

   – The user can define the restrictions combinations, the more restrictions the fewer the number of pathway archetypes

5. Primary heating system in F_i is allowed to be built in the year t_i
   – User is able to prevent certain heating systems being built, or force certain heating systems to be built

As set out in logic rule 4, the user is able to set the allowed primary heating system transitions. Table 5-17 outlines an example of a proposed structure for defining the allowable transitions.

Here, we define two axes on a heating system 'choice matrix', which are defined for each year in the period 2015 to 2050. The first axis distinguishes 'not zero carbon-compatible' systems and 'zero carbon-compatible' systems; the second distinguishes 'individual, non-heat pump-based', 'individual, heat pump-based' and 'network-based'. Retrofit events involving a primary heating

---

[22] This constraint could potentially be lifted, but that is likely to result in an intractable number of Pathway archetypes.

system conversion moving 'left' or 'upwards' across the table are not allowed, as described in Table 5-18.

It should be noted that the two axes of the heating system choice matrix can be defined according to the user's preference; the structure shown here is flexible. The structure shown here also allows the user to prevent or enforce the building of specific heating systems in specific years, by removing a heating system (or all other heating systems) entirely from the choice matrix.

Table 5-17   Heating system categorization to determine valid Pathway archetypes

	Not zero carbon-compatible	Zero carbon-compatible
Individual, non-heat pump-based	**Category 1** Individual Gas Boiler Individual Oil/LPG Boiler Individual Biomass Boiler	**Category 2** Individual H2 Boiler H2 Fuel Cell CHP Individual electric resistive
Individual, heat pump-based	**Category 3** Gas Ab/Adsorption Heat Pump	**Category 4** Individual HT ASHP Individual LT ASHP Individual GSHP
Network-based	**Category 5** Large Scale Natural Gas CHP DHN Biomass DHN	**Category 6** Shared Heat Pump Large Scale Heat Pump DHN Large Scale H2 DHN Waste Heat DHN Low Temperature DHN

Table 5-18   Allowed heating system transitions

Initial heating system category	Allowed Final heating system category
1	1, 2, 3, 4, 5, 6
2	2, 4, 6
3	3, 4, 5, 6
4	4, 6
5	5, 6
6	6

5.5.5.6.   *Cost of Pathway archetypes to pass to the POM*

The final step in 1-HOM is to calculate the annualised cost associated with each *Pathway archetype* for each year over the period 2015-2050. Since each *Pathway archetype* is defined by its component *Final archetypes* and the year of each retrofit event, this can be achieved by calculating

separately the contribution of each component *Final archetype* to the annualised cost in each year. We note that this will allow for a proper treatment of sunk costs in the case of replacement of technologies before the end of their economic lifetime.

Each conversion to a *Final archetype* is made up, in general, of a set of insulation measures and a heating system replacement. Since those components are likely to have different technical and economic lifetimes, the contribution of the insulation and heating system elements will be calculated separately before being combined.

5.6. Output validation checks

A number of output validation checks are proposed to be carried out on the HOM outputs as part of the application of the tool itself:

▶ Comparing ISO13790 outputs against the same properties modelled in other ISO13790-compliant packages, such as the Macro Distributed Energy (Macro-DE) housing sub-model

▶ Comparing dynamic outputs against observed load profiles where available.

6. Spatial Analysis Module (SAM)

6.1. Overview

The role of the Spatial Analysis Module (SAM) is to create a detailed representation of the local area that contains sufficient spatial data to enable the Networks Analysis Module (NAM) to assess the costs and feasibility of potential network reinforcements and new build, and the Pathway Optimisation Module (POM) to make well-informed choices on the optimum energy strategy for the area and elements within it (e.g. buildings, networks etc.). As part of its internal processes the SAM needs to be able to simplify the spatial representation by aggregating into larger (user defined or auto generated) regions to make the optimisation process tractable.

In order to provide the Networks Analysis Module (NAM) with the inputs it requires, the SAM is designed to develop geographic information system (GIS)-based layers containing detailed spatial and topographical information about the local area, synthesise a representation of the electricity distribution network serving the area and match domestic building archetypes (defined in the Household Options Module (HOM)) and non-domestic activity classes to each individual address in the area. The matched building archetypes are also passed back to the HOM to define the potential pathways for these addresses (i.e. the potential evolution of heat system and efficiency measures associated with the archetypes) which are then costed based on the mix of stock represented by each pathway.

Figure 6-1　　　**SAM context**
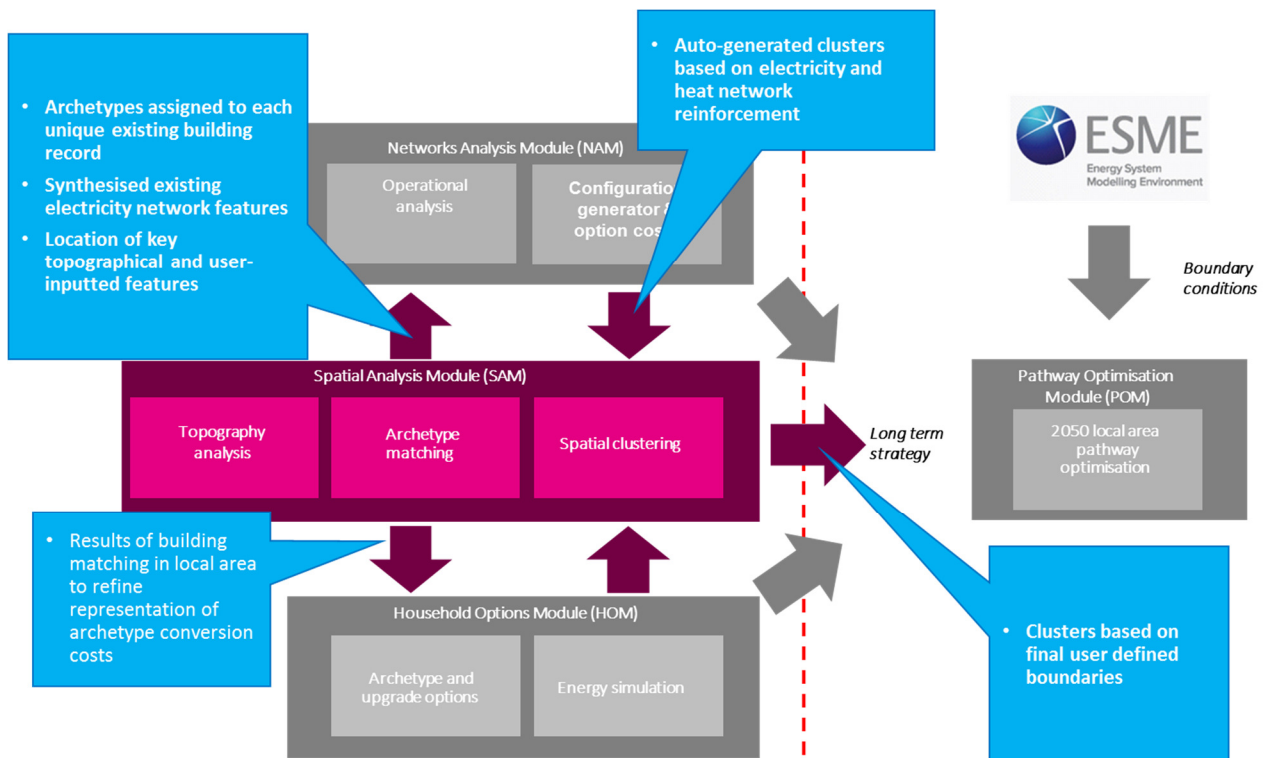
The topological network information and building archetype information is passed to the NAM, such that the NAM can calculate electricity and heat network upgrade cost curves, and gas network extension or hydrogen repurposing costs for individual network elements (e.g. pipes,

feeders, substations etc). The outputs of the NAM are then combined with the detailed spatial information and used to aggregate the cost of individual network elements into homogenous 'clusters' where the constituent zones are spatially contiguous and have similar network upgrade costs (in the automated clustering mode). The user then has the option to modify these clusters and the final spatial information of these clusters is passed to NAM to generate the final cluster network upgrade cost curves.

The spatial information associated with each of these clusters and the number of each building archetype contained within each of them are then passed onto the Pathway Optimisation Module (POM) for whole system energy cost optimisation.

In addition to existing spatial data, the SAM also provides an interface for the user to enter future spatial data (i.e. that will occur as part of a scenario) or future options that could be utilised or impact on the pathway design via the integration of ArcGIS into the EPN tool. This can include, but is not restricted to geothermal or biomass availability, conservation areas restricting archetype upgrades, possible sites for large heat source sites or embedded generation, sites for potential new build housing, existing PV uptake, EV uptake projections etc. This information is stored in the GIS layers and passed to POM for the pathway analysis.

The main steps undertaken in the SAM are summarised below and described in more detail in subsequent sections.

Define building stock data

The SAM allows the user to define building attribute and technology uptake at address level via ArcGIS interface. This includes the ability to define:

1. Known attributes for existing and new build domestic and non-domestic addresses

2. The existing uptake of PV per domestic address

3. The projections for existing and future EV uptake

4. Known sites of demolition

Assign building archetypes to addresses

The SAM uses address level data on the location and type of buildings and combines it with other data sources containing detailed building attribute characteristics to develop a detailed stock database that defines the key attributes of every residential and non-domestic building in the area. This information is used to match individual domestic addresses to their nearest representative base archetypes defined in the HOM. For non-domestic buildings they are matched to a single "activity class" identifier such as a commercial office or industrial warehouse, with associated energy benchmarks and potential retrofitting options (combined efficiency and or heating system options).

Synthesise a representation of the electricity distribution network

The SAM synthesises a representation of the electricity LV/HV network. GIS layers containing individual building locations and road network data are combined with data on the location of

distribution substations (which it is assumed the DNO can provide) and used by the SAM to synthesise a representation of the existing electricity distribution network. This uses user-defined connections between HV & LV substations or between LV substations and the buildings they service where available. Otherwise, standard GIS software network analysis packages are used to calculate the shortest road route between building and substation or LV to HV substation. The detailed topographical information of the network, including lengths of individual LV/HV feeders and the connectivity of buildings to different distribution substations, is sent to the NAM for heat and electricity network upgrade cost calculations.

Generate clusters

Within the SAM the user can define spatially contiguous "clusters", based on the more detailed information in the underlying zones and network components (e.g. individual electricity feeders and substations). This creates several clusters in the LA, where each cluster provides a more simplified representation for *all* building and network choices within the cluster (electricity, heat network, gas/hydrogen) to be passed to the POM. Manual and semi-automated processes (both using the ArcGIS interface) are available to help the user define appropriate cluster boundaries. The NAM (see section 7) also contains a fully automated process to help determine the set of clusters balancing appropriate network representation and the need to avoid artificially dividing key elements of electricity network connectivity.

Additional local information

The SAM also saves additional information about other local features of the energy system as defined by the user using external data sources (e.g. planning permission for new builds and embedded generation in a LA). This could represent existing features, those which are already planned for a given future timeperiod, or a set of constraints on possible options (e.g. a new CHP can only be located in zone X or Y, but Z due to planning restrictions). This additional spatial data is sent to POM for inclusion in the pathway analysis.

6.1.1. Key spatial elements

The key aim of the SAM is to collect and synthesise detailed data on buildings, network topology and resources at a zone level. The starting point is a 'unit' or an individually identifiable energy system feature within a local area (e.g. each individual building floor area polygon will be mapped using OS data). A 'zone' represents the highest geographical granularity at which detailed network cost calculations are performed in the NAM, typically a street which contains a number of building units, aggregated to a single road node.

The higher the geographic granularity, the more accurate will be the network upgrade cost calculations. However, increasing granularity results in larger numbers of costed zones in a LA area and increasing computational processing demand by the POM as it optimises for every zone. Therefore these zones are aggregated into clusters and represented by aggregate network upgrade costs, in order to ensure computationally feasible optimisation within the POM.
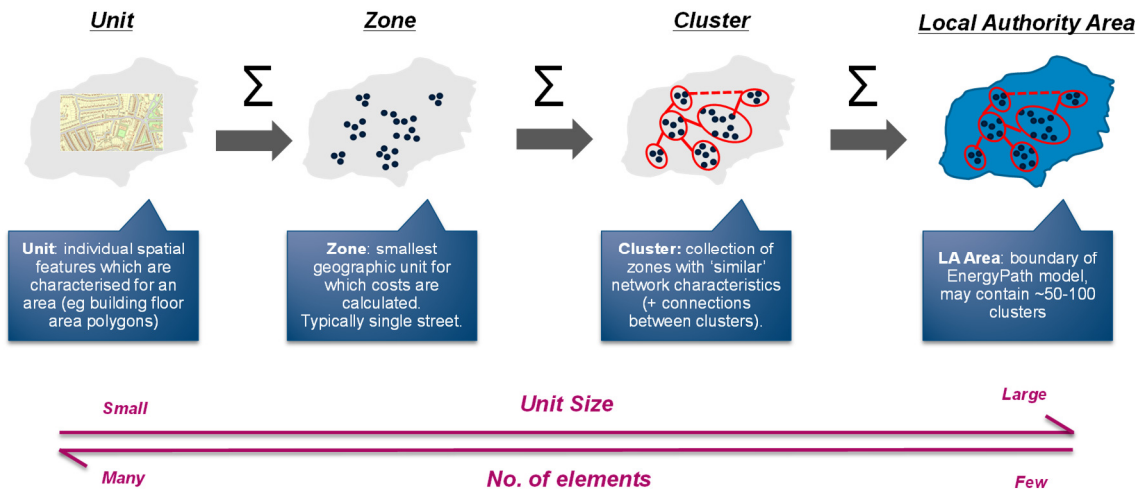
6.1.2. Zone definition

Zones represent the highest geographical granularity at which building, network, resource and topographical information is collected in SAM to be passed to the NAM. A single zone or aggregation of multiple zones could be included within a user defined cluster, provided the network topology rules are not broken. The NAM uses data on the network connectivity and network element characteristics (e.g. list of unique LV feeders, buildings connected to each, lengths of individual links in each feeder) to develop network upgrade option costs.

In order to facilitate the accurate definition of network upgrade costs in the NAM, zones are based on the smallest component of the network, which is aimed to be an individual street level LV feeder as indicated in the diagram below. However, it is important for the SAM to reflect how these zones are connected together at high levels of aggregation such that the overall network topology features are respected within any spatial analysis (see section 7.3.5 for further discussion):

A. Street level feeder that connect buildings on roads to the nearest distribution substation;

B. Low voltage network (Distribution substations that have several LV feeders)

C. 11kV feeder (HV feeders that connect several distribution substations to the nearest primary substation);

D. Primary substation that has several HV feeders.
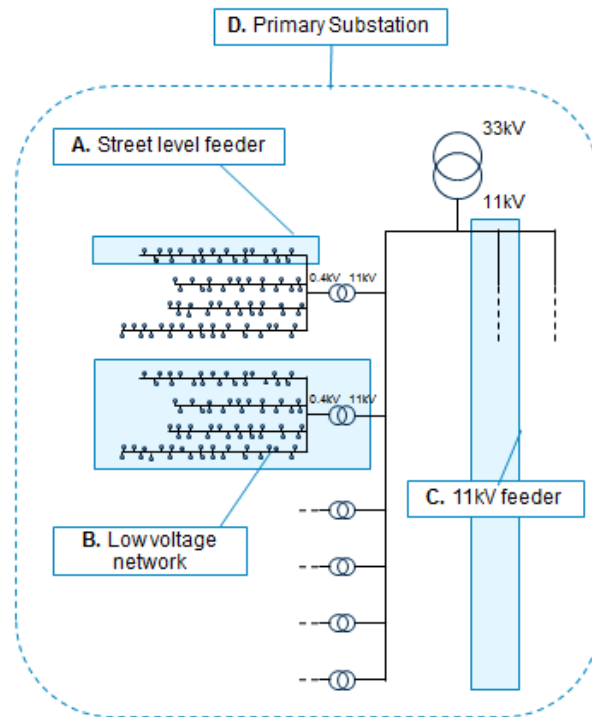
6.1.3. Cluster definition

Clusters represent grouped zones that are spatially contiguous, have homogenous network upgrade costs and are connected to the same electricity network component i.e. LV feeders connected to the same distribution substation, distribution substations (and all of their LV feeders) connected to the same HV feeder, HV feeders (and all of their distribution substations) connected to the same primary substation or multiple HV substations. The purpose of clustering is to create fewer spatial zones that have homogenous network upgrade costs, in order to minimize the computational burden on the POM.

The standard approach to defining a cluster is based around all buildings connected to an individual HV substation (multiple HV substations can also be included in the same cluster) as a way to avoid creating clusters which artificially cut across the electricity network connectivity. Where more spatial granularity is needed, for example if within the large cluster there exist areas with very high and very low heat network development costs, it is possible to sub-divide the larger cluster into two separate clusters. As a result the POM optimization would more clearly see the different costs of heat network reinforcement in the two smaller clusters as opposed to average costs across the larger cluster, which would refine its decision making.

Sub-dividing the clusters further must still respect the electricity connectivity rules shown in Figure 6-3. For example, it could be split by HV feeders connected to the HV substation (these new clusters could still contain multiple HV feeders as long as an individual HV feeder is not split), or even down to the LV substation or LV feeder level.

In addition to manually drawing the boundaries in ArcGIS, the SAM contains logic to allow the user to draw around individual objects (HV/LV substations or HV feeders) and automatically create a

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   102/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

cluster containing all individual network elements (and connected buildings) downstream of this (e.g. for an HV feeder all buildings on this directly, all connected LV substations and all buildings connected to these LV substations).

6.1.4. Module diagram

The diagram below shows the key logic / process steps for the SAM, the equivalent for the full tool covering all modules is shown in section 11.

Figure 6-4 Key SAM logic / process steps
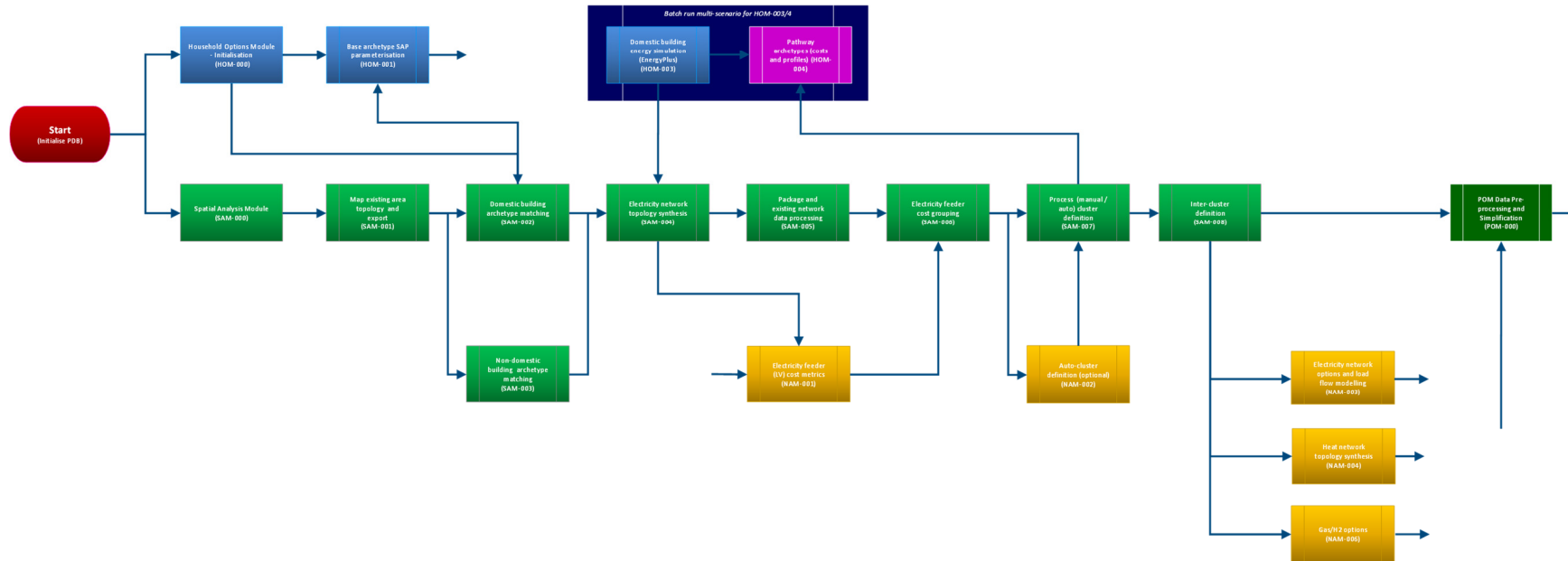
6.2. Key outputs

The main output of the SAM is a list of clusters, based on user defined boundaries or auto generated in NAM to represent regions with similar network upgrade cost functions, and its constituent network elements (LV /HV feeders and LV/HV substations) and buildings. Simple locational information is associated with each of these clusters, the number of each building archetype contained within each cluster and any technology suitability flags associated with the local topography or user-inputted data, are also passed onto the Pathway Optimisation Module (POM) for whole system energy cost optimisation.

In order to enable the NAM to carry out its functions, the SAM also produces a series of intermediate outputs, as listed below.

6.2.1. Building archetype database

The SAM outputs a record associated with each individual address (both domestic and non-domestic) within the LA area, with a building archetype assigned (i.e. matched) to each address, referencing the domestic building archetypes defined in the HOM and non-domestic activity classes which are provided by the user and referenced directly in the SAM matching process.

6.2.2. Synthesised energy network

Electricity network

SAM outputs a representation of the local electricity distribution network to the NAM. Where real-life data on the layout of the electricity network is not available from local DNOs, synthetic data on the network layout is generated within the SAM. Specific outputs generated to represent the network connectivity within the LA area include:

- ▶ ID of nearest road node for each domestic and non-domestic address

- ▶ Grid level of each domestic (assumed LV) and non-domestic address (based on activity class or peak demand)

- ▶ ID of LV/HV feeders

- ▶ Road nodes served by each LV/HV feeder

- ▶ Upstream LV/HV substation for each LV/HV feeder

- ▶ Length of links that constitute each LV/HV feeder

- ▶ Overlap between LV feeders

- ▶ Road node of each LV/HV substation

If the DNO can provide further information on which specific buildings are connected to each LV/HV feeder and LV/HV substation the network synthesis will incorporate this connection information into the final generated network. The additional network connectivity data could include one of the following:

- ▶ Connectivity of each address/LV substation to its upstream LV/HV substation and the connected LV/HV feeder

- ▶ Connectivity of each LV substation to its upstream HV substation and the connected HV feeder

- ▶ Connectivity of addresses to LV substations and the connected LV feeder where they lie within their buildings

The routing in all cases is still assumed to follow the shortest distance on the road network between two points (e.g. building and substation). However, it may be that the additional DNO information states that an individual building is not connected to the closest substation (the default assumed when fully synthesising the network), but one further away.

NAM uses this information to perform detailed load flow analysis and thus generate electricity network upgrade costs curves.

6.2.3. User-inputted spatial topology, resource and future planning

SAM creates a GIS layer with topology data for the LA. It also acts as the interface to store data on

- ▶ Local resource availability (e.g. geothermal, biomass, embedded generation)

- ▶ Future new building developments, planned demolitions and additional local characteristics (e.g. conservation zones) based on data provided by the user

- ▶ Existing or potential distributed generation sites for electricity and heat networks (e.g. a new CHP plant)

- ▶ Existing uptake of PV in domestic buildings

- ▶ Existing and future projections of EV stock

This data is passed to both NAM / POM, either in the form of flags associated with each cluster (e.g. altered technology suitability), or as individual elements for analysis.

6.2.4. Zone definitions

This is based on the smallest component of the electricity network to be costed in the NAM, as well as the outputs of the local electricity distribution network synthesis, the SAM outputs zone definitions assigning a zone ID to each component of the synthesised electricity distribution network.

The SAM would also provide further zone context information which could be used for the network costing within the NAM, such as the local area type (e.g. rural, urban, sub-urban) based on user defined data via the cluster definition layer as well as the network component types (e.g. above ground, underground).

Table 6-1   Output data summary

ID	Data field	Destination	Purpose	Data granularity	Uncertainty parameters
1	Matched domestic building base archetypes and non-domestic activity classes and floor areas	HOM / NAM / POM	HOM pathway archetype costings use information on types of existing base archetypes and non-domestic activity classes in the area NAM peak demand estimation for load flow modelling given existing buildings and potential for electrification POM building archetype conversion and non-domestic retrofitting option choices and supply / demand balancing	Locational information, building archetype or activity class, distance to nearest feeder and substation (including IDs for both) and zone ID for each building within the LA area.	-
2	Synthesised electricity network	NAM	Provides data required by the NAM to generate electricity network upgrade cost functions	Locational information, as well as Individual LV and HV feeder ID and length, and IDs for each additional network component, including LV and HV substations (and which feeders they are connected to).	-
3	Gas network road lengths	NAM	Provides data required by the NAM to generate cost for connecting off gas buildings to the gas grid, as well as for hydrogen repurposing	Lengths of road that are connected to gas grid and also which are off gas grid	
4	User-inputted data on building stock constraints	POM	Modifying the technical potential for building upgrades based on constraints defined for the user defined cluster boundaries e.g. to restrict or force pathways for domestic and non-domestic addresses.	Locational information provided for individual buildings or for selected geographical areas using GIS layers	-
5	User-inputted data on resource availability	POM	Determining the suitability of technology implementation based on local resources e.g. biomass, geothermal sites etc.	Locational information about the geographic extent of local resources, provided using GIS layers or excel based databases	-
6	User-inputted data on future development plans	POM	Providing the spatial and temporal information for future infrastructure plans e.g. new build, embedded generation etc.	Locational information provided using GIS layers for the LA and temporal information provided as a separate input	-

7	User-inputted data on existing embedded generation	POM	Determining the net loads on the network by POM	Locational information on the type and size of generation using GIS layers or Excel databases for the LA	-
8	Zone definitions	NAM	Provides data required by NAM to perform initial heat network based clustering	The zone definition is provided in the form of a zone ID assigned to each electricity distribution network component. Secondary outputs will include simple locational information, the total road lengths within each zone and the archetypes associated with each road.	-
9	Cluster definitions	NAM/POM	Needed by NAM to derive the aggregated electricity and heat network cost metrics. POM will use the clusters to perform system level optimisation	Simple cluster locational information, building archetype numbers within each cluster and any technology suitability flags associated with either local topography or user-inputted data.	-
10	Zone site context	NAM	Used to adjust the base network cost options e.g. if they are urban, suburban, rural. Based on user defined data	Zone level information	-

6.3. Key functional requirements

The key functional requirement of the SAM is to process a series of complex local building, electricity network and geographical datasets, and to combine these into a limited number (e.g. less than 50) of simplified 'clusters', containing all the information that is required for the POM to carry out its optimisation calculations. To do this the SAM is closely linked to the NAM with two-way input/output data flows, given the importance of spatial topology on the understanding of network reinforcement and new build options

In order to do this, it must be able to perform a number of processing steps, as defined below.

6.3.1. Understand the location of existing building archetypes

In order for the POM to make choices as to the balance between network upgrades or new build options and individual building upgrades, the SAM must output simplified data on the type and location of buildings within the LA area.

In order to do this and due to the lack of detailed building-level data for every building within a LA area, the SAM must combine detailed GIS-based local area representations, with other local or higher-level datasets on the attributes of the building stock in order to assign a distribution of standard building archetypes (defined within the HOM) to the address level data. The output is a database of standard building archetypes matched to each domestic address in the area, for which the POM can reference the HOM in order to understand relevant energy demand patterns. For non-domestic buildings each address is matched to activity class (and associated energy benchmark pathways) based on OS data on the base function, Scat or other external data sources, which the POM can reference along with the floor areas to determine the annual and peak energy demands.

The process used to develop the database of standard building archetypes must be flexible enough to accommodate existing or future local known building information.

6.3.2. Understanding the energy and retrofitting costs associated with non-domestic buildings

Modelling the retrofitting of non-domestic buildings offers additional challenges over doing the same for the domestic stock. Non-domestic buildings are significantly more diverse than dwellings in terms of their energy use, even within buildings of nominally the same type. For example, buildings identified or classified as offices range from low-rise naturally ventilated structures that have often been converted from former dwellings, industrial or agricultural buildings, to purpose built commercial tower units. The information technology, lighting, heating, ventilation and cooling (HVAC) strategies and systems employed in office buildings can vary widely with age, occupancy and tenancy, sometimes even on different floors or wings of the same structure. The result is that even when the form and function of two non-domestic buildings are similar, the energy demand and the available pathways for energy efficient retrofit are diverse and resistant to characterisation or archetyping.

At the time of writing, there are no large scale studies on the performance and costs of non-domestic retrofitting of UK buildings that can draw broad, robust generalisations about costs and or resultant performance improvements in a comparable manner to the way domestic building

base archetypes are created in the HOM, and the subsequent process for simplifying this detailed set of archetypes into smaller set for dynamic energy simulation in EnergyPlus.

Dynamic energy simulation has therefore not been used for non-domestic buildings. Instead a simpler benchmarking approach has been chosen:

▶ The user defines a series of non-domestic "activity classes" for which reasonable energy benchmark data is believed to exist and maps these activity classes to OS basefunction (i.e. type of building use), VOA Scat or other external source of building use descriptions. This is currently based on the UCL "Carb" energy benchmark dataset[23] with annual kWh/m2 figures and Valuation Office agency ratings dataset.

▶ The SAM matches the building to the activity class (based on OS basefunction, Scat or other external sources if available) and combines this with further information on floor area and number of storeys (from other GIS datasets) to infer total floorspace of the building and hence the floorspace of each address within the building.

▶ Finally the user can define a series of explicit retrofitting options (as well as the starting option) for each non-domestic activity class with different energy demands and costs that reflect potential efficiency improvements and/or changes in heating systems. The user also defines the valid transitions between these options (i.e. shift towards a less carbon intensive option).These are used to determine possible option transition pathways for each activity class. The user also defines cost categories for the retrofitting options of each activity class (i.e. low, central, high) and the energy consumption bands (MWh/year). The retrofitting option decision is made at a building level, however a building can contain activity class with varying starting and possible options as well as cost category, hence the collective cost category and retrofitting option pathways of each non-domestic building is based on the most dominant homogenous activity class i.e. those activity class that collectively have the largest share of the building energy demand. This is used to assign non-domestic buildings to unique 'groups '. A group is a collection of non-domestic buildings which are similar in terms of following attributes:

– Spatial location (i.e. are in same cluster)

– Cost category (based on the dominant homogenous activity classes)

– Demand category (based on total building energy demand)

– Retrofitting option pathways (based on the dominant homogenous activity classes)

These groups, their associated transition pathways and costs are then passed as choices to the POM in an analogous manner to domestic pathway archetypes described in 5.5.5.

The accuracy achieved by benchmarking energy consumption on an area basis is dependent on the quality of the input data, which are derived from empirical observations. Existing datasets are widely acknowledged to have shortcomings relating to the age of the collected data and the

---

[23] http://www.ucl.ac.uk/carb/

number of observations. This is an area for ongoing research and the EPN tool data architecture is flexible enough to accommodate new information as it becomes available.

It should be noted that the above approach does not preclude the use of more realistic demand profile or cost data as this becomes available from surveys within a given LA area, particularly for non-domestic public properties such as libraries, schools leisure centres or council offices.

6.3.3. Synthesise electricity network features where data is limited

In order for the NAM to evaluate and cost network upgrade options within the LA area, the SAM must first synthesise a representation of the existing local electricity network – to include an understanding of which buildings are connected to which low voltage (LV) and high voltage (HV) substations and the length of feeders between them.

The analysis must use available data for inputs, including basic information provided by local DNOs (e.g. substation location) and known local topographical features (e.g. building locations, road locations and lengths, etc.).

Again a flexible approach has been adopted to allow for the potential availability of detailed local electricity distribution network data, provided by local DNOs. The user can enter detailed address level data of network connectivity to represent the full network coverage (LV and HV), some LV connections where the substation is located within certain buildings or just between HV and LV substations. The user defined data is then incorporated where available while the remaining buildings are connected to the LV/HV substations using the synthesis approach i.e. connecting each load to its nearest substation along the road network.Allow the user to input a wide range of spatially-related data

The SAM must allow users to input location-specific data, which will affect the choices made in the POM. This may include:

▶ Data that could affect technology suitability for a given area, e.g. local biomass/geothermal resource, conservation areas, etc.

▶ Data that may prevent the implementation of certain building upgrade decisions, e.g. cluster level constraints that restrict the potential pathways available to domestic buildings

Ultimately these factors will be aggregated to the final cluster definitions and impact on the available choices for that *cluster* as a whole as seen by the POM (e.g. restricting certain types of building retrofit in a cluster that is known to reflect a conservation area or forcing the build of a planned energy centre).

6.3.4. Zone definitions

In order for the NAM to perform a heat network cluster analysis and to define electricity network reinforcement cost functions based on contiguous sections of the electricity distribution network, the SAM must first define a series of zones, which are used to break the LA area into manageable areas. Zone definitions must be passed to the NAM to enable the network costing analysis.

6.3.5. Clustering zones

In order to constrain the computational burden for the optimisation in the POM, the SAM must combine zones into fewer, larger clusters of zones with homogenous network upgrade costs, and which do not break the fundamental electricity network connectivity described previously. The boundaries of these clusters can be defined by the user using ArcGIS interface in a manual or semi-automated manner based on the electricity network topology.

6.4. Key inputs

Primary inputs

SAM uses several key external datasets to develop an enriched database of existing and future spatial elements in the LA. These datasets are listed below:

1. Ordinance survey (OS) datasets of the local authority including:

    - Mastermap layer containing detailed topographical data (e.g. location of buildings and their areas)

    - AddressBase Premium with location of individual addresses, their associated Unique Property Reference Number (UPRN) and adjacent roads

    - Integrated Transport Network (ITN) layer with detailed information about the existing road network

2. Data to informing matching of domestic base archetypes or non-domestic activity classes to individual buildings in the local area. For example,

    - English Housing Survey (EHS), Living in Wales (LiW) and Scottish Housing Condition Survey (SHCS) dataset for residential sector at government office region (GOR) level

    - Xoserve dataset for the availability of gas at postcode level

    - GeoInformation dataset (if available) for the age and type of domestic addresses

    - Valuation Office Agency (VOA) database for non-domestic buildings at address/postcode level

3. DNO data on the local electricity distribution network, including:

    - Location for distribution and primary substations

    - Number of LV and HV feeders per distribution and primary substations (if available)

4. If available mapping of HV to LV substations, buildings to specific substations or all addresses to their upstream substations to refine the network synthesis. Local user defined spatial inputs, to include, but not restricted to:

- Existing uptake of PV in domestic buildings as well as the existing and future projections of EV stock. Resource availability used to determine the suitability of technology implementation based on local resources e.g. biomass, geothermal sites, availability of "waste heat" from local industrial plant etc., defined using ArcGIS interface.

- Existing embedded generation used to determine the net loads on the network by POM, defined using ArcGIS interface.

- Future development plans that provide the spatial and temporal information for future infrastructure plans e.g. new build sites, defined using ArcGIS interface.

5. Non-domestic building energy and cost-related data

- Energy and cost benchmarks per m^2 for each non-domestic activity class reflecting existing and future demands (Arup)

- Non-domestic building height, combined with information about building footprint from OS sources to estimate total non-domestic building floor area

Secondary inputs

SAM relies on a number of databases that are generated internally within the EnergyPath^{TM} Networks Tool, in different sub-modules. These are:

1. List of archetypes and their attribute characteristics as defined in the HOM

2. Possible cluster boundary definitions generated via an automated process in NAM

3. List of non-domestic buildings and their attribute characteristics (demands, retrofitting options and associated costs) coming from the primary DB

The tables below describe what is technically required within the tool and how it is used. The relative importance of the primary data inputs and other factors such as costs and licensing restrictions, particularly where they must be purchased from an external provider, are discussed in the original deliverable (D4) Data Acquisition Plan. It should be noted that the ESC currently maintain the master set of input data for the current v2.1 EPN tool and this is not documented here.

Table 6-2   Input data

Data field(s)	Input type	Purpose	Granularity	Uncertainty parameters	Coverage	Source	Limitations
Building footprint area, address, UPRN, building use,[24]	Primary	Matching the building to external databases with building attribute characteristics	Defined for residence/hereditament at address level	-	All addresses in the UK	GIS layers of OS MasterMap, AddressBase Premium and potentially Points of Interest	-
Age, tenure, type, location, fuel, glazing, loft type/insulation, wall type/insulation, heating technology, number of storeys	Primary	Identifying the detailed domestic address attributes	Defined for sample residential dwellings at GOR level	Data available for sample buildings (around 16k for England)	All GOR in the England, Wales and Scotland	E.g. English Housing Survey , Living in Wales , Scottish Housing Condition Survey)	Building characteristic distribution in the sample data at GOR level is applied at address level
Type of non-domestic hereditament and floor area	Primary	Classification of non-domestic addresses for matching with activity classes in SAM	Defined for all hereditaments at address level	-	Whole of UK	OS / Valuation Office Agency	-
Non-domestic building height	Primary	Calculate total building floor area (rather than just footprint)	Defined at building level	-	Whole of UK	OS dataset	-
Existing heat source, embedded generation, etc.	Primary	Allow POM to optimise heat network	Defined for all known sites in the LA	-	LA	User	Bespoke layer generated by SAM for editing via ArcGIS interface
Planned new-build developments	Primary	Allow POM to consider new	Defined for all planned sites in the LA	-	LA	User	Bespoke layer generated by

---

[24] Ordnance Survey experimental height dataset may become available in future to help determine non-domestic building height

		developments in its optimisation of available options					SAM for editing via ArcGIS interface
Cluster level constraints	Primary	Identifying spatial constraints for application of retrofit/technology measures to building stock	Defined for all known constraint areas within the LA	-	LA	User	Bespoke layer generated by SAM for editing via ArcGIS interface
Road node coordinates, road length	Primary	Synthesising the LV and HV electricity network	Defined for all roads within the LA	-	All roads within the LA	GIS layer of OS Integrated Transport Network	
Distribution and Primary substation coordinate/address	Primary	Synthesising the LV and HV electricity network	All substations defined in the LA under consideration	-	Defined by the user for the area under study	User/DNO	
Electricity network topology data	Primary	Refine network synthesis process using knowledge of existing connections	Mapping of - HV to LV substations - Buildings to individual HV/LV substations	-	LA	DNO	Need to include UPRN to facilitate mapping of individual buildings with OS datasets
Domestic archetype definition	Secondary	Matching UPRN to an existing base archetype from HOM	All feasible combinations of attribute characteristics	-	Residential sector	HOM	-
Non-domestic archetype definition	Secondary	Matching UPRN to a user defined activity class in SAM	Activity classes mapped to energy benchmarks	-	Commercial sector	User	
Gas distribution network data	Primary	Location of gas transmission connections used to determine gas flows for cluster demands	Road Node	-	GB	National Grid operated DNOs from MAP viewer software, other DNOs provide directly	-

Non-Domestic Activity Classes	Primary	Equivalent of non-domestic archetypes for building matching	~100-200 activity classes, depending on data availability for subsequent items and for matching	-	National	User (based around UCL Carb dataset)	-
Non-Domestic Activity Class demand	Primary	Annual and peak demands for different energy vector	Each unique activity class	-	National	User (based around UCL Carb dataset)	-
Non-Domestic activity class retrofitting options	Primary	Used to define the starting and potential future retrofitting options that represent the combined effect of shifting heating technology and improved fabric efficiency	Multiple options per activity class with data by timeperiod and within year timeslices year	-	Per activity class	Arup	-
Non-Domestic activity class option costs	Primary	Provides cost of moving between different possible options for the valid pathways for each activity class	Per each possible option of activity class with data by timeperiod and within year timeslices year	-	Per activity class	User	-

6.5. Key SAM logic / process steps

6.5.1. Spatial Analysis Module [SAM-000]

This step initializes the SAM database with the key inputs e.g. the list of possible base archetypes, the cluster types and constraints, non-domestic retrofitting options, cost levels and demand levels etc. This step also creates the blank GIS layers that can be edited by the user later to define address level data for local building, technology uptake, generation etc.

6.5.2. Map existing topology area topology and export [SAM-001]

The first step is to map and configure all of the available spatial data for the local area as outlined in section 6.4 including:

1. Stock attribute layer (for defining building attributes en masse)

2. Demolitions layer (to define regions with planned demolition and the year)

3. PV layer (to define existing uptake in domestic households)

4. EV layer (to define existing and future projections of EV uptake)

5. Existing DHN (to define pipe sizes for regions with district heating)

6. Existing LV feeder (to define the size for LV feeders)

7. Existing HV feeder links (to define the size for HV feeder links)

8. Technology package (to define locations of potential embedded generation, energy centres etc)

9. Clusters (to define the boundaries of LA areas to be aggregated for optimisation in POM)

10. Cluster connection (to define the heat and gas/H2 connections between clusters for energy flows)

11. Meshing (to define the connections in electricity network for load sharing)

6.5.3. The user can edit these layers via ArcGIS interface to define spatial data at address level for the LA.Domestic building archetype matching [SAM-002]

SAM uses OS AddressBase Premium and MasterMap layers to populate a database of existing building stock with UPRN, address, floor area and building type. This provides detailed spatial information for all the residential buildings in an LA. At this stage the user is able to define attribute characteristics (include building age, type, tenure, location, fuel, glazing type, loft insulation, wall construction, number of storeys, heating technology, etc.) at an address level, where known, for residential buildings in a LA. This can be done by manually entering the attribute values for every known address via the ArcGIS interface in the 'Building' layer (for individual address) or 'StockAttributes' layer to define them en-masse.

The following sections assume that certain national level datasets are available, however, the underlying process of matching is generic such that data at different levels of spatial aggregation, covering different building archetype characteristics could be included within the process.

Next, GeoInformation is used to populate age and type attributes for residential addresses where available and where these have not already been defined by the user. This is done by matching the UPRN between the OS and GeoInformation datasets.

With the age and type now defined for all the buildings in the LA, Xoserve data (at postcode level) is used to identify the addresses with access to gas fuel , if it is not already defined by the user. *Any user-inputted data that has already been provided at an address level has precedence over these aggregated higher level datasets*.

Lastly, addresses with missing data on age, type and fuel as well as additional data on glazing level, loft insulation, wall construction, number of storeys and heating technology is populated for all the dwellings, by applying GOR-level distributions from the EHS, LiW or SHCS datasets, for all dwellings where these attributes have not been defined by the user. This distribution of wall type, glazing and loft insulation are applied based on the frequency of records with attributes of age, type, tenure, location and fuel in the used GOR level dataset.

Through the above steps, the detailed building attributes for all residential dwellings are defined at address level. These attributes are based on datasets available at address level as well as higher geographic granularity (e.g. postcode / GOR). The distributions of the attribute characteristics still conform to the aggregated distribution, based on the already populated attributes at lower geographic granularity e.g. GOR level distribution of wall type, glazing and loft insulation combinations is applied based on the known attributes of age, type, tenure, location and fuel type. Since the chosen attributes and the characteristic values are matched with those used in HOM, the resulting archetypal classification of residential building is always available in HOM database.

6.5.4. Non domestic building archetype matching [SAM-003]

The SAM uses OS AddressBase Premium and Mastermap layers to populate an existing building stock database with UPRN, floor area and type of use (based on high-level non-domestic classification of activity class).

The user can define non-domestic use type at address level, where known in an LA. This can be done by manually entering the use type for every known non-domestic address using the ArcGIS interface. The following datasets are used at address level, in order of hierarchy, to define the building use classification of all buildings (e.g. office, hotel, school etc) if not already defined by the user:

1. VOA data, based on matching the Scat category to the relevant activity class

2. Other external dataset (e.g. GeoInformation), based on matching the description to the relevant activity class

3. OS Addressbase, based on matching the base function to the relevant activity class

These classifications allow the annual and peak energy demands to be calculated for non-domestic sector.

Figure 6-5 **Schematic of the step-wise approach to defining attributes for each individual UPRN and assigning archetypes**
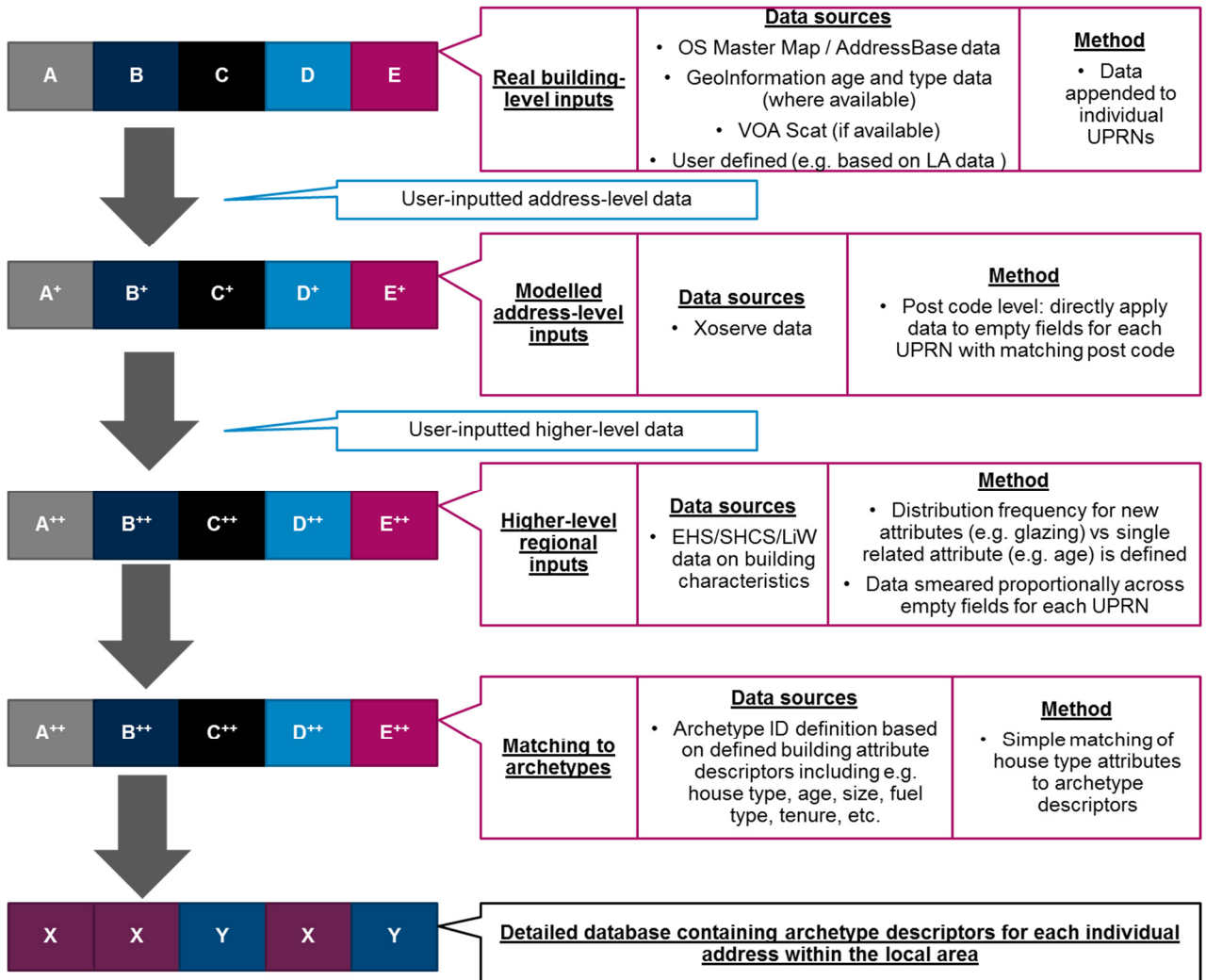
6.5.4.1. Validation

SAM performs a validation check to ensure the archetypal classification process results in an accurate representation of energy use in the LA based on historical metered energy demands at various levels of geographic granularity.

The implied annual energy (heating and electricity) demand for the LA is calculated based on the archetypal classification of the buildings and the energy demand defined in the HOM and SAM. This could then be compared with the annual energy consumption (and peak demand on characteristic days, if available) obtained from the actual metered data to inform the user of the difference between implied and actual metered data.

DECC currently release this data at LSOA level for domestic consumption and MSOA level for non-domestic, but discussions with DECC during Stage 1 have indicated that it may be possible to obtain this at postcode level, subject to commercial confidentiality in the non-domestic sector and resource to support the processing of the data.

If this difference is significant, this would signal to the user that additional, more localised data would benefit the analysis, in order to more accurately reflect the condition of the local building stock, e.g. LA-specific data on the distribution of glazing, wall type and loft insulation in the local area, to better reflect the LA rather than aggregated data at GOR level from EHS/LiW/SHCS. This should result in a better representation of LA building stock, matching more closely with the metered consumption data.

6.5.5. Electricity network topology synthesis [SAM-004]

SAM synthesises an electricity network for the LA to represent the connectivity of each building to an upstream substation and of LV substations to the HV substations. If detailed DNO data on electricity network connectivity is not available, then the network is synthesised based on the location of buildings, road network and DNO data on location of substations. Firstly, OS AddressBase Premium and ITN GIS layers are used in GIS software to identify the closest building road nodes. GIS network analyst tool functionality is used to connect each building road node and LV substationto its nearest LV and HV substation based on the road network. This information of the nearest upstream substation and the route used is processed to develop detailed network connectivity elements.

In addition to the location of the substations, the data set from DNO may also contain information on the peak load or number of connections of LV and HV feeders per each LV and HV substation respectively.

SAM also has the functionality to detailed network connectivity data, where available, to generate similar network element outputs that are consistent with the actual network. This data can be entered for:

1. Full network connectivity i.e. the LV/HV feeder used to connect each address/ LV substation to its upstream LV/HV substation

2. HV network only i.e. the HV feeder used to connect each LV substation to its upstream HV substation

3. Partial LV network i.e. the identification of those addresses which are connected to LV substations that are located within their buildings and only serve it

SAM uses the above DNO data on network where available and uses the closest distance routing for the remaining addresses to develop network elements that cover the whole LA.

Figure 6-6   **Electricity network synthesis process steps**
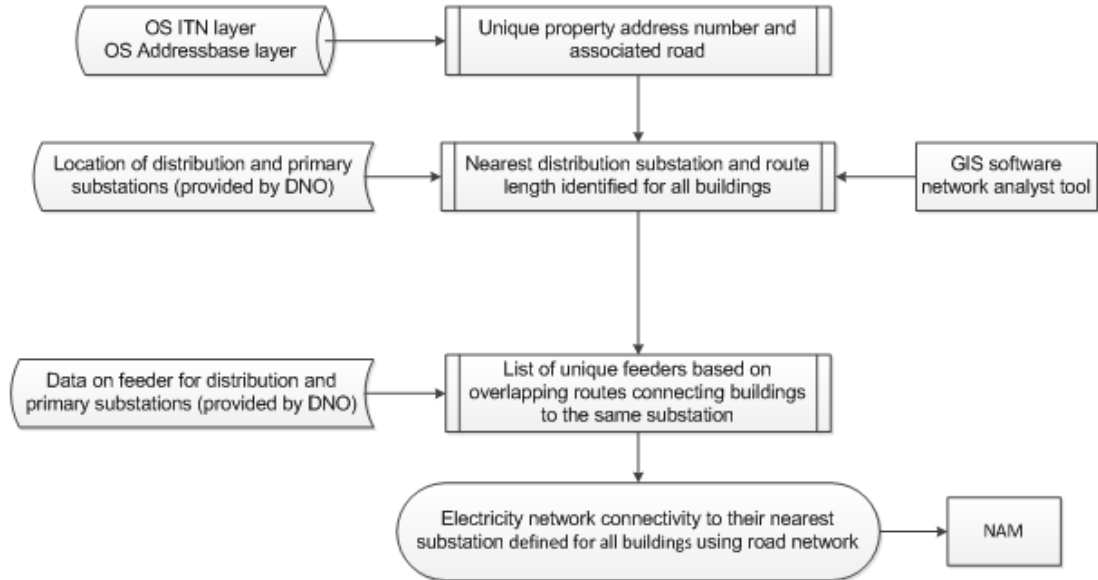
Figure 6-7   **Example of network synthesis**

Thus a detailed database is created which is passed to the NAM for electricity network cost calculation. This contains the following information each building in the LA:

4. Address level archetype ID

5. Closest road node ID

6. LV/HV feeder connected road node ID

7. LV/HV feeder connected upstream LV/HV substation ID

8. Length of links (network component with constant load) that constitute LV/HV feeders

9. Overlapping lengths between LV feeders

6.5.6. Package and existing network data processing [SAM-005]

The SAM allows the user to input location-specific data, which will affect the choices made in the POM. This is done through a simple ArcGIS based user interface, allowing the user to edit automatically generated layers. Data inputted may include the following information:

Data	Purpose
Building or technology option constraints	*Determining the technical potential for building upgrade based on constraints like conservation areas, listed heritage sites etc, or restricting possible sites for new embedded generation. These are defined in the 'Clusters' layer, up to 20 constraints can be defined per cluster.*
Resource availability	*Determining the suitability of technology implementation based on local resources e.g. biomass, geothermal sites etc. This is defined in the 'TechnologyPackage' layer.*
Existing embedded generation and storage	*Determining the net loads on the network by NAM / POM. This is defined in the 'TechnologyPackage' layer*

6.5.7. Cluster input layer creation [SAM-006]

This sub-module provides the user with the input layers on the electricity network topology showing the network connectivity at LV feeder, LV substation, HV feeder and HV substation levels. This, along with the layer showing auto generated cluster definition by NAM, provides all of the information layers in ArcGIS to inform the user defined clustering process in SAM-007.

6.5.8. Process (manual/semi-automated / automated) cluster definition [SAM-007]

Clustering is performed in the EnergyPath^{TM} Networks Tool to produce spatially contiguous regions with homogenous network upgrade costs.

An initial clustering process is performed in the NAM-002 that considers the cost of upgrading both electricity and heat networks. This aims to balance the representation of both heat and electricity network costs within the cluster boundary, but is led via the exploration of possible heat network configurations, given the greater degrees of freedom for a new build network. This process is described further in section 7.3.5.

These auto generated cluster definitions are passed to SAM and displayed to the user as a GIS layer for initial guidance, in addition to layers showing the network connectivity at varying levels i.e. each LV feeder, LV substation connected region downstream, HV feeder roads, HV feeder connected region downstream and HV substation connected region downstream. The user then has the option to define the final cluster definition in the 'Clusters' layer based on these or modify them to break some large clusters or combine some smaller clusters, however these boundaries must obey the electricity network topology rules i.e. a cluster must include all downstream elements and can only include two different elements on the same level (e.g. LV substations) if they have the same upstream element (e.g. HV feeder). These clusters can be defined by drawing boundaries around the whole region, or only at the highest network elements e.g. LV/HV substation nodes or HV feeder roads and all network elements downstream are also included, while also creating clusters of each element upstream.

This module also defines potential pathways for heating system transition in the non-domestic buildings and the annual cost for these pathways. The user is required to define the starting option and the potential options available for each activity class as well as the cost banding. The non-domestic buildings are then grouped based on cluster, similar option transitions, cost banding and demand banding.

6.5.9. Inter-cluster definition [SAM-008]

The user can define heat and gas/H2 connections between clusters to represent potential for energy flows between them. This is done by editing the 'ClusterConnection' layer via ArcGIS interface. In addition the user can also define potential meshing options for the load sharing of electricity network.

6.5.10. ArcGIS Spatial results [SAM-009]

The results of the optimisation in POM are then processed at cluster level and passed back to SAM for visualisation. These results include:

▶   Cluster connection metrics i.e. the energy flows between clusters

- ► Cluster metrics e.g. uptake of heating technology, insulation package, network upgrade costs etc.

- ► Polygon labels i.e. the technology installed for each polygon

- ► Heat network size i.e. the installed heat network within each cluster and the size of pipes to meet the thermal demands

7. Network Analysis Module (NAM)

7.1. Overview

The Network Analysis Module (NAM) provides the detailed analysis for quantifying potential network upgrade and new build options (their characteristics and costs) associated with an increase in energy peak demand requirements across the energy vectors under consideration: electricity, heat, gas and hydrogen. These options are then passed to the POM as part of the pathway analysis.

The key interactions between NAM and the other modules are shown in Figure 7-1 below.

Figure 7-1　　　**NAM context**
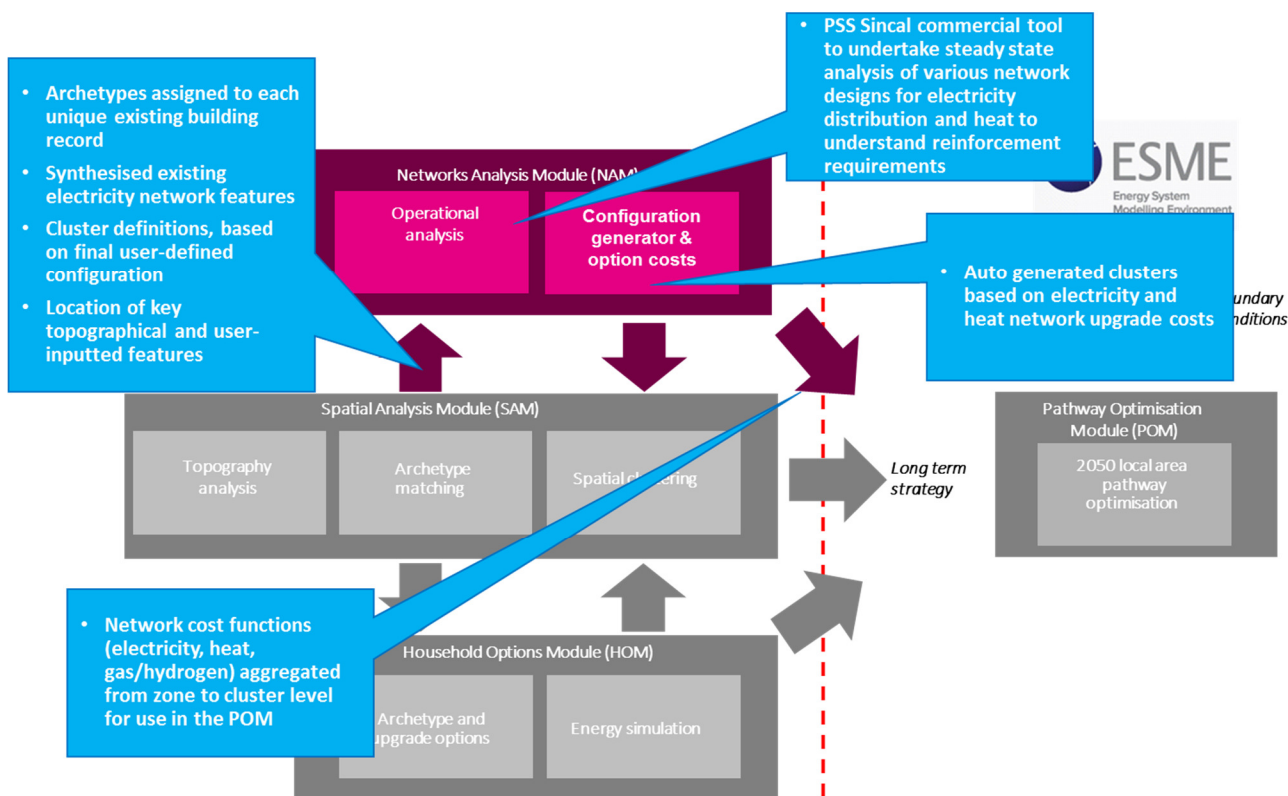
As shown in the Figure above, NAM contains two key components, each with a distinct set of requirements:

▶ The 'operational analysis' tool, which is mainly responsible for performing load flow studies using PSS Sincal[25] to simulate the operation of energy networks (existing and potential) in the area under investigation. The NAM interacts closely with the SAM and together they help to define the final area cluster boundaries, trying to ensure the cost

---

[25] http://w3.siemens.com/smartgrid/global/en/products-systems-solutions/software-solutions/planning-data-management-software/Pages/overview.aspx

of network options are as representative as possible across all vectors within a cluster, and that the total number of clusters is tractable within the POM optimisation.

- The NAM also contains an 'auto-clustering' process whereby the final cluster definition is generated automatically (by default this is user defined through the SAM) considering the trade-offs across both the electricity and heat networks (i.e. respecting electricity connectivity rules and considering the representation of network costs)

- Potential local heat networks are synthesised in the NAM using a similar process as for electricity network synthesis in the SAM

▶ The 'test options generator and costing tool', which is responsible for generating costed options for all zones, clusters and inter-cluster connections within the area under investigation. These cost functions are based on network upgrade or new-build cost functions depending on the reinforcement package to be considered;

In summary, the main requirements from NAM are the following:

▶ Materiality of load flow modelling: NAM performs load flow modelling studies across the different energy vectors (electricity and heat[26]) to ensure accurate allowable capacity limits are used when calculating reinforcement cost functions. All network constraints which can materially impact load flows (and hence costs) across each energy vector must be taken into account.

▶ Ensuring that existing and potential energy networks are accurately represented: NAM ensures that all existing and potential energy networks are represented to a sufficiently high accuracy in terms of their physical and geographical characteristics, together with an accurate spatial and temporal representation of potential demand and generation options.

▶ Producing cost functions for a range of network reinforcement options: NAM outputs cost functions for different network options for each energy vector. Cost functions compare network reinforcement costs and maximum allowable capacity (load/generation) which are then passed to the POM for system optimisation. Moreover, the NAM ensures that the number of network options tested by the scenario costing tool, as well as the number of network reinforcement options sent to POM, is tractable computationally.

▶ Materiality of network component costs: Similarly to the point above, NAM ensures that the cost functions produced for a given network reinforcement option accurately reflect the costs of component upgrades or additions, accounting for local factors (geographical, economic etc.) which may affect these.

▶ Producing information to inform the setting of clusters with specific boundary constraints: This can take the form of metrics passed back to the SAM (e.g. estimated

---

[26] Comparable analysis is not currently incorporated for gas given the existing capacity and limited likelihood of network upgrades (outside of some smaller extensions and/or repurposing to hydrogen).

electricity reinforcement costs) to help inform the user-defined cluster definition, or alternatively undertake an automated process to suggest appropriate cluster boundaries:

- It is important to note that the overarching process used to define clusters is trying to ensure that the network costs within them are as well represented as possible for **all** energy vectors and their associated networks and do not unfairly bias the solution in POM in a particular cluster (e.g. towards district heat ahead or electricity or hydrogen)

7.1.1. Module diagram

The diagram below shows the key logic / process steps for the NAM, the equivalent for the full tool covering all modules is shown in section 11.

Figure 7-2   Key NAM logic / process steps
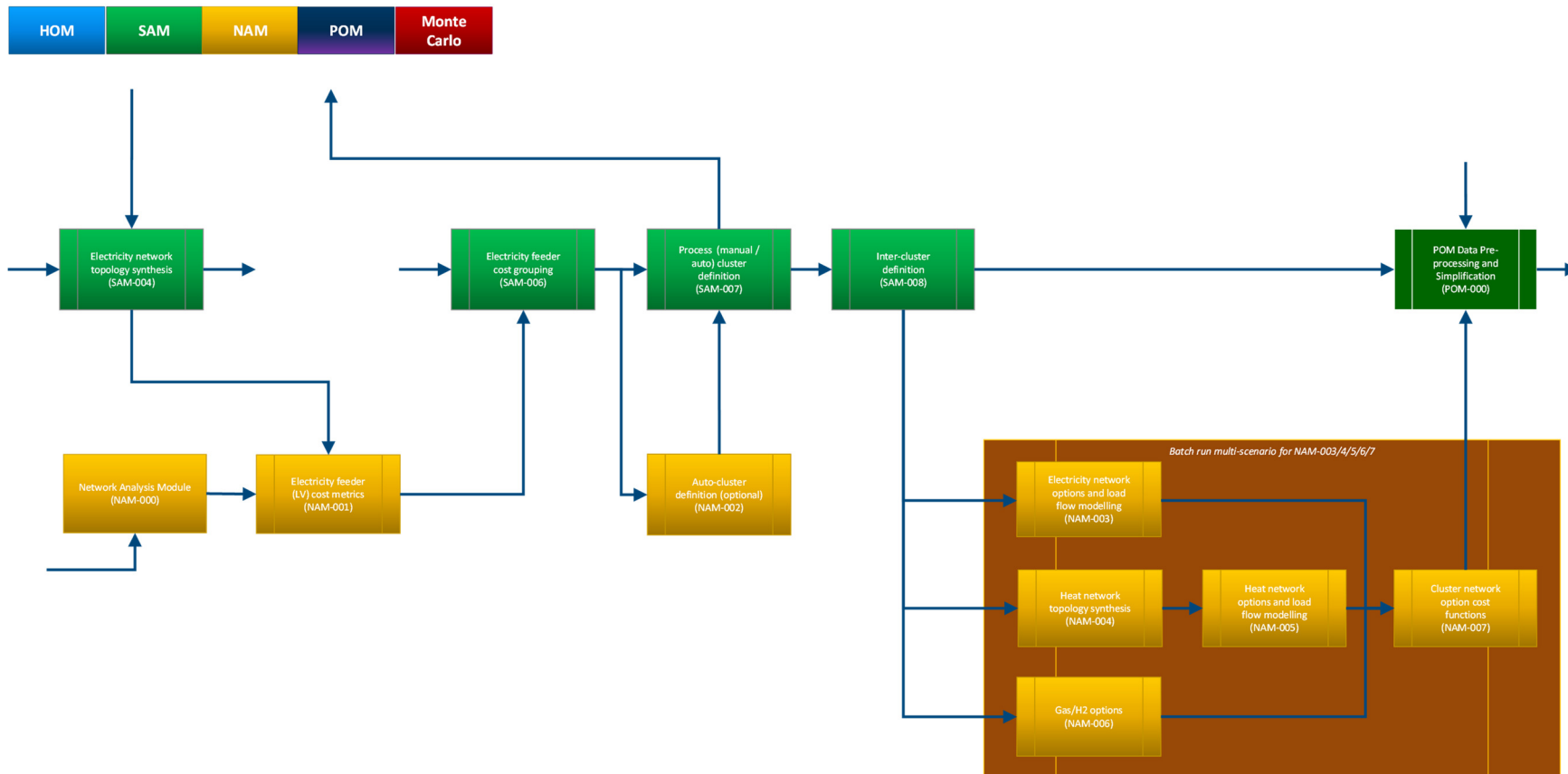
7.2. Key outputs

This section describes the key outputs from NAM, where they are used in the overall solution and for what purpose.

The key outputs from NAM can broadly be summarised in two categories:

1. NAM outputs related to cluster definitions: whilst the SAM is responsible for defining the final cluster and inter-cluster connections that are used by POM for the optimisation of the overall energy system the NAM provides supporting information to help define the clusters. The following outputs are included in this category:

    a. Simple metrics derived from zone-level electricity cost functions that were used to illustrate which zones have similar reinforcement costs to guide the user definition of clusters in the SAM[27]

    b. An 'auto-clustering' process, which provides a suggested set of final cluster definitions to the SAM, which the user can visualise in ArcGIS. The automated process considers the maximum number of desired clusters (given impacts on performance in the POM), electricity network connectivity rules and reinforcement costs across heat and electricity networks[28]

2. NAM outputs related to cost functions: The NAM is also responsible for outputting cost functions at the intra-cluster and inter-cluster level for each energy vector. These cost functions compare reinforcement/new build costs and maximum allowable capacity for each reinforcement option and are used by POM for optimisation of the energy system. NAM is also responsible for producing cost functions at the zone level (which are then used for determining intra-cluster cost functions) however zone cost functions are not sent to POM as they are not at a high enough level of aggregation. The following outputs are therefore included in this category:

    a. Intra-cluster cost functions for all energy vectors;

    b. Inter-cluster cost functions for all energy vectors.

NAM outputs are summarised in Table 7-1 below.

---

[27] This mode of clustering was used originally before the implementation of a more sophisticated clustering algorithm allowing for clustering all network components downstream from a chosen asset like a HV substation (see section 6.5.8). It is commented out in the latest version of EPN.

[28] I.e. trying to avoid suggesting clusters which contain significant variations in reinforcement costs, which would be averaged out by inclusion in the same cluster and potential distort the POM's decision making.

Table 7-1   Output data summary

Data field	Destination	Purpose	Data granularity	Uncertainty parameters

Auto-clustering definitions	SAM	Create suggested set of final cluster definitions for the SAM, which the user can also visualise in ArcGIS.	- A list with all buildings and relevant network components that have been identified as being part of a cluster. This includes:
			  ▪ Building ID;
			  ▪ Network component ID;
			  ▪ Cluster ID.
			The cluster definitions are timeperiod independent as they do not change over time	

Electricity zone cost metrics (Optionally used)	SAM	Provides user with information on groups of zones with similar electricity reinforcement costs to inform used-defined clustering process	Zone cost metrics are simple numerical metrics that describe how expensive electricity network reinforcement for a particular zone would be, based on the derived zone cost functions.	
Intra-cluster cost functions (electricity)	POM	Final intra-cluster cost functions (for electricity) to be sent to POM for the optimisation process	These represent options for allowable load/generation (kW) with associated costs (£) for different electricity network reinforcement options within the cluster – inter-zone meshing options also considered here at the MV level (but not at the LV level). See section 7.5.8. Costs are comprised by: - Annuitised capital expenditure (fixed) for intra-cluster electricity network reinforcement or expansion - Fixed operational and maintenance costs (fixed) for intra-cluster electricity network reinforcement or expansion - Variable operations and maintenance costs (dependent on utilisation and hence final values are calculated in POM) for intra-cluster electricity network reinforcement or expansion These cost functions are provided for each timeperiod vintage – i.e. as costs may change in future build years.	Distribution of network costs for Monte Carlo sampling as per section 7.3.4
Inter-cluster cost functions (electricity)	POM	Final inter-cluster cost functions (for electricity) to be sent to POM for the optimisation process	As above for electricity network meshing options	

Intra-cluster cost functions (heat)	POM	Final intra-cluster cost functions (for heat) to be sent to POM for the optimisation process	As above for heat networks [based on pipe size (mm)/temperature (°C)] & cost (£) and allow POM to make optimal decisions with regards to the sizing of the heat distribution network.	
Inter-cluster cost functions (heat)	POM	Final inter-cluster cost functions (for heat) to be sent to POM for the optimisation process	As above for heat networks [based on pipe size (mm)/temperature (°C)] & cost (£) and allow POM to make optimal decisions with regards to the sizing of the heat distribution network.	
Intra-cluster cost functions (gas)	POM	Final intra-cluster cost functions (for gas) to be sent to POM for the optimisation process	These represent pairs of maximum allowable load (kW) [based on number of customers not already connected to the gas network, and connection size] & cost (£) and allow POM to make an "all-or-nothing" decision on extending the gas network to cover off-gas grid buildings, to decommission it for operational costs savings, or to do nothing. Costs are comprised by: - Annuitised capital expenditure (fixed) for extending the gas network to cover off-gas grid buildings - Fixed operational and maintenance costs (fixed) for extending the gas network to cover off-gas grid buildings - Variable operations and maintenance costs (dependent on utilisation and hence final values are calculated in POM) for extending the gas network to cover off-gas grid buildings - Decommissioning costs for decommissioning the existing gas distribution network - Fixed operational and maintenance cost savings (fixed) for decommissioning the existing gas distribution network - Variable operations and maintenance cost savings (dependent on utilisation and hence final values are calculated in POM) for decommissioning the existing gas distribution network These cost functions are provided by time period.	
Inter-cluster cost functions (gas)	POM	Final inter-cluster cost functions (for gas) to be sent to POM for the optimisation process	These represent pairs of maximum allowable load (kW) [based on number of customers not connected to the gas network, and connection size] & cost (£) and allow POM to make optimal decisions with regards to inter-cluster gas network meshing options. Costs are comprised by: - Annuitised capital expenditure (fixed) for inter-cluster gas network meshing - Fixed operational and maintenance costs (fixed) for inter-cluster gas network meshing	

			- Variable operations and maintenance costs (dependent on utilisation and hence final values are calculated in POM) for inter-cluster gas network meshing These cost functions are provided by time period.	
Intra-cluster cost functions (hydrogen)	POM	Final intra-cluster cost functions (for hydrogen) to be sent to POM for the optimisation process	These represent pairs of maximum allowable load (kW) [based on number of customers in the network, and connection size] & cost (£) and allow POM to make an "all-or-nothing" decision on repurposing an existing gas distribution network to allow it to be able to transport hydrogen, or to do nothing. Costs are comprised by: - Annuitised capital expenditure (fixed) for repurposing an existing gas distribution network - Fixed operational and maintenance costs (fixed) for repurposing an existing gas distribution network - Variable operations and maintenance costs (dependent on utilisation and hence final values are calculated in POM) for repurposing a gas distribution network These cost functions are provided by time period.	
Inter-cluster cost functions (hydrogen)	POM	Final inter-cluster cost functions (for hydrogen) to be sent to POM for the optimisation process	These represent pairs of maximum allowable load (kW) [based on number of customers in the network, and connection size] & cost (£) and allow POM to make optimal decisions with regards to inter-cluster hydrogen network meshing options. Costs are comprised by: - Annuitised capital expenditure (fixed) for inter-cluster hydrogen network meshing - Fixed operational and maintenance costs (fixed) for inter-cluster hydrogen network meshing - Variable operations and maintenance costs (dependent on utilisation and hence final values are calculated in POM) for inter-cluster hydrogen network meshing These cost functions are provided by time period.	

7.3. Key functional requirements

The following sections describe the key functional requirements of the NAM in more detail. *The logic / process required steps, aligned to the individual sub-modules in the EPN tool, to undertake these are outlined in section 7.5.*

7.3.1. Capture most materiality issues within load flow modelling

The NAM can perform load flow modelling studies across the different energy vectors to ensure accurate allowable capacity limits are used when calculating reinforcement cost functions. This is because detailed load flow modelling cannot take place in POM as it needs to be done at a lower level of granularity. These capacity limits are determined by the local supply and demand conditions in the energy network under consideration, as well as by the identified key network constraints. For the majority of networks capacity limits are expected to be set by peak demand conditions, however it is also possible that network reinforcement or new build could be triggered by net exports when embedded supply outstrips demand (for example for a distribution network with very high penetration of small scale PV).

Table 7-2   Key constraints for electricity and heat networks

Network type	Component	Key constraints
Electricity	Network lines (LV, MV)	Voltage, Thermal
Electricity	Transformers (primary, secondary)	Thermal
Electricity	Meshed feeders	Voltage, Thermal
Heat	Network pipes (HT, MT, LT)	Energy delivered (e.g. due to maximum velocities), operating temperature, losses
Heat	Pumps	Energy delivered, operating temperature, losses

Electricity network modelling

Electricity network modelling focuses on electrical distribution networks up to (and including) 33kV/11kV substations as network reinforcement at higher voltage levels is likely to be impacted by additional factors that are outside the scope of the local area analysis for which the EPN tool is used. These factors include demand from large industrial customers, developments of large power stations etc. The baseline evolution of the 33kV network will need to be captured by exogenous scenario assumptions. However, as mentioned in 8.5.2 it will still be necessary within the POM to account for the ability to expand the network if driven by local area conditions.

The following constraints that can impact network reinforcement costs are considered, those which are not considered and the rationale for this are described further below:

▶ Voltage limits at the 33kV, 11kV and 230/400V level (as well as any other intermediate levels such as possibly 6.6kV for some UK distribution networks);

▶ Thermal limits of all three-phase and single-phase overhead lines and underground cables in the system;

- ▶ Thermal limits of all primary and secondary transformers (likely to be 33/11kV and 11/0.4kV) in the system.

These network constraints can be alleviated using a range of possible network reinforcement options. The following key network reinforcement options are considered by NAM:

- ▶ Replace, incremental upgrade or add a new distribution network line (underground cable or overhead line);

- ▶ Replace, incremental upgrade or add a new distribution transformer;

- ▶ Replacement, upgrade or addition or new meshed feeders[29].

- ▶ Distributed generation (including incurring costs to curtail those generators as a means of avoiding network reinforcement);

A range of other options, primarily 'smart' options which affect the level and shape of supply or demand on the network are captured as part of the POM pathway analysis and are assessed in parallel with the above options. They include:

- ▶ Demand side response;

- ▶ Energy storage management;

- ▶ Other smart grid technologies - substation automation and dynamic thermal rating systems[30].

Table 7-3   Network reinforcement options – electricity networks[31]

Option – Electricity networks	Voltage limits	Cable thermal limits	Transformer thermal limits
Replace, incremental upgrade or add a new network line	✓	✓	
Replace, incremental upgrade or add a new transformer			✓
Meshed networks	✓	✓	✓
Distributed generation	✓	✓	✓
Demand side response [in POM]	✓	✓	✓
Energy storage management [in POM]	✓	✓	✓
Other smart grid technologies [in POM]	✓	✓	✓

---

[29] Meshing options are only considered at the 33kV and 11kV level. A mesh can be created by closing one or more open points in the distribution network to resolve the network issue under consideration.

[30] In principle any option can be defined as long as it is possible to create the relevant cost function for the POM.

[31] Note: the load flow model will only consider generic load and generation and the range of load considered will cover existing to extreme conditions

The focus of the modelling described here is on steady-state conditions, with fault analysis not included as part of the detailed load flow simulations. NAM does, however, attempt to relate normal running criteria to n-1 contingency conditions. This is undertaken by reducing the headroom provided by the relevant options to reflect n-1 conditions to ensure the required build meets this standard.

Due to the focus on strategic local area design, only network constraints that can materially impact load flows (and hence costs) across each energy vector will be considered. The focus of the modelling described here is on steady-state conditions, with fault analysis not included as part of the detailed load flow simulations as mentioned above. Similarly, power quality, stability issues and harmonic analysis are also not considered here. The rationale for excluding these factors is outlined in Table 7-4.

Table 7-4   Materiality of electricity constraints that are not modelled in detail

Parameter	Definition	Key factors	Why is it excluded?
Steady-state voltage unbalance	In a three-phase system, voltage unbalance takes place when the magnitudes of phase or line voltages are different, when the phase angles differ from balanced conditions, or both.	– Phase distribution of power demand – Phase distribution of power generation – Power factors – Distribution line impedances	DNOs currently pay little attention to voltage unbalance (governed by Engineering Recommendation P29). For the purposes of the EnergyPath™ Networks Tool, it would be very difficult to know which phase existing customers are connected to, as well as to make meaningful assumptions with regards to which phase new customers may connect to.
Fault analysis	In a power system a fault is any abnormal electric current. Design of systems to detect and interrupt faults is the main objective of power system protection and can be an important design consideration.	– Fault currents – Type of fault – Location of fault	In general, fault analysis is more applicable for MV distribution networks (where traditionally the majority of faults have occurred) rather than LV distribution networks. Detailed fault analysis would be required in order to be able to predict whether fault levels in a distribution network have exceeded allowable limits. This would significantly add to the overall modelling complexity, and would also have cost implications in terms of the load flow software required.
Harmonic analysis	Power system harmonics are created by non-linear devices connected to the power system. High levels of power system harmonics can create voltage distortion and power quality problems.	– Non-linear power demand – Non-linear power generation – Distribution line impedances	Detailed harmonic modelling would be required in order to be able to predict whether harmonic distortion levels in a distribution network have exceeded their allowable levels (as governed by Engineering Recommendation G5/4-1). This would significantly add to the overall modelling complexity, and would also have cost

			implications in terms of the load flow software required.
			It should also be noted that if the harmonic producing loads are small in relation to total load, then harmonics will not present the most limiting network constraint.
Power system stability	The ability of a power system to regain a state of operating equilibrium after being subjected to a physical disturbance, with most system variables bounded so that practically the entire system remains intact.	– Magnitude of disturbance – Rotor angle stability – Frequency stability – Voltage stability	As with harmonic analysis, detailed stability modelling would be required in order to model the stability of a dynamic distribution network. Such analysis is considered to be outside the scope of the EnergyPath™ Networks Tool and would have significant implications in terms of complexity and cost.

District heat network modelling

District heat network modelling focuses on modelling the heat production facilities, pumps and pipes, storage facilities and connected buildings within the area under investigation. The focus is on thermal modelling rather than detailed hydraulic modelling, with turbulent fluctuations and leakage not considered in the modelling approach. Similarly, it is also assumed that fluid characteristics like density and heat capacity are constant. The rationale for excluding these factors is outlined in Table 7-5.

Table 7-5   **Materiality of heat network constraints that are not modelled in detail**

Parameter	Definition	Key factors	Why is it excluded?
Turbulent fluctuations	In fluid dynamics, turbulence refers to a flow regime characterised by chaotic property changes. The operating conditions of a district heat network are not perfectly stable and involve thermal and hydraulic transient regimes, in particular, temperature waves combined with temperature fluctuations	– Thermal transient conditions – Hydraulic transient conditions – Sudden temperature changes	The dynamic behaviour of district heat networks under varying operating conditions can be important in determining the safety and cost efficiency of the overall solution. Literature review suggests that simplified approaches provide a good approximation of overall operating conditions, particularly for representing the time delay in a system. Inaccuracies can occur, however, when trying to predict the temperature value at a specific time during the emergence of the temperature changes. However, this is considered to be outside the scope of the EnergyPath™ tool and would have significant implications in terms of complexity and cost.
Energy theft & leakage	Theft (for gas and electricity mainly) refers to energy losses that cannot be accounted for. Leakage refers to fugitive emissions from leaks in old pipes.	– Energy theft – Leakage	These typically form a very small part of overall energy delivered. Whilst they are very hard to model, some approximations may be used to take them into account.
Pressure drop caused by valves and fittings	A parameter affecting pressure drop in piping systems is pressure loss in the fittings and valves of the system. Typically, the calculated head loss caused by the valves and fittings within a pipe segment is expressed as an additional length of pipe that is	– Resistance coefficients and flow coefficients of valves and fittings	For piping systems within production facilities, the pressure drop through fittings and valves can be greater than that through the straight run of pipe itself. In long pipeline systems, however, which are the focus of the EnergyPath™ Networks Tool, the pressure drop through fittings and valves can typically be ignored with minimal loss in accuracy.

	added to the actual length of pipe when calculating pressure drop.		

The selection of pipes and pumps in a district heat network has important cost implications, but also important operational implications in terms of mass flow rates, velocity, operating temperature and heat losses, flow pressure drop and pump electrical energy consumption which need to be taken into account by the load flow model.

The aim of NAM is to determine, in terms of engineering feasibility, the minimum size of pipes and pumps that can be selected subject to a set of network constraints (e.g. typical maximum engineering velocities) and then pass the economic trade-offs to the POM, where the final size selection takes place. The following components are considered directly in the NAM as part of the network feasibility testing:

- ▶ Low Temperature (LT) pipes (temperature range 70°C to 85°C), Medium Temperature (MT) pipes (pressurised, temperature range 80°C to 115°C), High Temperature (HT) pipes (pressurised, temperature range 120°C to 200°C)

- ▶ District heating pumps.

The wider set of components, which make up the overall district heat network are considered in the POM (alongside the above network options produced by the NAM)

- ▶ Heat production plant (CHP or heat-only)[32] and overall temperature of the network

- ▶ Heat accumulators and storage facilities

- ▶ Wider topology of the network (radial, meshed, ringed) by connecting clusters within the local area together (within a cluster the network is assumed to be radial only)

The final network design is selected by the POM in accordance with the fundamental economic principles of heat network design shown the figure below, as some of the key factors that must be considered (e.g. the effective cost of the electricity to run the pumps) are only available as part of the pathway analysis undertaken by the POM.

It is important to note that temperature difference (ΔT) is assumed to be constant for all of the network infrastructure within a cluster as a whole, but the overarching temperature level can be selected by the POM given the heat supply options. As a result, this implies an interdependent set of network choices that needs to be taken into account when designing the overall process.

To a large degree, ΔT will depend on mass flow rates and on the velocity of the district heat network, which in turn will depend on the size of the pipes/pumps selected. Mass flow rates and velocity will in turn determine the flow pressure drop in the network, which is typically measured for the most remotely connected customer. The most common term used is "flow pressure drop per unit length" or "target pressure loss" (TPL) which refers to the combined

---

[32] For the purposes of the NAM modelling a free, unlimited heat source is considered for the purpose of testing network feasibility, whilst the POM optimises the choice of supply sources alongside building and network choices.

pressure loss of the supply and return piping. Typically pumps and pipes are also selected based on maximum TPL for the most remote consumer.

As Figure 7-3 shows, smaller pipes have lower capital costs but higher operating costs (including losses). Maximising ΔT at design conditions can reduce total investment – this is because reducing flow temperatures can significantly reduce heat losses in the figure. Similarly, however, it is important to ensure that savings in heat losses are not offset by increased pumping costs.

Figure 7-3   Economic design principle for district heat networks
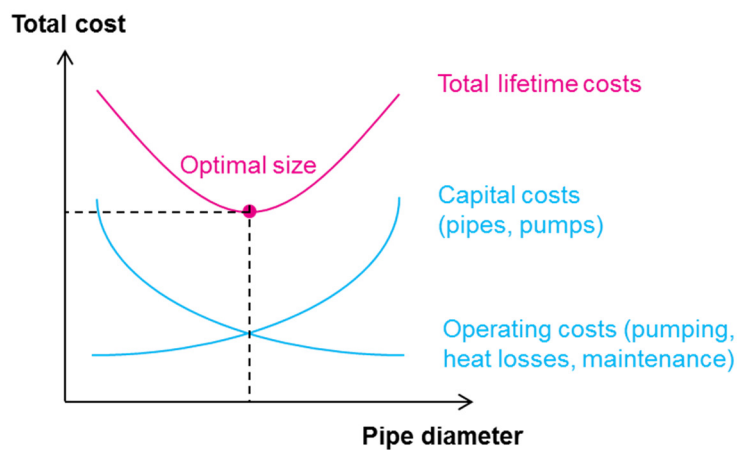
In terms of overall network topologies, two-line radial systems are predominantly used for small and medium-sized district heat networks. Meshed or ringed networks allow for incorporation of several heating stations and offer reliability benefits however due to longer and larger pipelines the investment cost is also typically greater. The choice between different network topologies (radial, ring or meshed) at the inter-cluster level (i.e. heat transmission) also takes place in POM based on the economic trade-offs described above, whilst the network within the cluster (i.e. distribution to the end-use building) is assumed to be radial.

In summary, the following key heat network options are modelled as outlined in the table below.

Table 7-6   Network reinforcement options – district heat networks

Option – Heat networks	Energy delivered	Operating temperature	Losses
Replace, incremental upgrade or add a new network pipe	✓	✓	✓
Replace, incremental upgrade or add a new pump	✓	✓	✓
Heat production plant and temperature of the overall network [in POM]	✓	✓	✓
Heat storage [in POM]	✓	✓	✓
District heat network topology (radial, ring, meshed) [in POM]	✓	✓	✓

Gas and hydrogen network modelling

NAM does not undertake detailed flow modelling for gas and hydrogen network to determine their suitability for a given cluster. Instead, the proposed approach is to represent these options, along with their associated costs and benefits, in POM by considering:

▶ Extension of the gas distribution network to cover off-gas grid buildings;

▶ Decommissioning some parts of the existing gas distribution network;

▶ Repurposing an existing gas distribution network to allow it to be able to transport hydrogen (potentially with expansion to cover off-grid buildings);

Dedicated new build hydrogen networks are not included in v2.1 of the EPN tool, with gas network repurposing considered as the only potential option, given the expected costs of new network, but these could be included in future versions.

Given the binary nature of decisions on decommissioning[33] an existing gas distribution network or repurposing it for transporting hydrogen, these are evaluated for a given cluster as an "all-or-nothing" decision and is not considered at the zone level. It is also possible to reflect other combinations, such as the impact on decommissioning costs if a heat network is being installed at the same time.

7.3.2. Ensuring that energy networks are accurately represented

The NAM should ensure that all existing and potential energy networks are represented to a sufficiently high accuracy in terms of their technical, physical and spatial characteristics (using the detailed topographical and siting information from the SAM), together with an accurate spatial and temporal representation of *the potential* demand and generation options in the area under investigation.

All network data are provided to NAM by SAM as explained in Section 6.2. In practice this involves the provision of the following key data types:

1. Distribution of archetypes across individual buildings in local area;

2. Synthesised physical layout of local electricity distribution networks (and data to synthesis physical layout of local heat networks), based on available data and GIS algorithms;

3. Existing and planned discrete sites for e.g. CHP deployment, new developments;

This information is used for two key purposes:

▶ Ensuring that all spatial information about the area under investigation is taken into account – this is important so that all parameters that can materially affect network component costs are included (Section 7.3.4);

---

[33] Decommissioning a gas network will result in capital costs but also in operating cost savings and these economic trade-offs will be simulated in POM.

▶ Ensuring that all physical information about the area under investigation is taken into account – this is important so that suitable boundaries (for example building blocks, rivers, cliffs, railways etc.) are used during the clustering process but also in order to capture load flows from production to consumption as accurately as possible;

It is important to note that the level of profile of embedded supply and demand (and associated peak demand or maximum export conditions) are ultimately determined in the POM, and are contrasted against the range of network reinforcement new / build options from the NAM.

Hence the NAM needs to model the full spectrum of load / supply conditions across the different parts of the network, such that a set of network options are available for any conditions that could be observed in the POM. The NAM can use data from the SAM/HOM to help inform the boundary of these conditions – e.g. it will know the number of each building archetype by zone, and the maximum heat load without any diversity effect to estimate the highest peak load that could reasonably be seen (e.g. if all buildings used electric resistive heating).

In practice, the network modelling requires engagement with the local relevant parties according to the specific energy vector modelled (e.g. DNOs for electricity networks, heat network operators for existing district heat networks etc.) to ensure that the existing networks modelled in NAM are calibrated to a sufficient level of real world detail[34].

As gas and hydrogen networks are not modelled in detail, the key information required by NAM is whether the area under investigation is currently being supplied by a gas network, and how close to one is it if not. Depending on potential energy requirements, the tool also considers the optimal size of pipes required to extend the gas network to cover off-gas grid buildings, or to repurpose existing gas networks to be able to transport hydrogen (albeit without the use of steady-state gas flow analysis in v2.1 of the tool).

7.3.3. Producing accurate cost functions for a range of network options

One of the key requirements of NAM is to be able to output cost options for different network reinforcement and new build options and for each energy vector (electricity, heat, gas, hydrogen). Cost functions must compare network reinforcement costs and maximum allowable capacity (load/generation) for system optimisation in POM e.g. when trading off reduced emissions through increased use of heat pumps versus the cost of reinforcing electricity networks to accommodate increased electrical load.

For existing network expansion the cost functions need to characterise the expansion costs plus any material operational costs and losses. For new build networks the full costs of installation are included, along with the operational costs and losses and the costs (where relevant) for later expansion or increased geographical penetration.

---

[34] If data is not available for a particular network component, generic values will be used based on expected network design.

The values for network options that will be calculated by NAM are[35]:

- ► Capital costs;

- ► Fixed operational and maintenance (FOM) costs;

- ► Variable operations and maintenance (VOM) costs;

- ► Operational lifetime and age of the asset (if existing).

- ► Construction/deployment period where this is >1 year.

FOM can be added to a capital expenditure and considered as one metric. On the other hand, however, VOM is dependent on utilisation (rather than peak capacity) and as such its final value will be determined by the POM depending on system dispatch optimisation. Finally the age and technical lifetime of reinforcement options is required in order to accurately represent the trade-offs between load-driven and age-driven reinforcement.

Three important issues need to be taken into account when producing cost functions:

1. Network costs must accurately represent the cost of covering each MW of local load (or local generation for situations where network reinforcement could be triggered by net exports to the network);

2. Where network reinforcement options have dependencies or are mutually exclusive this must be taken into account;

3. The number of network options tested by the NAM, as well as the number of network reinforcement options sent to POM, must be kept relatively low to ensure a tractable optimisation, but with flexibility for the user to easily scale up and down the number of options. As such, due to the complexity of the optimisation problem, only network constraints that can materially impact costs are considered.

7.3.4. Capturing material uncertainties in network component costs

The final cluster cost functions produced by the NAM (see section 7.5.8) are constructed from individual component costs (e.g. the cost of a wire or the labour cost of installation). These costs may be uncertain due to a number of factors, which have been identified, largely based on the ETI 2050 Energy Infrastructure Outlook project, and can be summarised as follows:

- ► Site context: the site context is important in determining land or access rights costs, transportation costs, as well as other costs such as costs related to street works,

---

[35] The component costs of the option are intended to reflect the economic resource costs of the options (see section 8.5.1) and are not on the same basis as e.g. the price control costs that are submitted to Ofgem, which include factors such as pension costs (which may vary by DNO). However, it would be possible to post-process the network component costs from the EnergyPath[TM] Networks Tool to account for these additional factors.

planning and consents costs etc; and from the above project is driven primarily by whether the area is classed as urban, suburban, rural or London[36].

▶ Material costs: these refer to the costs of purchasing the network component under consideration and can depend on commodity prices, global/national/regional supply and demand conditions, foreign exchange fluctuations, learning curves for future costs etc.

▶ Labour costs: Labour costs depend on regional labour costs, the skills availability in a particular region, as well as on the complexity of the installation under consideration.

▶ Plant costs: The availability of suitable plant to support the installation of the network would also have an impact on overall costs.

▶ Installation costs: The scale of installation could also have an impact on costs and this will depend on a complex inter-relationship between the sizing and capacity of the overall installation and the various system costs. In general, large installations could bring system costs down through economies of scale and through avoiding duplication of some costs and labour.

▶ Ground conditions – excavation difficulty: degree of difficulty expected to be encountered in the excavation of trenches and holes during construction.

▶ Ground conditions – ground contamination: excavated material which requires specialist handling and disposal as a result of chemical contamination contained within the soil.

▶ Ground conditions – ground water conditions: water requiring intermittent or continuous pumping during construction operations to keep excavated areas safe and dry.

These factors are summarised in Figure 7-4 below. For heat distribution networks we have also identified the height difference between the heat source and the supplied customers as an additional factor that can potentially impact total costs. Indicatively, for a hilly area where the difference between the heat source and the last customer is 100m, costs for the heat distribution network would be expected to increase by roughly 3% due to the increased need for suitable heat exchangers. This has a relatively low impact on heat network costs, but as an approximation the user can define a typical height distance between supply source and end-users on a cluster by cluster basis. In principle this could be made more detailed in future by, for example, incorporating Ordnance Survey contour layers into the representation of the local area.

---

[36] Rural (typically defined as locations with <30 dwellings per hectare); Suburban (typically defined as locations with 30-60 dwellings per hectare); Urban (typically defined as locations with >60 dwellings per hectare); London.

Figure 7-4   Key identified cost scalars
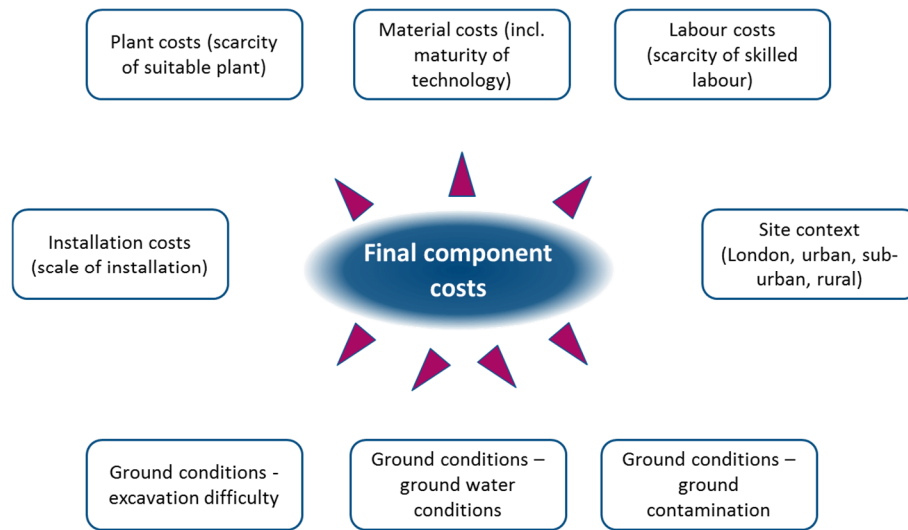
Table 7-7 summarises the approach for how the NAM takes the identified component cost modifiers into account:

▶ Material costs, labour costs, plant costs are all grouped together with the ability to define an uncertainty distribution around them – i.e. to create multiple sets of network options with different costs. These can be explored in the POM via Monte Carlo simulation of the network costs. *The approach to uncertainty across the EnergyPath^{TM} Networks Tool is discussed further in section 9.*

▶ Installation costs reflect possible economies of scale from large versus small-scale installation. This can be reflected in the decision making trade-offs in the POM via the use of MIP (Mixed Integer Programming) see section 8.5.2 for further details.

▶ Site context is modelled as a separate cost scalar, with information from the SAM (e.g. user defined data on rural, urban, semi-urban areas and component types e.g. under ground and over ground) used to flag the appropriate cost scalar for the NAM.

▶ Ground conditions (excavation difficulty, ground contamination and ground water conditions) can affect some network components more than others (for example underground cables more than overhead lines) and for those that they do, costs could vary very significantly. This is not captured explicitly in v2.1 of the tool and require the user to manually adjust the cost data on a cluster by cluster basis. For future versions this input would ideally be included as a GIS layer flag (or direct cost scalar) for different zones by the Local Authority. The information would then be used to adjust the network component costs.

Table 7-7   Treatment of identified cost scalars

Parameters	Allowable Values	Treatment	Data exchange
Site context	Rural Sub-urban	Capture range of uncertainty by using different cost scalars for London, urban areas, sub-urban areas and rural areas.	SAM to send information to NAM regarding the site context of the area

	Urban London		
Material costs	Can apply distribution around central value	Capture range of uncertainty by modelling a range of possible outcomes (representing different market variances) with regards to material costs .	Can use Monte Carlo process to simulate multiple sets of network costs for use in the POM
Labour costs	No skilled labour scarcity Central Skilled labour scarcity	Capture range of uncertainty by including a range of possible outcomes (representing different market variances) – these will be captured under the "material costs" category.	As per material costs, combine high across material, labour, plant costs
Plant costs	No scarcity of suitable plant Central Scarcity of suitable plant	As with "Labour costs".	As per material costs, combine high across material, labour, plant costs
Installation costs	Large scale of installation Central Small scale of installation	Scale of installation aims to capture economies of scale. This is can be captured in the POM via the use of MIP (see section 8.5.2)	Cost component for installation separated from other network cost functions and captured within the POM to reflect economies of scale
Excavation difficulty	Soft ground – no rock or hard material Intermittent rock / hard material Prolific rock / hard material	This will not be explicitly modelled at this stage as this information would not be available from a GIS package – the range of uncertainty will be captured under "material costs". It will be possible, however, to better reflect these costs as appropriate once this information is available.	Flagged by the Local Authority who would then adjust costs appropriately
Ground contamination	Clean and inert ground Mildly contaminated ground Heavily contaminated ground	As with "Excavation difficulty".	As with "excavation difficulty"
Ground water conditions	Little or no ground water Intermittent dewatering required Continuous dewatering required	As with "Excavation difficulty".	As with "excavation difficulty"

For illustrative purposes, Figure 7-5 below show the annualised cost of an 11kV underground cable (rated at 6 MVA) and a 300mm underground heat network (with an operating temperature of 120°C). The annualised cost is calculated as the initial cost plus repetitive refurbishment costs at replacement cycles, plus abandonment cost, and assuming a 40 year asset lifetime.

It can be seen that a significant range of uncertainty may exist for both components considered here, particularly if they are installed in urban areas. The challenge will be to include this uncertainty in the modelling approach in order to allow for different market variances to be

represented (see section 9). However, this should be narrowed down to the extent possible using the most accurate technical, spatial and physical information for the area under investigation.

Annualised cost of an 11kV underground cable (rated at 6 MVA) and a 300mm underground heat network (with an operating temperature of 120°C)
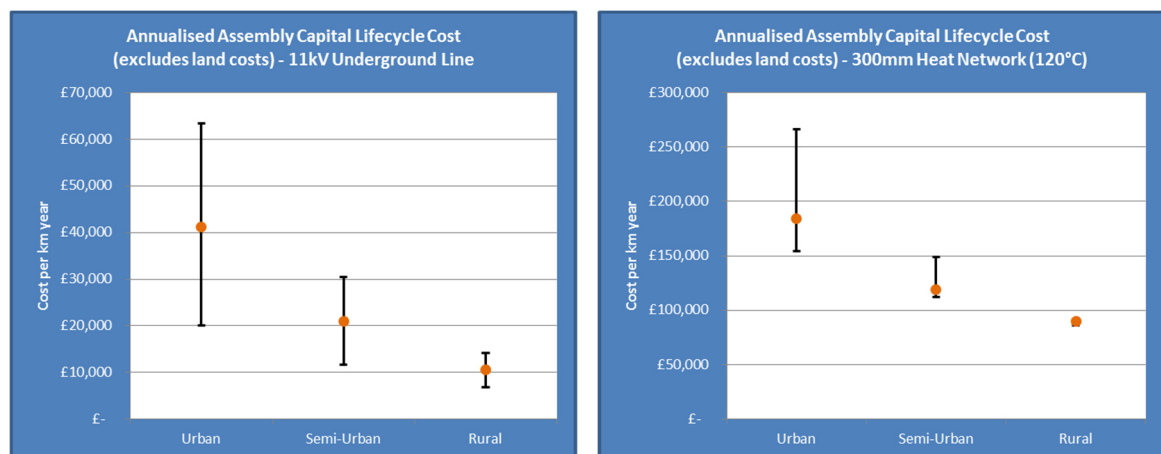
7.3.5. Producing 'automatic' cluster definitions with specific boundary constraints

Aggregating spatial zones into a smaller number of clusters is required in order to reduce the number of options available to the optimiser, thus ensuring that the optimisation problem is tractable given finite computing power. The final clusters used in POM should contain a reasonable representation of new and existing network options for both electricity and district heat networks within the same geographic boundary (and should also include a set of simpler options for gas grid extension, decommissioning or repurposing to hydrogen) so as not to bias the network and other trade-offs assessed in the POM.

Whilst the user can manually define clusters through the SAM, the NAM can provide a suggested set of cluster definitions that aims to balance the trade-off between the representation of heat and electricity network costs. This uses a combination factors to define the final boundaries of the different clusters as part of the automated process. The process is described in more detail in section 7.5.3, but the factors considered include:

▶ The maximum number of clusters

▶ The implied cost for developing a heat network within the cluster

▶   An equivalent cost for upgrading the electricity network within the cluster which can be contrasted with the heat network value in parallel to help ensure both costs are representative (e.g. such that cluster level cost functions do not inadvertently average very high and low costs of reinforcement in the same cluster, as this would indicate the need to produce two separate clusters

▶ Physical boundary constraints which respect existing electricity network topology

As described in sections 6.1.2 and 6.1.3 zones are the smallest building blocks for clusters, and are defined using electricity network characteristics: these could be the LV distribution feeder

supplying a street, the LV distribution network (which would include a number of LV distribution feeders) etc. Derived clusters must ensure that they do not breach boundary constraints set by the topology of the electricity network.

A high-level schematic of the approach is shown in Figure 7-6 below showing a variety of acceptable and unacceptable clusters based on the methodology described above.

Figure 7-6   **Cluster boundary constraints using building blocks**
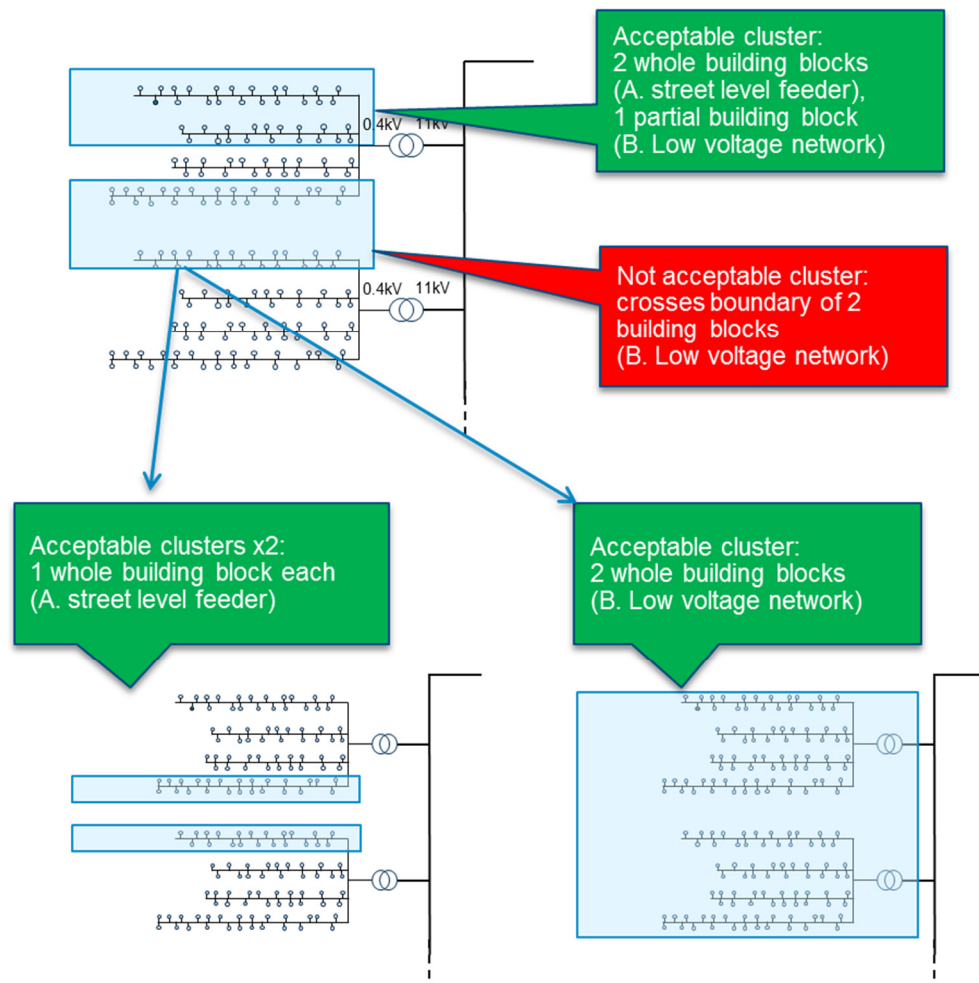
7.4. Key inputs

This section describes the key inputs to NAM, where they are used in the overall solution and for what purpose. The key inputs to NAM can broadly be summarised in four categories:

1. Primary inputs: These are primary inputs to the NAM and do not require interactions with other modules. The following inputs are included in this category:

   a. Generic network component database, i.e. a database containing all relevant information (costs and operational parameters) for the network model components modelled in NAM;

b. Any applicable cost scalars or flags, i.e. a list with any applicable scalars that will be used to inflate or deflate the baseline component costs depending on the factors identified in Section 7.3.4. Some of these e.g. urban/suburban/rural classification and network component type e.g. above ground and below ground are defined in the Sam at cluster level;

c. Statutory network constraint limits, i.e. a list with any limits to be applied on the identified network constraints after which network reinforcement or upgrade is considered to be required, such a drop below the minimum acceptable voltage level;

2. <u>Inputs from SAM:</u> These are inputs from SAM related to geographical information about the area under investigation. The following inputs are included in this category:

a. Overarching area information, i.e. spatial and physical information about the area to be used to determine appropriate cost functions and clusters;

b. Intra-zone information, i.e. geographical and technical information about existing and potential network components in the zone, as well as the number of each building archetype within the zone;

c. Inter-zone information, i.e. geographical and technical information about existing and potential inter-zone connections for all energy vectors. This could reflect the ability to mesh the two zones together if there is no physical reason they could not be connected;

d. Final cluster and inter- cluster definitions, i.e. a final definition of all cluster and inter-cluster connections (i.e. existing or feasible clusters which can be connected together) in order for NAM to derive cost functions for the final clusters.

3. <u>Inputs from HOM:</u> These are inputs from HOM related to demand profiles. The following inputs are included in this category:

a. Archetype demand profiles to help estimate existing and maximum plausible load per zone;

b. Archetype generation profiles to help estimate maximum likely embedded generation per zone.

The tables below describe what is technically required within the tool and how it is used. The relative importance of the primary data inputs and other factors such as costs and licensing restrictions, particularly where they must be purchased from an external provider, are discussed in the original deliverable (D4) Data Acquisition Plan. It should be noted that the ESC currently maintain the master set of input data for the current v2.1 EPN tool and this is not documented here.

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   147/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

Table 7-8   Input data – primary inputs

Data field(s)	Input type	Purpose	Granularity	Uncertainty parameters	Coverage
Generic network component database	Primary	A database containing all relevant information (costs and operational parameters) for the network model components modelled in NAM.	Electricity networks: Costs and operational parameters for distribution lines, transformers, generation units, energy storage units and smart technologies Heat networks: Costs and operational parameters for pipes, pumps, heat production plant and heat accumulators and storage facilities Gas and hydrogen networks: Costs and operational parameters for gas and hydrogen pipes. The generic network component database is based on the ETI 2050 Cost Database.	See below	Include all modelled network components, across all energy vectors and across the UK
Any applicable cost scalars	Primary	A list with any applicable scalars that will be used to inflate or deflate the baseline component costs depending on the factors identified in Section 6.3.4.	Cost scalars for the key factors ("component cost rate modifiers") that can materially affect network reinforcement costs. These include: - Site context - Material costs, labour costs, plant costs, installation costs - Ground conditions (excavation difficulty, ground contamination, ground water conditions) The cost scalars are also based on the ETI 2050 Cost Database.	Uncertainty distribution can be applied to core network component costs	Include all modelled network components, across all energy vectors and across the UK
Statutory network constraint limits	Primary	A list with any limits to be applied on the identified network constraints after which network reinforcement is considered to be required.	Statutory limits for the modelled network constraints, including: - Statutory voltage limits at the LV level (electricity) - Statutory voltage limits at the MV level (electricity) - Operating temperatures (heat)	-	Include all relevant network constraints, across all energy vectors and across the UK

			- Target pressure loss (heat)		
			- Operating pressure (gas, hydrogen)		
Technical characteristics of existing network components	Primary	SAM contains spatial information on where components are located, but NAM needs information on current headroom in current and future years to understand spectrum of requirements for network reinforcement options	Technical information about all existing network components across all energy vectors (cable/pipe sizes, transformer ratings, remaining lifetime etc.)	-	All existing network components

Table 7-9   Input data (from SAM)

Data field(s)	Input type	Purpose	Granularity	Uncertainty parameters	Coverage
Overarching area information	Secondary	Providing spatial and physical information about the area under investigation in order to be used to determine appropriate cost functions and clusters.	<u>Physical information:</u> providing physical information around suitable boundaries for clusters according to the four electricity building blocks identified (street level feeder, low voltage network, 11kV feeder, primary substation) Spatial granularity by zone	-	Include all modelled network components and buildings within the area under investigation
Intra-zone information	Secondary	Geographical and technical/physical information about existing and potential network components and buildings in the zone;	<u>Geographical information:</u> providing geographical information about the zone, which includes: - <u>Electricity</u> - Total street distance & feeder length for feeders connecting buildings to distribution substations (for all buildings); - <u>Heat</u> – Road network information for use in heat network synthesis - <u>Gas/Hydrogen</u> - Total street lengths that are on and off gas distribution system; - <u>All energy vectors</u> - Potential locations for new-build network components (for ex. potential location for a new transformer or a new heat source etc.). - <u>Building archetypes</u> – number of each archetype within the zone - <u>Embedded generation</u> – applicability of different types and maximum potential size of embedded generation <u>Technical/physical information:</u> providing technical/physical information about the zone, which includes:	-	Include all modelled network components and buildings within the area under investigation

Data field(s)	Input type	Purpose	Granularity	Uncertainty parameters	Coverage
			- Information around whether the buildings in the zone are currently connected to the gas distribution system.		
Inter-zone information	Secondary	Geographical and technical/physical information about existing and potential inter-zone connections for all energy vectors	<u>Geographical information:</u> providing geographical information in terms of existing inter-zone connections and potential new inter-zone connection options and the distance between two meshing points following road network (including connections to zones located in different clusters) <u>Technical/physical information:</u> providing technical/physical information with regards to the ratings of these connections and any other relevant technical characteristics	-	Include all modelled network components and buildings within the area under investigation

Final cluster and inter-cluster definitions		A final definition of all cluster and inter-cluster connections in order for NAM to derive cost functions for the final clusters.	The final cluster definition includes a list with the following data:
			- A list with all clusters (and the archetypes across buildings in each cluster) as identified following the user defined cluster definition.
			    ▪ Building ID;
			    ▪ Network component ID (e.g. feeder, substation, road node, road link);
			    ▪ Cluster ID
			    ▪ Technology package.
			- A list with all existing and potential new inter-cluster connections;
			    ▪ Cluster ID (on both ends of connection);
			    ▪ Type of connection.
				-	Include all clusters and inter-cluster connections within the area under investigation

Table 7-10   Input data (from HOM)

Data field(s)	Input type	Purpose	Granularity	Uncertainty parameters	Coverage
Local demand profiles	Secondary	Estimate likely maximum load conditions possible within zone to help bound load flow modelling	Building archetype half hourly- demand profiles for different characteristic days	Range of discrete demand profiles representing e.g. high/medium/low drivers of demand	All buildings within the area under investigation

| Local generation profiles | Secondary | Estimate likely maximum embedded supply conditions possible within zone to help bound load flow modelling | Half hourly-supply profiles for different characteristic days | Range of discrete profiles representing e.g. high/medium/low output | All generation units within the area under investigation |

7.5. Key NAM logic / process Steps

The next sections describe the key steps within NAM that fulfill the modelling requirements described in Section 7.3, for each specific NAM sub-module [NAM-00X].

7.5.1. Network Analysis Module [NAM-000]

Relevant spatial and physical information is first received from the SAM (see section 7.4) with regards to the area under investigation, at a zonal level of detail, along with information from the HOM on the demand and generation profiles associated with archetypes in the area. Zones are the smallest building blocks for clusters, and are defined using electricity network characteristics.

We have identified 4 levels of the zone building blocks, increasing in size, which must be respected as part of any final cluster definition (see section 7.3.5, Figure 7-6):

A. Street level feeder that connect buildings on roads to the nearest distribution substation;

B. Low voltage network (Distribution substations that have several LV feeders)

C. 11kV feeder (HV feeders that connect several distribution substations to the nearest primary substation);

D. Primary substation that has several HV feeders.

These blocks are shown schematically in Figure 7-7 below.

Figure 7-7   **Identified zone building blocks levels reflecting electricity network connectivity**
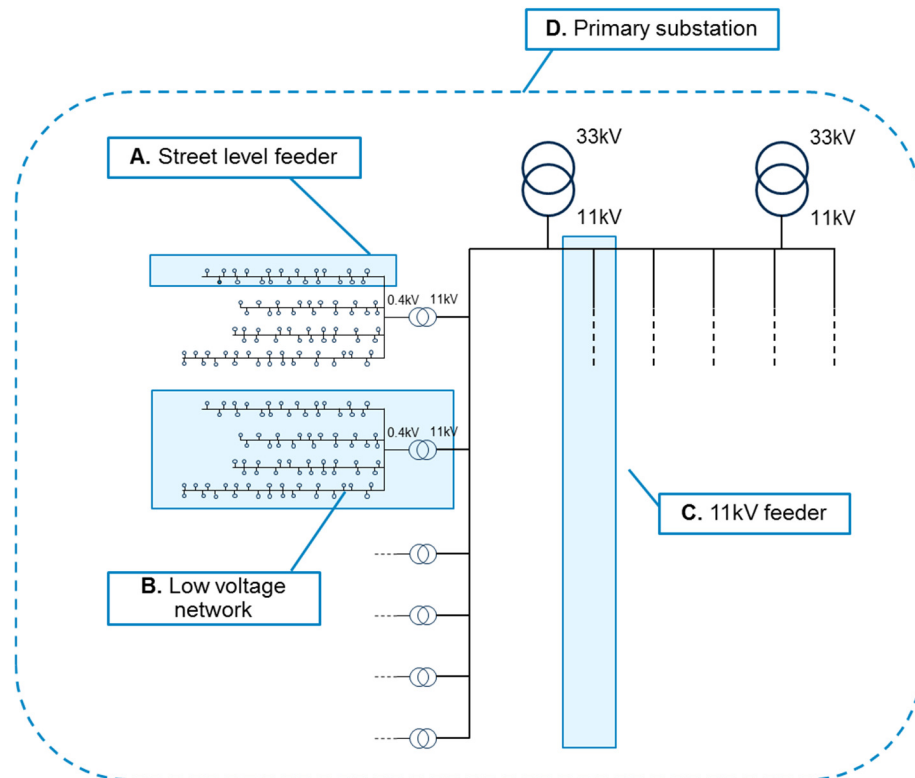
The aim is to maintain the base zone definition for the load flow modelling at the lowest level of spatial granularity (i.e. at street level) within the NAM in order to maximise accuracy of the derived cost functions. However, given the number of load flow studies that need to be undertaken to test the full range of network reinforcement and new build options, it is likely to be necessary to simplify the network representation slightly, as described below.

Electricity network modelling

For electricity this is undertaken first by collapsing the representation of individual buildings (units) on a street-level LV feeder (rated at 0.4kV) to an aggregated load at each road node, as shown in Figure 7-8.

Figure 7-8   Simplified street level feeders
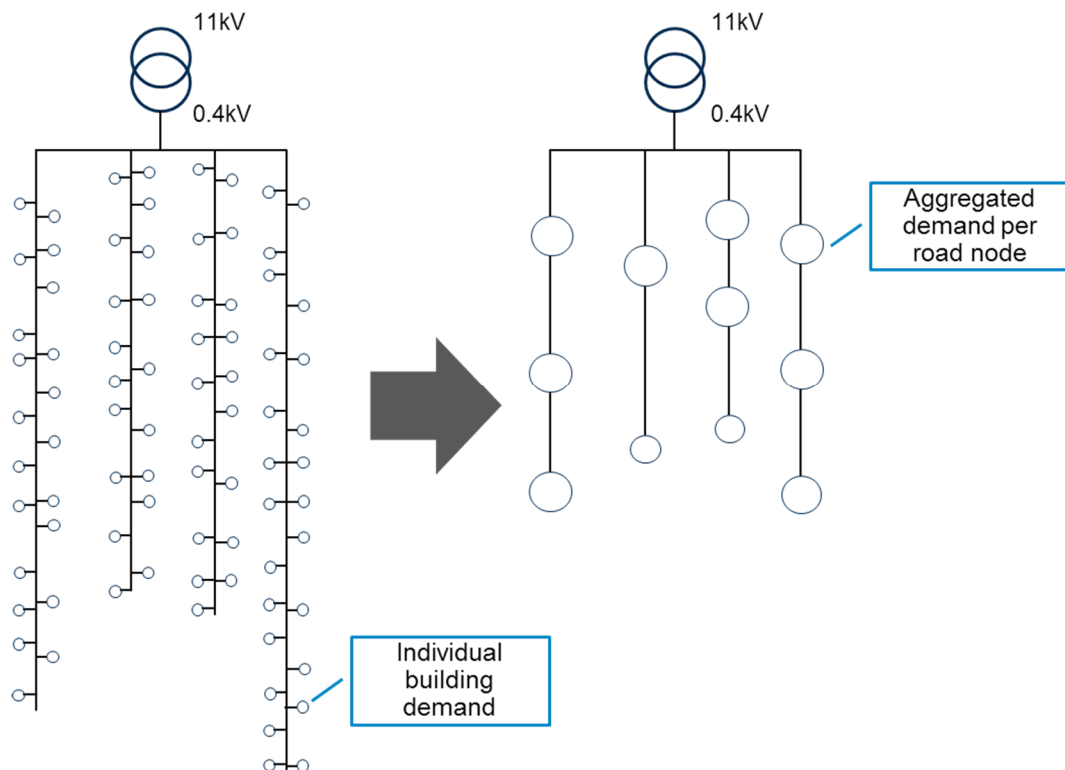
In this simplified representation load flow calculations can be performed on each feeder independently using SINCAL to assess peak load / cost characteristics for incremental reinforcement options for each feeder cable.

As part of the pre-processing in this sub-module, the maximum possible load/supply conditions per zone are estimated (e.g. by using the maximum non-diversity adjusted building archetype demand profiles on the most extreme cold-weather characteristic day) to help bound the number of load flow studies to be considered.

7.5.2. Electricity feeder (LV) cost metrics [NAM-001][37]

To inform the user-defined clustering process in the SAM the NAM can create a series of electricity cost metrics to identify where upgrade costs at a zonal level appear similar. These metrics are used within a GIS spatial statistical grouping technique in the SAM - outlined in section 6.5.7 - to visualise these. To enable this, the electricity cost functions created for each zone are converted into a single cost metric, as the GIS grouping functionality is limited one metric, to describe how expensive the costs of reinforcement are.

Possible zone network reinforcement cost metrics could include (see Figure 7-9):

---

[37] This is now deprecated (along with a part of SAM-006). It used to provide visual help for the user to define clusters manually in ArcGIS. It is now replaced by an automated process where the user can select clusters originating from electricity assets (e.g. HV substation and downstream).

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   154/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

▶ The area under the derived cost function, potentially normalised by the MW values (a);

▶ The electricity network reinforcement cost at potential peak load (b);

▶ The derived cost function gradient (c);

In addition, the definition of a cluster cannot change between time periods within the POM hence if the zone cost functions and subsequent metrics evolve significantly over the pathway to 2050, the metrics need to account for this. For example, averaging across all time periods under consideration.

These metrics (and potentially others) have been explored with real world sample data as part of early testing of v1.0 of the EPN tool and the overarching aim is to embed the most representative 'shape' of the cost of reinforcement. A) and c) are generally more appropriate than b), and a) is generally less sensitive to the value used to fix the biggest reinforcement option than c).

Figure 7-9   **Possible zone network reinforcement cost metrics**
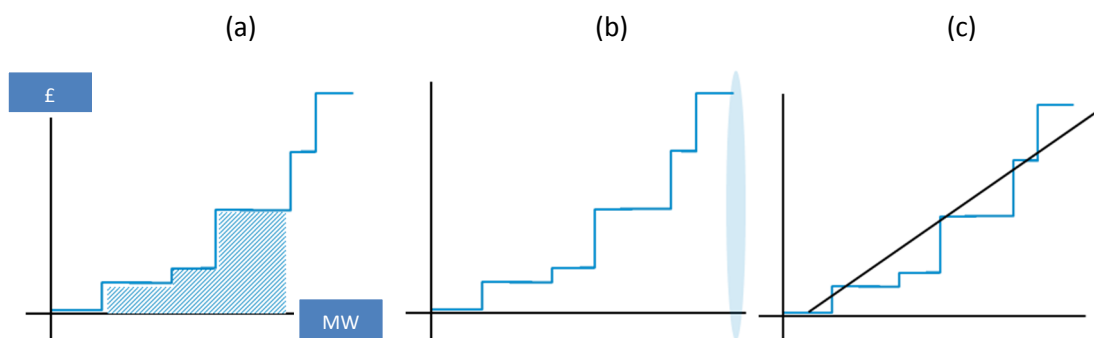
For v2.1 option a) has been implemented, averaging the calculated metrics across all time periods to create a single metric.

7.5.3. Auto-cluster definition (optional) [NAM-002]

This section explains the auto-clustering process that takes place in the NAM to provide a suggested final cluster definition that balances the trade-offs between the representations of heat and electricity network costs and associated topology/connectivity issues. The broad methodology can be summarised in the follow four steps:

1. Choose geographical "**shapes**" to test

2. Test metrics in each shape

3. Combine shapes into clusters based on metrics and constraints

In the sections below we cover each step in more detail.

1) Choose geographical "shapes" to test

We will start with the smallest possible unit size – an LV feeder. However, allowing clusters of this size may result in far too many clusters in the modelled area making the subsequent POM

optimization intractable. The user can choose to limit the smallest shape size to being LV substations and downstream elements, or indeed HV feeders or even HV substations. The methodology is robust to this constraint on the smallest shape size.

Many of the "good" cluster shapes (when we are only allowed a small number of total clusters) will likely be HV feeder groups where one of these contains a large anchor heat load. As mentioned below, it is possible to identify adjacent or overlapping network objects as well via the roads of the buildings being served by the network elements being clustered.

2) *Test metrics in each shape*

The algorithm tests each shape to provide metrics that will be used to assess whether shapes can be sensibly clustered together. Network investment decisions are made at a cluster granularity in the POM (all or nothing decision for each cluster) and so it considers metrics of network reinforcement costs when trying to cluster "similar" shapes. It does not include any metrics for building archetypes, as these can vary within cluster when being considered in the POM. I.e. only the cost of the direct network infrastructure (pipes, feeders, substations, heat exchangers) is considered and not building related costs, heat/electricity sources of supply

The cost metrics for electricity and heat networks are tested separately.

1. **Electricity** – re-uses part of the methodology for the electricity metrics in NAM-001, however, the key difference for auto-clustering is that a single cost curve combining all components relevant to the test cluster is created before being normalised, rather than just LV feeders. For example, for an HV feeder test cluster this is HV feeder + LV substations + LV feeders. This methodology gives a metric that describes the average £/MW reinforcement cost for possible future increases in demand, and includes any "free" head room. This metric will vary depending on the available free head room on components, and the density of demand (MW demand per m cable length), and can be combined between components and network levels. By normalising the curve to give an average £/MW reinforcement cost this avoids any scaling issues as the potential cluster size increases. A cap is placed on the expected maximum peak demand level (MW) in the cost curve used to calculate the metric. If this is instead bounded by the maximum component capacity, the metrics for a similar shape with very low demand and a shape with very high demand may look very different. This methodology relies on undertaking the load flow calculations for all parts of the electricity network.

2. **Heat** – to avoid using the heat network synthesis process and SINCAL modelling (see sections 7.5.5 and 7.5.6 respectively) for each test shape which is highly time consuming, a far simpler method is used to cost the heat network:
   a. Extract populated road length in shape
   b. Cost = StandardPipeCost * RoadLength + StandardHeatExchangerCost
   c. Extract peak demand in shape
   d. Metric = Cost / PeakDemand

   This method gives high £/MW costs in very small areas (1 heat exchanger and low MW demand), lower costs in medium areas (1 heat exchanger + small pipe costs and medium

MW demand) but high costs in large less dense areas (1 heat exchanger + very high pipe costs and medium/high MW demand).

Using only 1 standard pipe size assumes this meets demand in all shapes tested. On average this may be true, but this may overestimate the costs of small networks (where smaller pipes may be used) and larger networks (where larger pipes must be used) or where pipe sizes are tapered moving from source to demand. This metric can be refined in future however, the overarching principle to quickly test combinations of shapes that may look "good" for heat networks without having to run the detailed synthesis and load flow process. *This detailed process [NAM-004/NAM-005] always needs to happen after the clusters have been defined regardless of whether these are manually or automatically defined.*

3) Combine shapes into clusters based on metrics and constraints for testing

It is computationally intractable to test all possible cluster definitions, based on different combinations of "shapes" within these. Instead a "growing" algorithm is performed, that iteratively adds in neighbouring shapes to create a "test cluster" and re-evaluates the electricity and heat metrics in this "new test cluster" against both an absolute condition and relative to the "previous test cluster".

The "depth" of the search can be varied to allow 1, 2, 3 or more additional shapes to be added and tested at each point, to find appropriate cluster boundaries. For an area with N shapes at all levels, and a search depth of D, this will result in at most $N*(1+D)$ calculations of the metrics (test each shape alone, and test the addition of an additional D shapes to each)

This algorithm relies on having a method of deciding the starting shape, and deciding the next shape to try adding on at each point. The heat metric is used to choose the starting shapes as low values are viewed as a good proxy for an "anchor heat load" around which a heat network could be constructed.

At each step, both heat and electricity metrics are checked to see if the "new test cluster" is acceptable subject to stopping criteria which are described further below.

The overall process is illustrated graphically in Figure 7-10, note that here the test "shape" has been defined to be the smallest zone definition in the EPN tool (i.e. an LV feeder).

Figure 7-10   Illustration of auto-clustering process and metric calculation
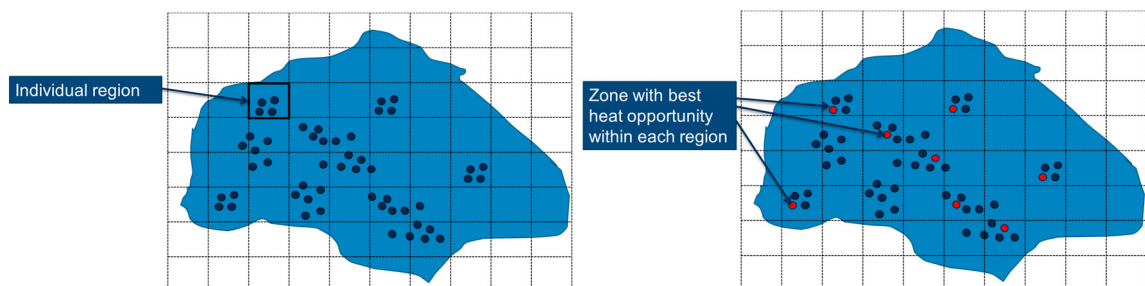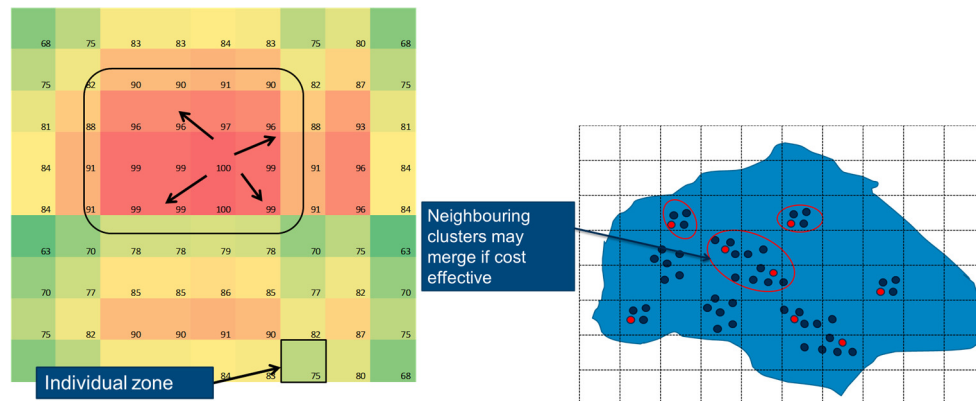
When testing the "new test cluster" the following conditions are evaluated:

▶ **Heat** - Does the combined shape have a heat metric below a certain user-defined threshold[38]? In this case the cluster is acceptable. If the starting shape that is being joined to has a heat metric below the threshold we can cluster regardless of the new combined shape heat metric. I.e. it is acceptable to cluster where the metric pairs of original shape and combined shape follow Good-Good, Bad-Bad, and Bad-Good. It is not possible to cluster Good-Bad.

▶ **Electricity** - Does the electricity metric of the combined shape vary from the metric of the starting shape (i.e. as a ratio) by less than a user-defined threshold (+/- in both directions)?

If both are true then add a new shape is added to the test cluster and the process continues until either are false or the depth testing limit is reached, the test cluster which still passes the conditions is then fixed as a cluster. A new search begins again with a different starting shape (selected through the lowest heat cost metric), but now excludes test shapes that have already been assigned.

As part of growing the different test clusters a number of constraints must be observed with respect to the shapes that can be combined together for testing. These include:

1. Electricity network topology must be observed (the same rules as currently used in the SAM semi-automated cluster definitions – i.e. objects can only be clustered if they share the same upstream electricity network object, and upstream objects can only be clustered with other upstream objects if all downstream objects are also clustered together in the same cluster)

2. Locational proximity – clusters must be physically "close" to one another to be clustered.
   a) Shapes are considered "close" if they share a boundary via linking roads
   b) If no shapes share a boundary a shape is chosen with the closest heat load weighted centre (as the crow flies)

---

[38] This based on a top-down engineering estimate of the typical heat density and associated cost above which are heat network is 'unlikely' to be economic.

3. Physical boundaries – rivers etc. These are not considered automatically at present in v2.1 of EPN, but the user has the ability to sense-check the automatically generated clusters

The final cluster definitions from the auto-clustering process are passed back to the SAM. The user can visualize these and decide whether to accept them or modify them further e.g. due to the physical boundaries mentioned above. Regardless of where the clusters have been manually or automatically defined the downstream process in the rest of the EPN tool in terms of the remaining SAM/NAM/POM steps remains the same.

It is important to note that the aim of the auto-cluster definition is not to define a sufficiently large enough cluster that represents an *entire network*, but a *potential component cluster of this*. By seeing all trade-offs between all clusters simultaneously, the POM can then make the decision to develop heat networks across multiple clusters, connect these clusters together, as well as to an appropriate heat source; all as part of creating an overall economically viable heat network (e.g. see section 7.5.8).

In practice, predicting the optimal set of network choices before modelling these decisions in POM will be challenging and to some extent depending on the topology of each Local Authority area. Expert user judgment is still generally needed to define the initial clusters, examine the POM results and refine the clusters definition as part of the overall process of using the EnergyPath™ Networks Tool. The above process provides another route to enable the user to undertake this as efficiently as possible.

7.5.4. Electricity network options and load flow modelling [NAM-003]

For a given spatial area (e.g. zones at the different building block levels for electricity) this step generates a series of network reinforcement/new build options which must be tested within the load flow modelling tool. The components are drawn from a master component database list which is valid for each relevant section of the network (e.g. cable upgrades for feeders or transformer upgrades for substations).

Via an automated process, all valid combinations of reinforcement options are added to the network configuration and a series of iterative load tests[39] are undertaken on that configuration until network constraints on the relevant part of the network are breached (e.g. thermal limits are breached or voltage drops below a given threshold).

Where these limits are breached reflects the maximum allowable load or embedded supply export, enabled by these network options. The options are then mapped to the underlying costs and characteristics within the component database to understand the associated:

▶ Capital and installation costs (these may be adjusted to reflect the locational and other uncertainties outlined in section 7.3.4 – as result the final options may have an uncertainty distribution applied to them which can be explored as part of the overall Monte Carlo analysis in the POM – this is discussed further in section 9)

▶ Fixed operational and maintenance (FOM)

---

[39] From a maximum net export position, where relevant, through to maximum load.

▶ Variable operations and maintenance (VOM)

▶ Operational lifetime of the asset.

Testing all possible combinations of network components in all feasible configurations to their failure point is clearly a computationally intensive process, however, this process is simplified in a number of ways:

▶ As outlined in section 7.3.1 dynamic / active network options (such as storage) are handled in the POM, hence the number of options to be examined in the NAM is reduced.

▶ More careful selection of the number of load tests via use of an interval/binary search algorithm.

▶ By separating out testing of network options, which are effectively independent of others (e.g. radial lines and transformers) from others which are interdependent (e.g. meshing) and hence must be tested in conjunction with other network options.

The process of creating the final cost functions/options for use in the POM from this basic process is complex and is discussed further in section 7.5.4.2.

For the vast majority of distribution networks it is anticipated that network reinforcement will be triggered during periods of 'maximum demand – minimum embedded generation'. This is because the most extreme system conditions in terms of network flows are likely to be experienced during Winter peak demand, although for some areas (particularly for areas with high penetration of air conditioning units in the South) this could also be during Summer peak demand.

Conversely, for zones with significant potential for distributed generation it is possible that network reinforcement (particularly to prevent high voltages) could be triggered during periods of 'minimum demand – maximum embedded generation' if the total installed capacity of these units is greater than peak demand.

For example this could be the case during a particularly sunny day in an area with significant penetration of small-scale PVs. These zones should be identified and the network reinforcement options produced for them should cover both 'maximum demand – minimum embedded generation' as well as 'minimum demand – maximum embedded generation' system conditions as shown.

For non-building scale embedded generation it will be important to use information from the SAM on the potential location and maximum installable capacity (e.g. of small scale wind) to test network reinforcement configurations, which are likely to be required in significant net supply cases. Where it is clear that an electricity network reinforcement option is only being driven be larger-scale embedded supply, this could ultimately be linked to peak supply from certain technologies within a cluster in the POM, as opposed to the peak demand net build-scale generation, which would be used to drive the majority of network option decisions.

Aggregating electricity load moving from lower to higher level zone building blocks

When performing the load flow studies, it is necessary to distribute the incremental load (for example due to heating) amongst the individual zones connected at a lower level building block to the next level up (e.g. a street level feeder connected to a low voltage distribution substation). For electricity networks this is required in order to accurately test voltage and thermal constraints on the connected lines.

This is necessary because when the final cluster boundaries are determined (which reflect the different zone building block levels within them) the POM has no understanding of where load is developing spatially within the cluster, as this information has been lost to simplify the problem. It only understands the aggregate peak load (net of embedded generation) due to choices it is making within the cluster and must trade this off against the cost of network reinforcement to support this peak (see Figure 7-11).

Hence the way the underlying network cost functions are constructed for a cluster (based on the zones within it) must superimpose a similar approximation to ensure that the way the POM ultimately interprets the cost functions is consistent.

Figure 7-11   Example of aggregating load from lower to higher level zone building blocks
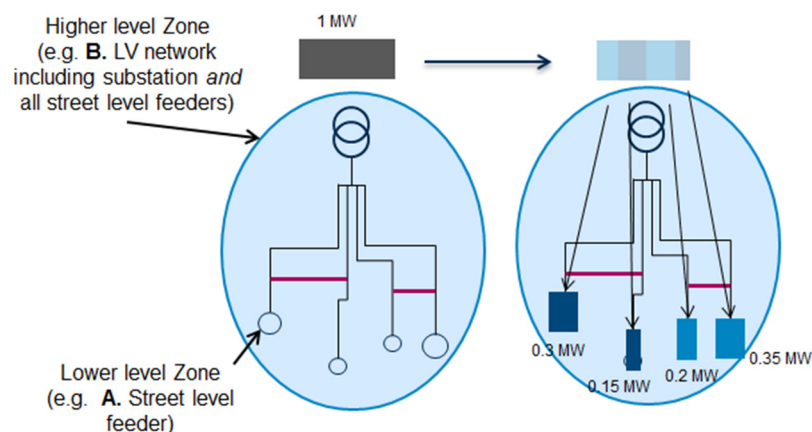
There are a number of ways to approximate this load distribution ex-ante, for example:

▶ Lower level Zones with lowest cost first – this would be comparable to a best-case scenario where incremental load would accrue at the parts of the networks with the lowest reinforcement costs (for example due to DNO incentives);

▶ Lower level Zones with highest cost first – this would be comparable to a worst-case scenario where incremental load would accrue at the parts of the networks with the highest reinforcement costs;

▶ Equally distributed amongst lower level Zones – this would assume that the same level of incremental load would be added to each zone within the cluster;

▶ Based on zone's potential peak load share of cluster – this would assume that the level of incremental load added to each zone would be based on its potential peak load share (which in turn is likely to depend on the number of customers connected in that zone);

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   161/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

▶ Distributed based on a lower level Zone's current load share – this would assume that the level of incremental load added to each zone would be based on its current load share (which in turn is likely to depend on the number of customers connected in that zone);

The same basic principles for aggregating load apply as you move further up the zone building block levels – i.e. from **B.** Low Voltage Network to **C.** 11kV feeders to **D.** Primary substations

In v2.1 of the EPN tool the last option (based on current load share) is implemented as it is assumed to be the least distorting default in terms of the final POM solution. Ideally a metric which reflected the likely propensity to electrify by zone would be used (i.e. trying to pre-judge the outcome of the pathway analysis in the POM). However, creating this is complicated by the myriad of factors which are ultimately traded-off to determine the degree of electrification in the local area.

7.5.4.1. Load flow modelling

The steady-state load flow modelling for electricity helps to understand how network constraints (such as thermal limits or voltage as outlined in section 7.3.1) are affected under different configurations of network reinforcement / new build options under different load conditions. For the majority of networks capacity limits are expected to be set by peak demand conditions, however it is also possible that network reinforcement could be triggered by net exports when supply outstrips demand.

▶ For simple radial networks (i.e. for networks where meshing options are not considered) load flow calculations can take place in the network reinforcement 'wrapper' rather than in a dedicated network load flow modelling software. This allows the overall process to be considerably faster.

▶ For systems which exhibit meshing (i.e. at intra-cluster and inter-cluster level) dedicated network load flow modelling will be used to ensure that load flows are modelled as accurately as possible.

7.5.4.2. Zone level electricity network option cost functions

Once the electricity network reinforcement / new build options have been tested via the load flow modelling they need to be converted into cost functions for the relevant zones. These need to consider the relevant components at each *level* of the zone building blocks A to D in section 7.5.1 (this is important because all final cluster boundaries definitions have to respect this underlying topology/connectivity definition).

These zonal functions are aggregated later to a cluster level function and passed to the POM for use in the pathway analysis (as described in section 7.5.8). *For heat this process is different* as there are intermediate steps to first define sensible cluster boundaries and then evaluate the heat network cost options within the cluster as a whole (see section 7.5.5) rather than individually for a zone and then aggregate up.

Five possible methods were originally identified for collating the electricity network options at a given peak load of increasing accuracy and complexity. These options are:

1. Incremental cost function (lowest cost reinforcement option to cover each MW of capacity);

2. Discrete additive options (both for £ cost and for MW capacity headroom);

3. Discrete additive options with careful selection of cost order of packages – i.e. where we need to represent dependency (X needs to be done before Y) - X must be cheaper than Y to operate properly within the POM's least cost optimisation;

4. Separate 'sets' of cost options e.g. cable options vs transformer options being driven off same peak MW – this would allow us to resolve the £/MW headroom interdependency issue for everything but meshing;

    b. Potentially this could be overcome by separately testing two configurations of the network: (i) with meshing; and (ii) without meshing.

5. Separate discrete options with interdependencies (e.g. X -> Y -> Z) as best way to capture interaction on £ and MW headroom across different packages of options particularly meshing.

*At present v2.1 of the EPN model implements option 4b. It is possible for the user to define mesh links **within** individual clusters (as opposed to between clusters) in the SAM and the NAM will automatically include the impact of this meshed link when generating the final cluster cost curves. By running the NAM (and subsequently POM) with and without the mesh links it is possible to understand their impact.*

Option 5 has not been implemented to-date, but would instead capture the choice over whether to mesh endogenously within the POM (rather than via exogenous scenarios in option 4b). This is technically feasible, but would require the use of integer conditional constraints, which are likely to lead to a sizeable performance impact in the POM.

Method 1: Incremental cost function

The first method is to consider a simple cost function where all the available network options are considered to be incremental. The only strength of this option is reduced complexity, however its fundamental weakness is that it does not account for dependencies between different network reinforcement options due to its incremental representation.

For example, having selected reinforcement option #2 in an early timeperiod (i.e. to reinforce a network line and also create a mesh) to meet capacity D1, we would not dismantle the mesh to select reinforcement option 3 (line and transformer) to meet capacity D2 in a future period. In this regard reinforcement options 2 and 3 should be mutually exclusive and importantly the full cost of option 3 should be incurred to implement it, however this cost function representation would misleadingly only consider the incremental cost of moving from option 2 to 3.

Figure 7-12   Incremental cost function (lowest cost reinforcement option to cover each MW of capacity)
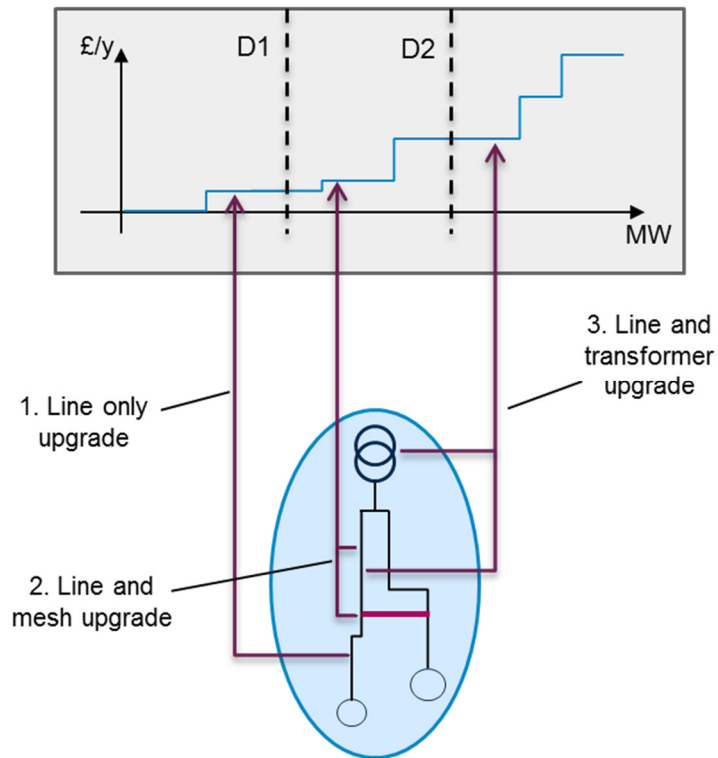
Method 2: Discrete additive options

The second method is to only reflect a set of discrete network reinforcement options in the function, which are genuinely additive in terms of their headroom – i.e. no options can be included which are mutually exclusive. The benefit of this is that the incremental costs are more reflective of the nature of the upgrade costs.

However, the key weaknesses are that this limits the available network options which can be represented and that an incremental upgrade may be lower cost compared to a pre-requisite enabling upgrade. For example in the figure below, on an incremental basis option #2 (i.e. to reinforce a network line and also create a mesh) is cheaper than option #1 (i.e. to reinforce a network line only) and as a result the optimiser will always choose option #2 before option #1, however, in reality option #1 must be undertaken *before* option #2 is available.

Figure 7-13   Discrete additive option (both for £ cost and for MW capacity headroom)
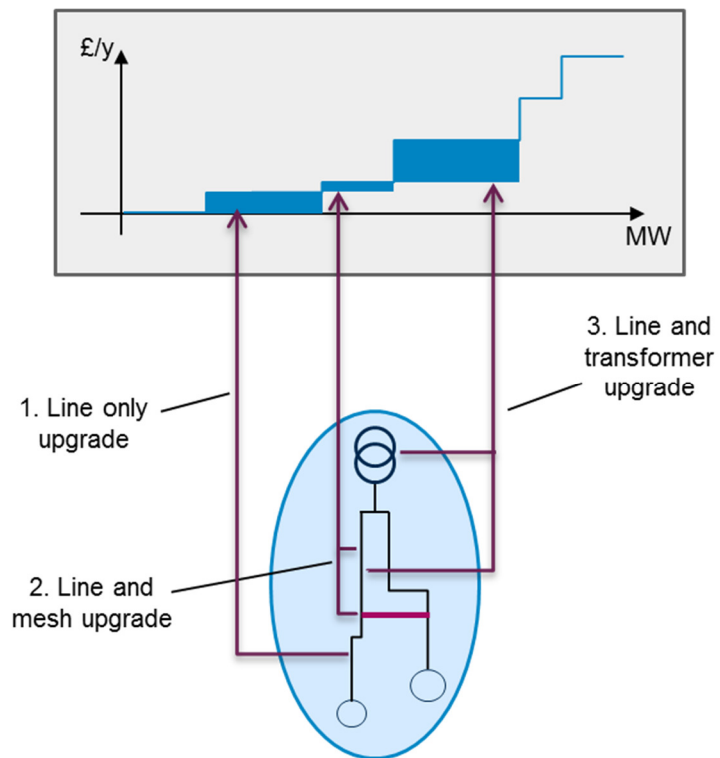
Method 3: Discrete additive options with selection of cost order of packages

This is a refinement of Method 2 by removing or adjusting previous options to ensure that network reinforcement options are monotonically increasing in terms of costs and the appropriate order of network upgrades. As shown in Figure 7-14, there are two main ways to achieve this:

▶   On the left hand side - aggregate both the enabling option #1 (Line only) and the incremental upgrade option #2 (Line and mesh upgrade) into a single option (thereby losing the differentiation between them)

▶   On the right hand side – remove option #2 the incremental mesh upgrade

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   165/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

Figure 7-14   Discrete additive option with selection of cost order of packages
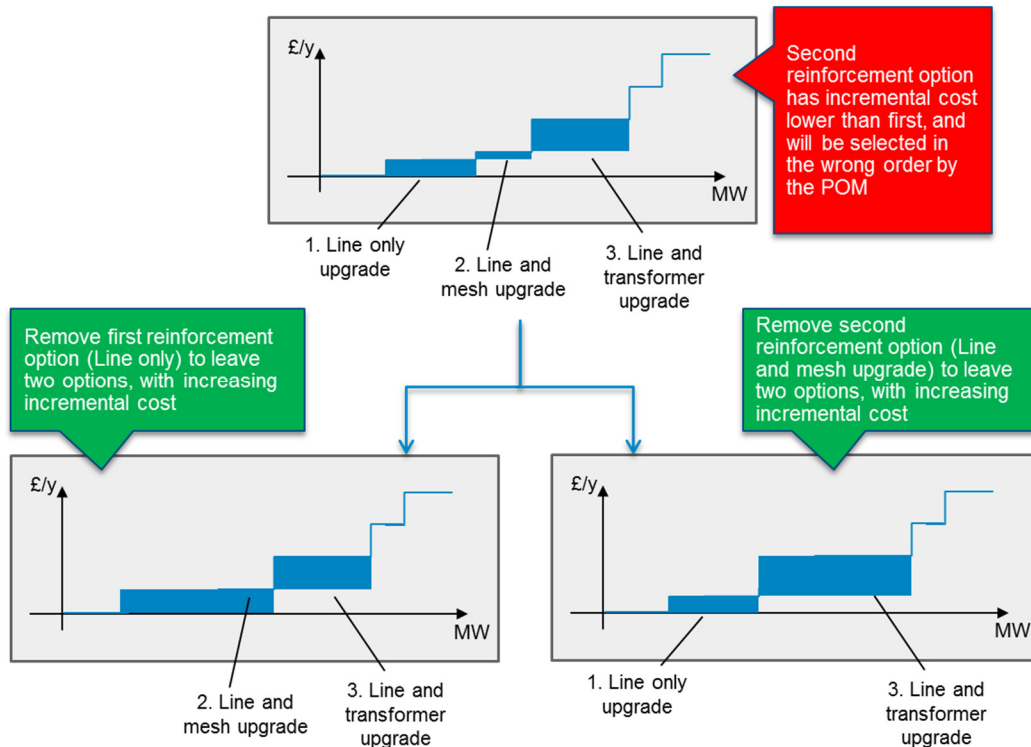
There are still two key issues with this method:

► The first is that this process could remove a large number of potential network reinforcement options (especially for relatively densely populated areas where radial networks are used but where meshing could represent a viable alternative). This could therefore push up the cost of reinforcing the network from the true least cost solution.

► The second issue is that this option also does not account for dependencies – for example the new combined options #1 plus #2 are still mutually exclusive with #3 due to meshing, however they appear as additive options.

Method 4: Separate 'sets' of cost options for the same peak MW in cluster

This method is an extension of Method 3 based on the notion that network reinforcement functions for some network components can be considered independently – for example for radial lines and transformers. This is because each of these components will 'fail' (i.e. will need to be reinforced) at a load flow independent of the other network components that are connected in series. For each independent component, reinforcement options are assessed, and ranked by cost and allowable capacity.

Hence within the POM, a given cluster peak load is effectively driving reinforcement across a number of independent network reinforcement functions *in parallel.* The way the peak load drives each function will need to be scaled based on the distribution of load as discussed in section 7.5.4.

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   166/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

The key issue with this approach is that where meshing is present the component costs are no longer independent – the peak allowable load is a function of the reinforcement of radial lines and meshing links (e.g. where voltage drop constraint is binding). As a result, this means that meshing as a network reinforcement option needs to be considered ex-ante by the tool.

Figure 7-15   Separate 'sets' of cost options for the same peak MW
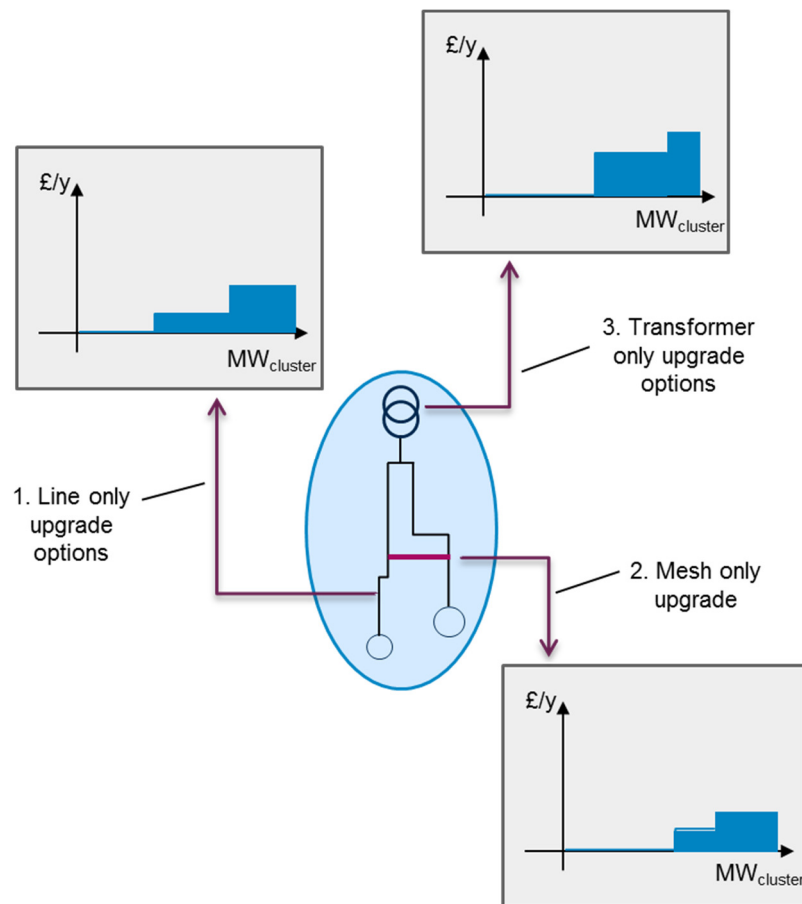
Method 4b: Separate 'sets' of cost options for the same peak MW – with/without mesh

As an extension of method 4 it would be possible to test the impact of meshing by creating two sets of network cost functions in all relevant cases (i.e. with and without meshing) and run two discrete sensitivities in the POM.

Figure 7-16   Separate 'sets' of cost options for the same peak MW – with/without mesh
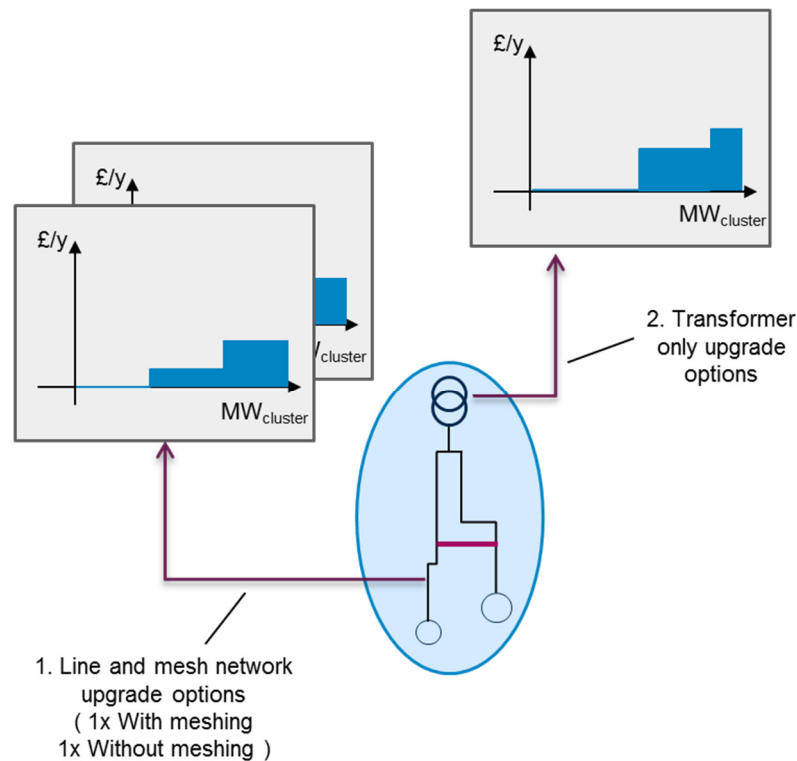
Method 5: Separate discrete options with interdependencies between years

This method is based on the notion that for many network reinforcement combinations the costs and/or increased capacity headroom associated with the upgraded components cannot be summed linearly as they are inter-dependent (e.g. the options are either mutually exclusive or a higher cost reinforcement option is first needed to enable a lower cost truly incremental upgrade). This can be illustrated using the following examples:

▶ Reinforcing a network line would give an incremental increase in capacity headroom of H_{line} at a cost C_{line}

▶ Creating a meshed link would give an increase of H_{mesh} at a cost C_{mesh}

▶ Performing both options, however, would not result in the sum of the above in terms of capacity headroom, i.e.

– $H_{line + mesh} \neq H_{line} + H_{mesh}$

▶ In terms of costs, this example would likely lead to similar costs being accrued (i.e. $C_{line + mesh} = C_{line} + C_{mesh}$), however there are other situations where it is possible that costs would not be the same, for example if savings were realised by reinforcing both a transformer as well as a line (i.e. $C_{line + transformer} \neq C_{line} + C_{transformer}$)

The incremental head room H_{mesh} from creating a meshed link is dependent on whether the line has been upgraded or not. A tree of "if" statements (reflected by conditional constraints using Integer Programming) could therefore be used to capture these dependencies within the

optimiser. As a result this is the most accurate approach for reflecting network cost functions, but may require a significant number of conditional values to be created, which may not be computationally tractable.

Figure 7-17   Separate discrete options with interdependencies between years
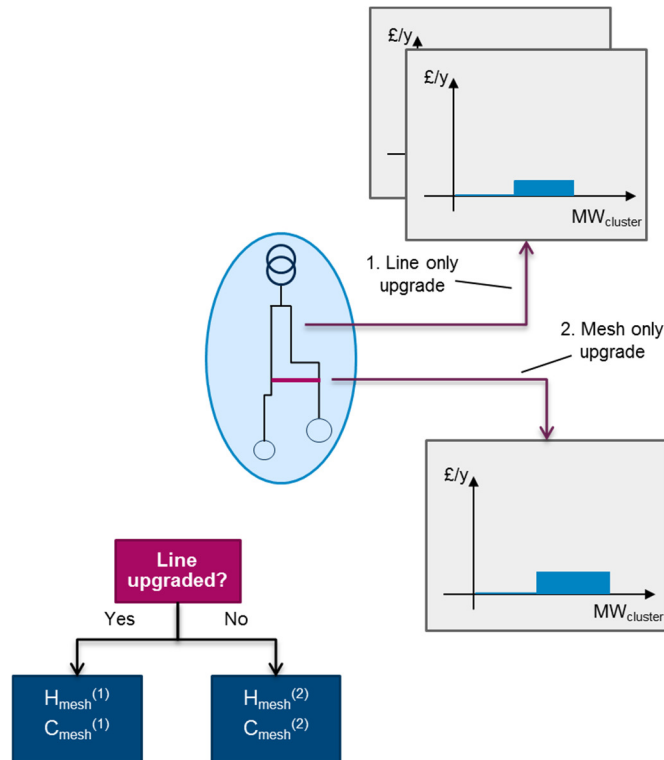
7.5.5. Heat network topology synthesis [NAM-004]

In a similar manner to electricity, a number of pre-processing steps are undertaken to help generate a representation of a potential heat network configuration. The first step is to calculate the maximum heat load-weighted centres for each, with the constraint being that the load-weighted points must be located on a road, which represents the route of the assumed future heat network in that zone.

These are then used to calculate the maximum load-weighted centres for a number of zones within a cluster, again with the constraint being that those points must be located on a road within the cluster (see Figure 7-18). The key point here is that the cluster load-weighted centre is a function of both the number and shape of the zones within the cluster itself, and will need to be re-calculated if either of these change (i.e. if the user re-defines the cluster in the SAM); the zone load-weighted centre is, however, static.

Figure 7-18   Simplified heat network – total load and load-weighted centers
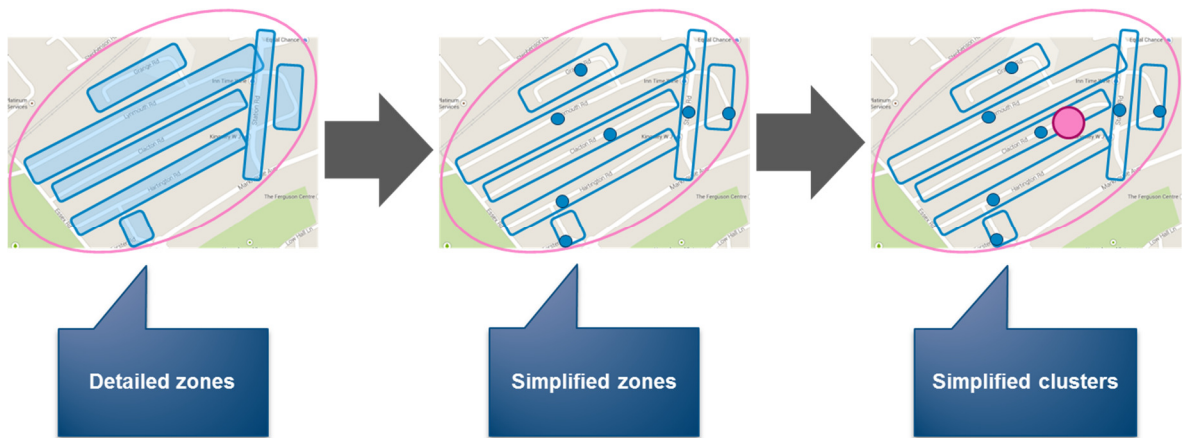
It is assumed that the heat distribution system (which could be an inter-cluster source) or the heat source directly supplying the cluster connect at the cluster load-weighted centre point and then supply all the connected zones within that cluster in a radial network (i.e. assuming no meshing within the cluster).

A skeleton heat network topology is then formed by developing pipe networks from the cluster load-weighted centre point to individual zone load-weighted centre points using minimum road distance. The network reinforcement/expansion options use this skeleton heat topology, but the POM then evaluates a range of different pipe/pump sizing options to determine optimal sizing for the cluster[40]. Finally, it is also assumed that the size of the intra-zone network does not change. In practice, this means that the size of the pipes supplying the customers connected along a zone is the same as shown in Figure 7-19.

Figure 7-19   Simplified heat network – heat network skeleton

---
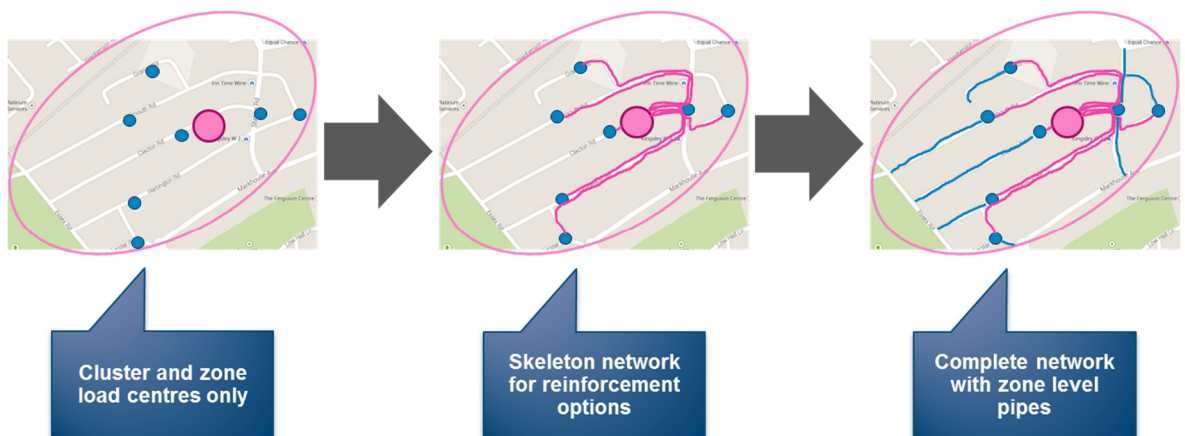
[40] It is assumed that full details of the original network outline would be stored, so that this can be refined further as part of the detailed network analysis of higher-level solutions from the POM.

Once the intra-cluster network topology has been established, the load weighted centres which reflect the likely location of a heat exchanger are then assumed to be the connection points for inter-node heat transmission between clusters as illustrated in Figure 7-20.

Figure 7-20   Developing cluster reinforcement cost functions for heat
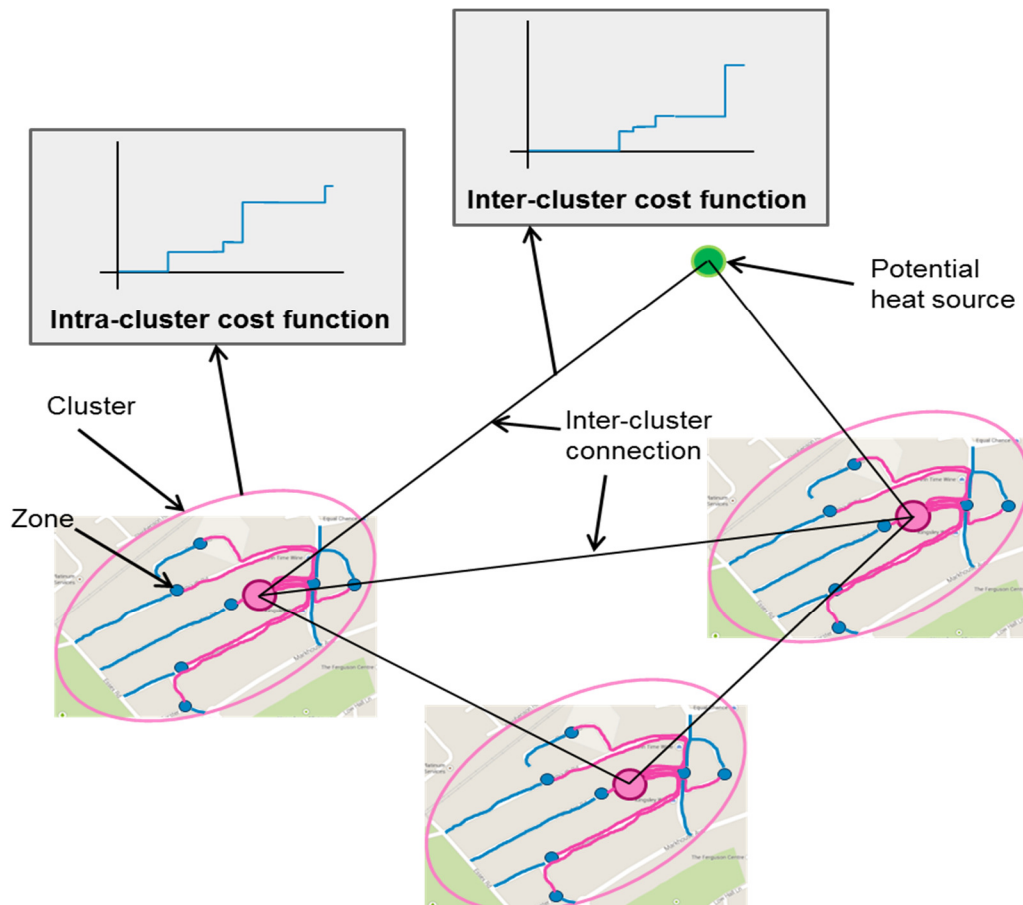
7.5.6. Heat network options and load flow modelling [NAM-005]

After the heat network topology has been defined for both within and between cluster connections a process analogous to that for electricity described in section 7.5.4 (but using the district heat module of PSS Sincal) is used to undertake dedicated network load flow modelling in order to ensure that all physical parameters underpinning heat networks are accurately modelled when testing different network designs.

This is separated into:

▶   A process to test heat network configurations within each cluster independently and generate the required reinforcement cost function considering e.g. required pipe length and size, heat exchangers, pumps, etc.

▶   A process to test the required pipe sizes for inter-cluster connections and generate the corresponding reinforcement cost function.

As noted previously, the decisions over what heat supply technologies to build and where to site them, along with the decision to connect the buildings to heat networks, are undertaken in the POM. At this stage the NAM is only providing a set of costed network reinforcement options to provide the POM with a holistic view of all the potential costs associated with deployment a heat network.

7.5.7. Gas/H2 options [NAM-006]

Unlike electricity and heat, v2.1 of the EPN does not undertake steady-state network flow modelling for gas and hydrogen (although this could be added in future drawing on the same basic process described above for electricity and heat and using the 'PSS Sincal – Gas module'). Instead the NAM produces "all-or-nothing" costed options for each cluster:

▶ Do nothing – continue to use the current gas grid and incur ongoing fixed operating costs.

▶ Extending the gas network to cover off-gas grid buildings. This is undertaken by using data from the SAM to understand the number of off-gas grid buildings in a zone and the total lengths of roads that are on and off the existing gas grid. The cost per metre is multiplied by that of the relevant pipe size (assumed to be the same as the connection point) adjusting for any double-counting of distance where multiple off-gas grid buildings would be connected via the same route.

▶ Decommissioning the existing gas network for operational costs savings. As above, data on the existing network (e.g. length) within the zone would be combined with per unit estimates of decommissioning costs to create a cost for the zone as a whole.

▶ Re-purposing the existing gas network to allow it to be able to transport hydrogen. As per decommissioning data on the existing network within the zone would be combined with per unit costs for re-purposing to hydrogen (the costs for changing appliances is estimated within the HOM and based to the POM as part of the building archetype data and potential hydrogen supply sources are evaluated in the POM).

▶ As above, but extend the repurposed gas network to include off-grid buildings.

7.5.8. Cluster-level network option cost functions [NAM-007]

As the POM only operates at a cluster level of spatial granularity the final step is to aggregate any remaining zone level functions within the final cluster boundaries. For heat networks and gas/hydrogen options this is not necessary as the final cost functions are only assessed at the cluster level as described previously.

For electricity, the final cluster cost functions are constructed using the *independent* sets of zone cost functions that have already been developed as well as the cost functions for all inter-cluster connections (existing or potential) as shown in Figure 7-21.

Figure 7-21   Developing network cluster cost functions for electricity
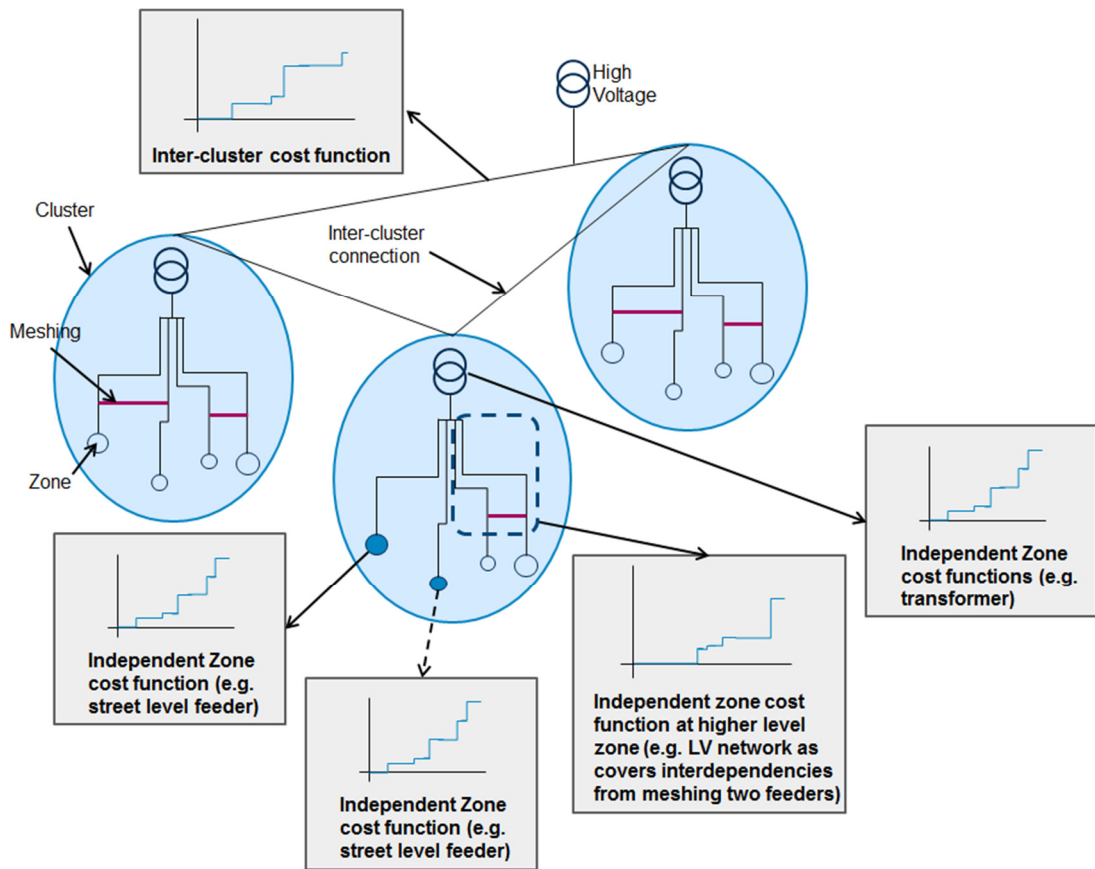
As described in section 7.5.4, when undertaking the original load flow modelling, it is necessary to make a simplifying assumption with respect to how load at different levels of network connectivity (e.g. street level feeder) is distributed when aggregating up to the next level of the network (e.g. LV network feeders and substation).   This is because the POM only calculates a single peak load (net embedded generation) value for the cluster as a whole.

For network options represented by cost functions at lower network levels (e.g. LV network feeders) driving reinforcement directly from the cluster-level peak load is not appropriate and would over-estimate the required reinforcement. Hence some of the zone cost functions themselves have to be adjusted before they used at the cluster level so that their interpretation as a function of cluster level load is appropriate.

If cluster load is assumed to be distributed amongst zones using a fixed % for each zone, the zone cost function can be "stretched" to be converted from MW$_{zone}$ to MW$_{cluster}$

▶   For example if zone load = X% of cluster load, then $MW_{cluster} = \frac{MW_{zone}}{X\%}$ for each zone.

Importantly the load distribution used *must reflect that used in section 7.5.4 as part of the original load flow modelling.*

7.6. Output validation checks

Output validation checks must be undertaken to determine whether the outputs from the NAM are fit-for-purpose for the overall modelling solution. This is because it is necessary to determine whether the derived cluster definitions and cost functions appropriately represent all energy vectors considered here, and particularly so for both heat and electricity which are modelled in detail.

In summary, the key checks that must be undertaken include:

▶ Ensure that all derived clusters respect the boundary constraints set by electricity network characteristics (as explained in Section 7.3.5). Clusters must be a collection of whole building blocks, or part of a single large building block, but cannot be a collection of partial building blocks;

▶ With regards to the derived cost functions it is also necessary to validate their accuracy in terms of: (i) materiality of load flow modelling; and (ii) materiality and uncertainty of network component costs. In practice, the former means checking that all network constraints that could have a significant impact on costs are taken into account for each test load flow simulation, whilst the latter means checking that any other material physical or geographical factors that may affect component costs are included in the modelling approach, which is likely to be undertaken in conjunction with the Local Authority.

8. Pathway Optimiser Module (POM)

8.1. Overview

The EnergyPath™ Networks Tool supports strategic local area energy system prioritisation and planning, focused in particular on choices for buildings and networks, alongside local area energy generation and storage, over the pathway from now to 2050.

As described in the previous sections the HOM / SAM / NAM modules help to create a detailed spatial picture of the existing local energy system, along with the *options* available for evolving the system over time such as building retrofits or installing a new district heat network. The other modules also create the necessary data about the options (cost, performance and various other characteristics) so that the POM can compare and trade-off these options against each other. However, to make the pathway analysis a tractable problem the level of detail needs to be simplified, for example, by reducing the spatial detail from individual units or zones (see Figure 6-2) to a smaller set of clusters.

As shown in Figure 8-1, POM is effectively the automated decision-making engine focused on

▶ **What *options* should we deploy, where and when?**

▶ For some *options*, **how** should we utilise these once deployed?

The process needs to consider whether the options that are chosen as part of an overall viable energy system pathway for a local area are collectively appropriate (e.g. cost-effective and feasible) and satisfy various other goals (ensure households are provided with sufficient comfort, is consistent with meeting our climate change targets)

Figure 8-1   **POM context**

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification

175/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.
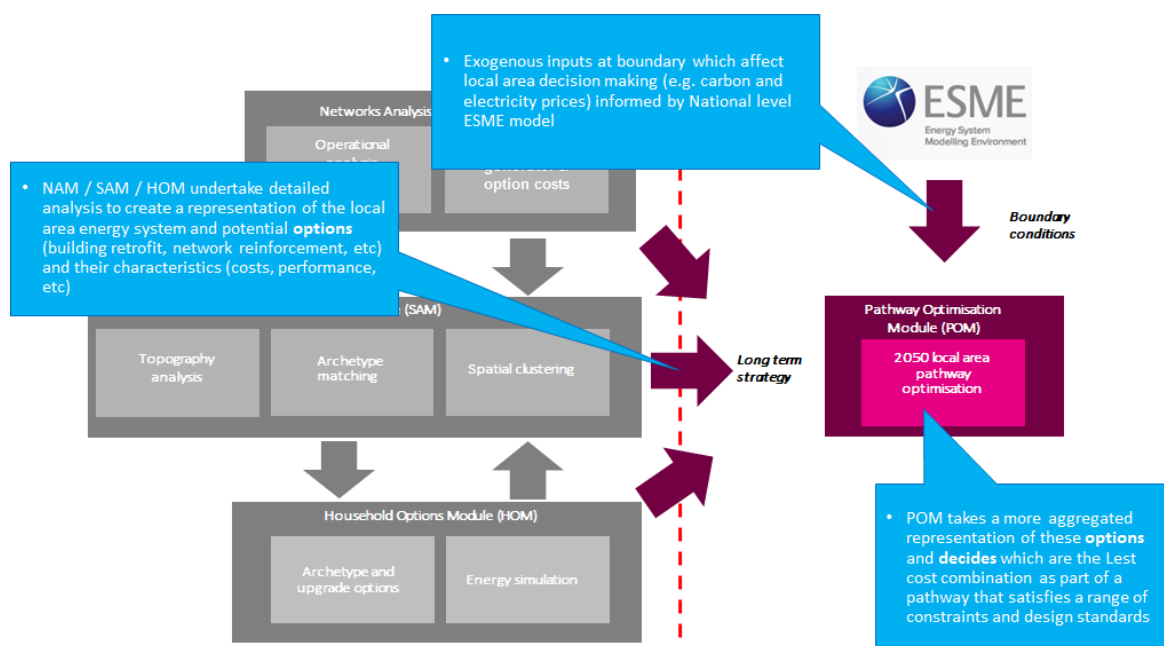
To make the problem of comparing available multiple options more tractable, their representation in the POM must be simplified relative to the other modules, for example, dividing the area into relatively small number of clusters (e.g. <50 clusters versus 1000 zones). Simplification naturally involves a loss of detail and may then affect the options chosen. To counter this two key features are incorporated:

▶ In-built flexibility to vary easily the level of aggregation/simplification and understand how the solution changes as you move along this spectrum e.g.:

- Number of time periods on the pathway to 2050 (e.g. annual to 10 year steps)

- Number of time slices within year (e.g. half hourly demand profiles to peak and annual average)

- Number of spatial clusters in the SAM

- Number of building pathway archetypes in the HOM (i.e. set of future choices for building efficiency and heating system upgrades) and analogous set of user-defined non-domestic pathways in the SAM

- Number of discrete network investment options for electricity and heat as defined in the NAM. For gas and hydrogen the number of options are always fixed per cluster[41]

8.1.1. Module diagram

The diagram below shows the key logic / process steps for the POM, the equivalent for the full tool covering all modules is shown in section 11.

---

[41] Maintain existing gas network, decommission, extend network to off gas-grid, repurpose gas network to hydrogen, extend and repurpose to hydrogen

Figure 8-2   Key POM logic / process steps
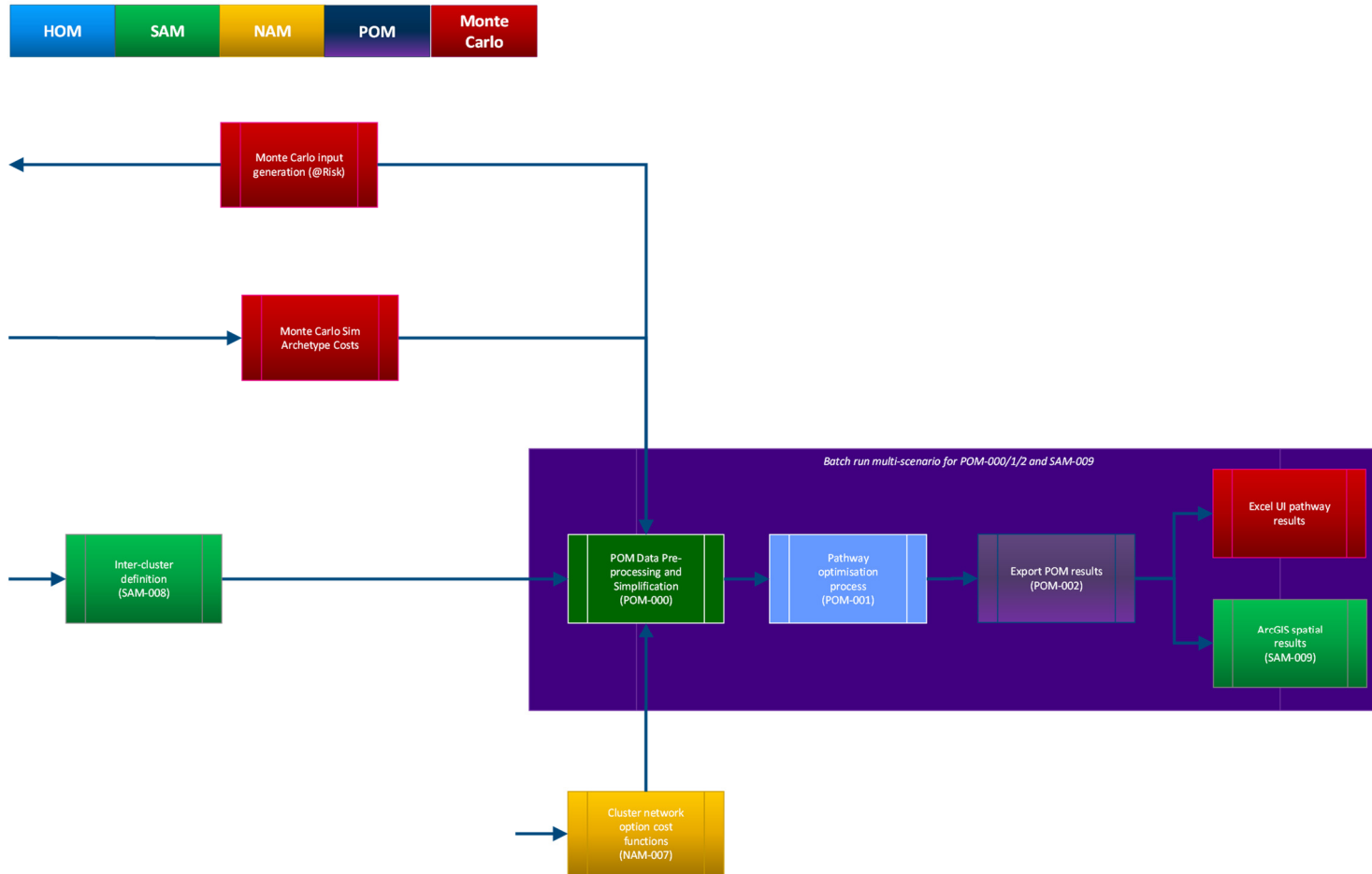
8.2. Key outputs

The POM provides the main outputs from the EnergyPath™ Networks Tool, which are ultimately used to help inform decision making by Local Authorities and other key stakeholders. This is discussed in more detail in section 4, but at a high-level the key outputs from the POM must be able to be re-packaged and interpreted such that they help understand:

▶ What are the high-level features of the overall area viable energy pathways?

– Deployment of specific network, building, other options (or more broadly *'types'* of pathway design such as heat network focus versus electricity reinforcement)

– At what point should they be built over the pathway to 2050?

– What are the direct impacts such as on emissions or energy 'costs'?

– When POM outputs are combined with socioeconomic or other data, what are the distributional impacts, e.g. the potential effects on fuel poverty?

▶ What are the key geographical and underlying features of the pathway, in particular where in the local area are specific options being deployed?

▶ What are the key areas of uncertainty (this is discussed in more detail in section 9)?

– How much confidence is there that an option is 'low regrets' within a pathway (in terms of the cost of the energy system or other metrics), as well as how resilient or sensitive the option is to changing key input assumptions (such as the fossil fuel prices rising faster than anticipated)

– How can the decision maker understand what the value / impact is of reducing uncertainty by undertaking further primary research (e.g. surveys to better understand the cost of retrofitting local buildings)?

▶ What does the pathway design mean in terms of specific projects?

– The outputs need to contain sufficient information on individual options, or groups of options, that can be packaged into potential projects. This should cover data on the type of option, capital and operating costs, size or number, deployment dates, etc. This would be used to target more detailed FEED (Front End Engineering and Design) studies to enable the specific projects to undertaken.

The table below outlines the fundamental outputs of the POM, which can be processed and combined in a number of different ways to provide the insights required and generate some of the more specific outputs described in section 4.

Essentially, all information regarding all the options and choices made in the POM (see section 8.5.2) are available in the outputs (e.g. what, where and when a technology option is deployed). These can then be combined with other static input parameters (e.g. costs of deployment or

emission factors) to understand the impacts of the selected options (e.g. total costs of deployment of pathway or CO2 emissions).

Further optimisation-specific information can also be generated such as the shadow (marginal) price of key variables such as electricity, and the slack remaining on constraints (see section 8.5.3) to understand which of these are *binding* for a given pathway in the POM.

Table 8-1   Output data summary

Data field	Purpose	Data granularity

Stock of building pathway archetype choices	• Choice of building pathways reflecting decisions about efficiency improvements and changes to heating systems (including switch to district heat networks)
	• Covers both domestic and non-domestic pathways
		Stock of archetypes that exist in given timeperiod as well as timeperiod of conversion of archetype in each
		   • Cluster
		   • Constraint Class

Stock of network options and network new build options	• Components of Pathway design	Stock of network options that exist in each timeperiod as well as timeperiod of build in each
		   • Cluster
		   • By each energy vector (electricity, heat, gas)

Energy / supply demand balances	• Components of Pathway design	Energy inputs and outputs to each technology option (building, embedded generation, etc) by
		   • Cluster
		   • Timeperiod
		   • All timeslice
		Intercluster network flows by
		   • Timeperiod
		   • All timeslice
		By extension other necessary results can be calculated
		   • energy consumption,
		   • resource use
		   • CO_2 emission results,
		   • Peak demand
		Note that level of timeslicing may vary by product (e.g. full within day for electricity/heat, daily for gas)

Stock and new of other energy system features (CHP, DHN boilers, network storage, etc.)	• Components of Pathway design	Stock of technology options that exist as well as year of build each
		   • Cluster
		   • Time period
		By each energy vector (electricity, heat, gas)

Operation of other energy system features	• Operation within pathway design	For flexible technologies, the level of utilization by
		
		    • Technology vintage
		
		    • Cluster
		
		    • Timeperiod
		
		    • Timeslice (level of timeslicing dependent on the products the technology sees as an input/output)

System costs	• Understanding of pathway impact	Underlying cost components of deployed options (e.g. CAPEX, FOM, VOM, fuel costs) can be calculated for each option by
		
		    • Cluster
		
		    • Timeperiod

System shadow prices	• Understanding of pathway impact • Detailed analysis (test whether simplified operating profile chosen appropriate given overarching system parameters such as electricity costs)	Data granularity defined by granularity of constraint which generates shadow price – i.e. if imposed by timeperiod, cluster, timeslice – results will be available on this base An example is electricity within the main S / D balancing constraint, which would be available at the most detailed level of temporal and spatial granularity
Constraint slack	• Understand which constraints are *binding* the pathway design, and if not how close are they to binding	Data granularity defined by granularity of constraint – i.e. if imposed by timeperiod, cluster, timeslice – results will be available on this base

8.3. Key functional requirements

8.3.1. Optimisation

For a given set of input assumptions to the POM (cost of available options or wider exogenous assumptions such as fuel prices) there are likely to be too many trade-offs to explore all possible combinations of options manually, to understand what the 'best' combination is for a pathway.

An analogy is the DECC 2050 calculator pathway calculator[42], which contains approximately 40+ levers and 4 basic settings for each lever. Rationalising this limited set manually is time consuming enough without adding further complexity to the choices such as when or where over the pathway an option is undertaken, which is of critical important for the EnergyPath™ Networks Tool. To add further complexity, the significant uncertainty over multiple input parameters over the pathway to 2050 means that resolving the 'best' combination of options needs to be repeated under multiple sets of input assumptions.

Optimisation is considered to be the most appropriate technique to help resolve the huge number of possible trade-offs. It is effectively just a 'calculator' that can resolve all possible combinations of choices simultaneously, whilst maximising or minimising some objective, such as profit or cost. Constraints can be added to restrict inappropriate combination of choices to reflect real world issues or design standards that must be achieved, e.g.:

- ▶ "No building of technology X before year Y in location Z"

- ▶ "CO2 emissions must remain below limit of X"

- ▶ "What would happen local area network requirements if all buildings were forced to achieve a minimum of retrofit standard X"

Provided the underlying optimisation model is a good 'representation' of the problem this then allows the user to focus on how the optimiser design output varies under different sets of inputs assumptions. For example, in the context of the EnergyPath™ Networks Tool a good representation of a local energy system is one which ensures that supply and demand are balanced across the system appropriately (meeting peak demand requirements) and, as a direct result, ensure that householders' energy service demands and comfort requirements are met.

8.3.2. Efficient uncertainty analysis

'What-if' analysis will be a fundamental part of the process of using the POM, for example:

- ▶ Undertaking single deterministic runs changing one or more input assumptions and exploring the impact on the pathway design

- ▶ Extending this simple what-if analysis to understand the resilience of a pathway design (by fixing key elements of the solution such as the deployment of a heat network) and

---

[42] http://2050-calculator-tool.decc.gov.uk/pathways/111111112111111111111111141411111111111111111111111111/primary_energy_chart

understanding how the outputs of the POM (e.g. the cost of the energy system) change in response to different input assumptions (higher gas prices)

However, there a large number of possible input parameters that could be changed and a wide range of potential uncertainty on these parameters over a 30+ year pathway, in particular the cost of technology or fuel prices.

The POM allows the user to explore a wide set of uncertain input parameters in an efficient manner via the use of Monte Carlo simulation of a number of key input parameters (described in more detail in section 9). This creates a number of *sets* of input parameters each of which is optimised deterministically. The outputs then comprise a *set* of results from which distributions, rather than discrete numbers can be established.

A formal *stochastic optimisation* within the POM, as opposed to the *probabilistic simulation* process we have described above, has not been proposed for two main reasons:

▶ The rationale for applying a formal stochastic optimisation is to understand the optimal 'hedging decision' in the near term given future uncertainty, and recourse strategies should the future turn out to be different to that expected[43] - i.e. the best decision making given what we *are expecting* to happen. By contrast the uncertainty which needs to be explored in the EnergyPath™ Networks Tool is subtly different and is primarily about understanding the overall pathway design that a local area should be *aiming for*, over a long-term pathway, considering the inherent uncertainty over this extended time horizon in the round

▶ The problem size increases exponentially, quickly becoming intractable and limiting the number of variables that can be considered stochastically, whereas a probabilistic simulation scales linearly with the number of discrete optimisations to be undertaken (and which can potentially be solved in parallel)

8.3.3. Accurate energy system representation

The more accurate the representation of energy system choices (and by extension trade-offs) in the POM the better the optimiser solution, in particular the accuracy with which the costs and the fundamentals of energy system operation are reflected.

A more accurate representation is generally achieved via combination of more granularity and adaptations to the structural representation of the optimization problem, such that it mimics reality as closely as possible:

▶ **Granularity** – increased granularity can be achieved by having a more detailed spatial representation (more clusters each of which covers a smaller area), more timeperiods (i.e. annual rather than 5- or 10-yearly steps on the pathway), or more timeslices within year (more characteristic days and periods within day)

---

[43] It is often used for decision making around energy storage – i.e. do I inject or withdraw now given my expectations about future energy prices

- Finer spatial clusters offer more accurate information on network reinforcement/new build costs

- More timeperiods offers a more accurate representation of how costs will evolve over time and the impact on the timing of investment (as opposed to averaging these costs over a number of years)

– More timeslices within year offers a more accurate representation of energy balancing and by extension the costs of operating the system. In particular there is a trade-off between POM choosing how best to operate a flexible device given the wider system conditions and pre-judging its operation. This is important to understand the suitability of more flexible technologies such as hybrid boilers, storage, micro-CHP and the impact on the demand profile, net peak demand[44] and emissions

– More granularity in pathway archetype choices provides a more accurate reflection of the existing building stock and possible upgrade options for both domestic and non-domestic buildings

– More granularity in network investment options for electricity distribution and heat networks.

▶ **MIP (Mixed Integer Program)** representation for some of the optimisation variables which can provide a better representation of the cost of options than an **LP** (Linear Program) representation, for example:

- *Lumpy* investment of some options (e.g. new large scale CHP) as under an LP formulation the POM could build any idealised fraction of a new plant it wanted, which would likely underestimate the costs

- Discrete (and potentially interdependent) choices such as being able to fully repurpose the gas grid in a cluster to use H2, which in reality is likely to be 'all-or-nothing' decision. The alternative under an LP representation is an idealised partial repurposing of the network in a cluster, which is not feasible from an engineering standpoint

The basic representation of an energy system structure within the POM is outlined in Figure 8-3 and is essentially the same as that within the ESME model or MARKAL/TIMES energy system models. Starting from the left hand side are a number of individual energy resources (e.g. biomass, gas, wind, etc). These can be used directly in an end-use technology (e.g. gas in a building archetype) to produce the required energy services (e.g. space heat, hot water, etc) or converted via intermediate routes into a number of energy carriers (e.g. gas in a power plant to produce electricity) which are then used in the final end-use technology. A number of intermediate conversion steps can exist (e.g. the earlier electricity step could be used to electrolyse hydrogen which is then used in a boiler or fuel cell). As part of the transport of

---

[44] I.e. peak demand in the cluster net of embedded generation, and the resulting implications for network reinforcement and new build

resources and energy carriers across the system network technologies may also exist, such as an electricity distribution network.

In theory the conversion technologies can reflect a many-to-many relationship with inputs and outputs. I.e. they can take in one or more inputs (resources or energy carriers) and produce one or more energy carrier outputs (e.g. in a co-fired CHP plant). Similarly, end-use technologies may be able to use more than one input (e.g. in a hybrid gas boiler). In addition, storage technologies exist which can temporarily store and release an energy carrier or resource (e.g. electricity or gas).

The POM decides what resources to use, what technologies to build (conversion, storage, network, end-use), how these are operated (and indirectly what energy carriers to produce), such that *all* energy service demands specified within the scenario are satisfied. The key constraint that must be satisfied across the entire energy system structure as part of this is that supply must also equal demand for all other energy carriers and resources.

For example, if 100 units of energy carrier 2 are required by end-use iii (to produce sufficient quantities of energy service 2) these 100 units must be produced by conversion technology 3. By extension sufficient inputs of energy carrier 1 must be available to feed into conversion technology 3, with the exact quantity dependent on conversion technology 3's efficiency. The process extends back through the chain such that, a sufficient quantity of energy carrier 1 must have been produced from elsewhere and so on, back to the original resources, which must also be available in sufficient quantity.

Figure 8-3   Example of basic energy system model structure
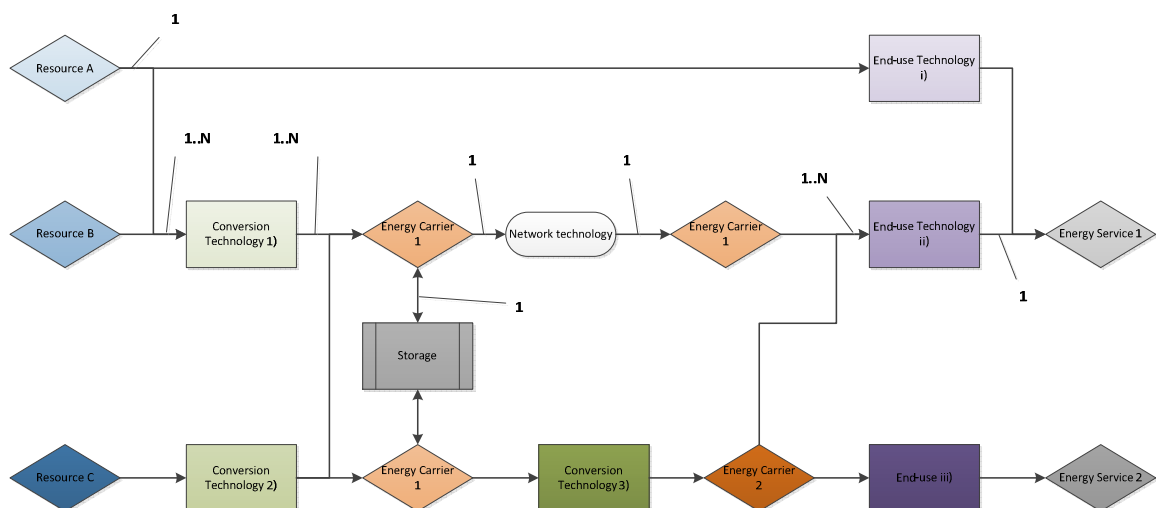
For further details of a standard energy system modelling structure please refer to the ESME Functional Specification document.

8.3.4. Other

Two other key related requirements that have informed the POM decision module are that it is

▶   Easily scalable with respect to

- Adding more 'options', which can be considered as part of the trade-offs in the creation of a pathway (e.g. building pathway archetypes, network options, fuel types, etc)

- Granularity in terms of the numbers of spatial clusters, timeperiods and timeslices. This has significant implications for POM performance, but flexibility to simplify (or increase) the complexity of the problem is built directly into the EnergyPath[TM] Networks Tool and is discussed further in section 8.6.1.2. This flexibility is important because it is difficult to say ex-ante where additional detail matters most and may vary by local area given differences in underlying data.

▶ Reasonable performance limits

- The POM is by definition trying to resolve a highly complex optimisation problem, but to allow 'reasonable' user interaction this leads to overnight solving (~15 hours) as a typical working threshold[45]. For probabilistic simulations there is also the ability to distribute the optimisation problems across 3 pre-configured machines in parallel[46], with the problem setup and consolidation of output results managed by a single 'master' machine

- The scalability described above along with other routes to simplify the optimisation problem, allows performance to be tailored and the POM used in different ways. For example, a 'simpler' representation of the local area energy system can created and more simulations explored to understand the implications on the pathway design due to a wider range of uncertainty, versus a very detailed representation for 1 deterministic scenario.

- MIP optimization in particular can lead to a significant performance decrease and we propose to retain a parallel MIP/LP representation for key variables as a further simplification option. But, as a result we are not considering the use of **NL-MIP** (Non-Linear Mixed Integer Programs) even though this may in some circumstances allow us to better represent reality[47].

- A number of performance enhancements have also been implemented in v2.1 to speed the solving time of the MIP problem including the use of SOS (Special Ordered Set) structures where possible, warm MIP starts[48], tuning of the solver settings, defined branching priorities for variables[49]

---

[45] This is broadly comparable to an ESME pathway simulation run with ~100 simulations
[46] As used by the ESC as part of the programme of analysis, each machine has the relevant set of software licences
[47] For example representing a non-linear cost function in its original form versus a piecewise-linear approximation of this function.
[48] I.e. reusing a previous solution to "kick-start" a new solve
[49] I.e. using an EPN model-specific understanding of what the most important variables are likely to be in driving a least-cost solution

8.4. Key inputs

The SAM / NAM / HOM modules provide the majority, but not all of the input information necessary for the POM, which can be categorised into 3 main areas:

▶ Direct inputs from a single module, which provides all necessary data

– This may include the parameters necessary to undertake Monte Carlo simulation of the inputs

▶ Partial inputs from other modules, which need to be combined to provide a full set of necessary data

– The SAM primarily provides spatial IDs/Flags for the components of the energy system (existing or potential options) that are contained within a cluster such as the number a given archetype. These are combined with other datasets (e.g. from the HOM) to create all the information that is needed by the POM (e.g. the costs and demand profiles associated with the base archetype and possible conversion options)

▶ Primary data inputs including

– Those necessary for pre-processing of POM options (such as diversity factors or other technology information on existing non-building, non-network energy system features)

– National pathway boundary conditions from ESME

– Others user-defined scenario inputs which are not already generated as part of other modules such as technology / option build rate constraint values or Monte Carlo correlation factors

The tables below describe what is technically required within the tool and how it is used. The relative importance of the primary data inputs and other factors such as costs and licensing restrictions, particularly where they must be purchased from an external provider, are discussed in the original deliverable (D4) Data Acquisition Plan. It should be noted that the ESC currently maintain the master set of input data for the current v2.1 EPN tool and this is not documented here.

Table 8-2 **Input data**

Data field(s)	Input type	Purpose / sub-fields	Granularity	Uncertainty parameters	Coverage	Source	Limitations
Cluster-specific ID / flag information	Secondary	Provides the POM with the data on number of clusters existing / possible interconnections between clusters and ID/ flag information on • # of each existing domestic & non-domestic building archetype in cluster X • Cluster & building level restrictions on archetype retrofit / other options as inputted by the user within the HOM/SAM • Other existing non-building / non-network energy system component (IDs only) • Future energy system components that *will* be constructed (e.g. new build housing zone and base archetype) • Future energy system options for siting of new technology that *could* be constructed (e.g. new CHP, energy storage, etc) or with possible local resources, which could be exploited (e.g. biomass, waste, geothermal) I.e. the SAM provides information on where energy system features are sited by cluster (existing, future deployment or potential future option) which can be combined with other data to create the full set of characteristics	Data by cluster, by timeperiod	N/A	By cluster and inter-cluster	SAM	Cluster definition naturally loses spatial information within it
Database of option characteristics for non-building, non-network options	Primary	Combined with the ID flag provided by the SAM this creates the list of technology characteristics used to define the option in POM • Costs (CAPEX, FOM, VOM)	Data by technology option / resource product, by timeperiod (different	Capital costs for future vintages are simulated in the Monte Carlo sub-module.	Covers embedded generation/CHP, other heat network sources, network storage	*Existing* data will have to be compiled in conjunction with LA and DNOs	N/A

(*both* existing and new)		• Type (flexible profile, fixed profile, storage) • Modes of operation • Input / output products by mode • Efficiency by mode / (losses for storage) • Lifetime (economic, technical, construction period) • Annualisation factor • Max availability (for flexible profile plant) • Load factor (for fixed profile plant) Separately this also need to cover resource availability by cluster	characteristics by vintage), by cluster		(electricity, heat), etc	*Future* data from various ETI sources (e.g. UK ESME database or specific technology programmes)	
Building archetype pathway option parameters	Secondary	Combined with SAM ID for number of each existing archetype in each cluster to create all archetype pathway characteristics needed in POM • Viable 'archetype pathways' for domestic and non-domestic buildings – i.e. for each pathway containing the set of final energy demand profiles and costs associated with that pathway • Constraints can be applied to restrict retrofitting options at the building or cluster levels	Data by archetype cluster, by timeperiod Demand profile(s) for each half hour in each characteristic day Annualised cost per timeperiod for each pathway	Archetype conversion costs, final energy demand profiles (driven by heating system efficiency) are simulated in Monte Carlo submodule.. Demand profiles (driven by target comfort temperature) also vary by scenario.	All relevant archetypes	HOM	# of remaining domestic archetypes used to define area are subject to simplification in HOM
Archetype demand diversity scalars	Primary	Used to adjust HOM base demand profiles for diversity effects on the network to avoid overestimating peak demand, separate scalars for • Heat-related electricity demands • EV electricity demand • Other electricity demand • Heat network demand	By archetype, by characteristic day timeslice	N/A		Existing engineering estimates (e.g. WPD, Ramboll) Analysis via Monte Carlo simulation of demand profiles in HOM	

					Survey data	
Network reinforcement and new build options	Secondary	All data necessary to define network options within POM • Costs (CAPEX, FOM, VOM) • Lifetime (economic, technical, construction period) • Efficiency (nominal input & output quantities) • Availability of nameplate capacities Options already contain data on existing network features	Data by network option, by timeperiod (different characteristics by vintage), by cluster and by inter-cluster connection	Capital costs of network and transmission lines' materials and overhead (as described in section 7.3.4) are simulated in the Monte Carlo sub-module. Cost scalars can also be applied to capital costs on a deterministic run.	Electricity, heat, gas, hydrogen	NAM	Cluster definition naturally loses spatial information within it
National Pathway boundary conditions	Primary	To help ensure pathway design at local level aligned with national level pathway. Range of possible conditions, but is likely to contain at a minimum • Carbon shadow price • Cost, availability and carbon intensity of transmission connected electricity and heat at boundary of local area • Cost of increasing available peak supply at boundary • Fossil fuel and biomass prices • Relative cost trends by timeperiod for key technologies	By timeperiod By timeslice, by cluster where relevant	EPN maintains consistency between national pathway simulation results and local area simulation e.g. for product market prices, technology capital costs variation range, correlations.	Local Area	ESME	More limited resolution (spatial, temporal, technology aggregation) means some intermediate processing is needed to translate direct ESME values into the EnergyPath™ Networks Tool boundary conditions Need to ensure consistency of national level inputs (e.g. fuel prices) which generate a given set of national outputs (e.g. carbon shadow price)
Other user defined scenario assumptions	Primary	This includes other necessary scenario assumptions not already defined through data in other modules passed to POM. In particular	By technology option, by timeperiod, by cluster, by	Some constraint values could in principle be simulated	Various	Various – ETI, LA, etc.	N/a

| | | Local area industry final energy demands over time (by fuel, profile) Key pathway constraints such as design standards ('peak winter' de-rated reserve margin), cluster level restrictions on what can be deployed (building, network, energy supply) or forced deployment requirements, etc Correlation factors for Monte Carlo simulation | timeslice as relevant | Non-domestic energy demands and retrofitting costs can be simulated.

Non-domestic energy demands can also vary by scenario. | | | |

8.5. Key optimisation components

8.5.1. Objective function

The core cost components of the POM objective function to be minimized are

- ▶ Option costs (e.g. building archetype retrofits, network reinforcements, etc) covering CAPEX, FOM (Fixed Operating and Maintenance), VOM (Variable Operating and Maintenance) costs.

- ▶ Resource costs for fossil fuels, biomass, and energy inputs that occur at the boundary of the local area (e.g. national transmission connected electricity prices).

- ▶ Carbon price (note that POM will also include the ability to cap absolute emissions)

The impact on individual agent incentives or wider distributional impacts, and by extension the consideration of profit or consumer/producer surplus maximising behaviour, is not considered directly within the primary objective. *This means that taxes, subsidies and other distributional transfers are not included.* By extension the cost of land is not included where this represents a straight transfer from buyer to seller. However, it may be useful at a later stage to value the opportunity cost from the perspective of 'society within the local area' of converting land from one use to another – e.g. the conversion of productive farmland versus a brownfield site to an energy centre. This could be undertaken within the modelling framework by adding a location specific 'resource cost' for development. The values themselves are difficult to estimate and it becomes important that such costs are applied appropriately across the local area so as not to bias the cost optimisation.

The objective function considers the minimisation of costs across the NPV (Net Present Value) of pathway costs from the base year until 2050, rather than trying to minimise the costs in each 'spot year snapshot'. As part of this process the CAPEX is annualised (with varying discount rates by option[50]) to avoid pathway boundary issues such as incurring the full CAPEX cost of investment made in 2050 even though the investment will last beyond this point.

In addition, an overarching societal discount rate is applied reflecting the preference for delayed costs. The benefit of minimising the total pathway costs is that this better captures the trade-offs around timing of investment such as early oversizing versus incremental upgrade.

8.5.2. Key decision variables

The decision variables represent the choices that POM is able to make as part of determining the least cost pathway design given a set of input assumptions, these are summarised in the table below.

---

[50] This could be used to indirectly test consumer preference with respect to technology choice, e.g. domestic building retrofit options could have a much higher discount rate than non-domestic options or network upgrades (particularly as the latter are operated under pre-defined rate of return business models)

Table 8-3   Summary of key POM decision variables

Item	Timeperiod	Location	Timeslice*
Building archetype pathways (domestic and non-domestic)	Choice of pathway for the overarching time horizon from now to 2050 from available valid pathways. This includes possible application of building or cluster level constraints.	Separate choice of archetype (resp. non-domestic) pathways in each cluster, archetype (resp. non-domestic) group and possible constraint classes.	The energy demand profiles associated with the pathways are averaged to reflect the final timeslice definition used in the POM and are considered as part of the overarching supply/demand balancing
Networks build and operation	Choice in each timeperiod for reinforcing or deploying new network features for various energy vectors (electricity / heat). For gas discrete choices between ongoing use, expansion to off-gird, decommissioning or repurposing to hydrogen (potentially also with expansion)	Different choice of network option by cluster as well as options between clusters (except for gas, which is assumed large enough to supply current and future demand).	Additional decision variables to track the effective supply/demand balancing of network energy flows within & between clusters, for each relevant timeslice (for electricity and heat this covers both characteristic days and within day, but for gas only daily balancing is necessary)
Resource use	Choice of quantity of resource to be used. The degree of timeslicing will be dependent on the product – e.g. biomass, waste, geothermal and coal only need to consider supply/demand balancing at the annual level	Resource use may be cluster dependent in some cases – e.g. for biomass or geothermal, whereas others such as coal choice is at aggregate local area level only	Gas supply/demand balancing is considered at the daily level
Boundary heat / electricity / other energy vectors such as hydrogen	Choice in each timeperiod for use of resources given available capacity Choice in each timeperiod to reinforce boundary to access more peak demand and annual supply	Choice in clusters that are effectively connected at the boundary, network flows then ensure supply/demand balancing between clusters to point of demand	Choice for each timeslice within each characteristic day
Other network embedded generation / storage technologies for e.g. electricity and heat	Choice in each timeperiod for build of technology	Choice within each cluster	Choice of dispatch by timeslice for flexible technology (e.g. CHP / storage) Fixed profile dispatch for others (e.g. embedded wind / solar)

A number of additional complexities exist with respect to the deployment of a particular option, be it building archetype conversion, network or other technologies:

▶ For some options (e.g. network build or embedded generation) the variables are also *vintaged* by the year of deployment. For example, CHP plants may be built in 2015 and 2020 and both will still be operational in 2025 given operating lifetimes. The 2015 and 2020 vintages of the plant will likely have different characteristics (e.g. the new plant

may be more efficient) and hence the choice of how to operate them in 2025 may be different depending on the vintage.

- ▶ In addition, for some option build variables (such as large scale network upgrades) there can be a user-defined construction period effect, whereby if development takes more than one year to complete some costs are incurred from the start of construction, but the operating ability of the option (e.g. to flow energy through the network as part of supply / demand balancing) may not be available instantaneously. When combined with the optimisation across the full pathway, this helps to understand better the real timing of investment, particularly where network reinforcement needs to precede additional load.

As discussed in the table above a number of the variables associated with the supply / demand balancing are timesliced at different levels (annual, characteristic day, within day) to simplify the problem optimisation problem. For example, it is not necessary to understand the supply / demand balance of coal at finer than an annual level as we can reasonably assume that stock-piling and supply chains are sufficiently developed to ensure that supply is available where it is needed. However, given the issues and cost associated with large-scale electricity storage and the need to more realistically account for peak demand as the key driver of capacity requirements it is necessary to track the supply / demand balancing of this product at the most granular within day level for each characteristic day. The number of periods represented within day in the POM is flexible as discussed in section 8.6.1.2.

Finally, it is important to note that the more focused scope of the EnergyPath^{TM} Networks Tool means that a number of more 'traditional choices' found in national level energy system models are not appropriate or covered. All of these factors will still need to be replaced by exogenous input assumptions informed by national pathway modelling from ESME or other sources (with variations tested via sensitivity) as they still indirectly impact on the local area energy system pathway. Factors not reflected by explicit decision variables include:

- ▶ Choice over the deployment of different vehicle types (e.g. H2, EVs) in the local area as this should be informed by developments at the national pathway level. Separate EV charging profiles are layered into the building archetype electricity demand profiles subject to external assumptions about the shape of the profile, ownership of EVs by cluster, diversity impacts in charging profiles

- ▶ Deployment of new large scale transmission connected electricity generation as this is outside the boundary of the network system modelled (up to 33kV/11kV substations as outlined in section 7.3.1). However, as indicated in the table above, the choice to use this external supply (at a given cost) and expand the implied peak capacity of supply (at a given cost) at the boundary is factored into the decision variables, so that this can be traded off against alternative options within the local area such as efficiency measures, fuel switching or more embedded generation

- ▶ Choices around industry energy use such as fuel switching, building of new boilers or on-site generation within the boundary of the industrial site. The potential supply/cost of electricity / waste heat at a large industrial site boundary are assumed to be exogenous.

8.5.3. Key constraints and 'design standards'

Constraints restrict the possible combinations of variables (options) that can be explored by POM to find a least-cost solution to mimic real-world issues. These can broadly be divided into core constraints (which must apply at all times to provide a sensible energy system representation) and optional constraints, which further restrict the possible solution (and be extension increase the costs of the pathway design):

▶ Core constraints

– Supply must equal demand in all timeperiods, timeslices and clusters. This implicitly guarantees a feasible working energy system under normal operating conditions and by extension that household comfort requirements are met by the pathway design[51]. In addition, this supply / demand balancing requirement ensures that sufficient network capacity is developed to cover *net* peak demand (i.e. peak demand net of embedded supply within a cluster)

– Restrictions on the maximum (or optionally forced minimum) operating availability of options such as CHP or network components – i.e. this limits the maximum output or flow that can be produce or accommodated, respectively

– Reflection of mutually exclusive choices across the model in terms of building archetype pathways (i.e. for a given set of similar buildings in a cluster only 1 pathway can be chosen), network investment options (e.g. a binary decision to repurpose to hydrogen), or other local area technology options which would occupy the same land (e.g. the choice to build a CHP plant or solar farm, but it is not possible to build both in the same space).

▶ Optional constraints

– Restrictions on building archetype / network / other technology deployment. These can reflect a maximum absolute limit (including zero) or rate of deployment in cluster X or the total local authority area by timeperiod.

– It is also possible to invert the above constraint to force a minimum absolute limit or rate of deployment, for example, to calibrate deployment in the near term in response to a non-modelled policy incentive

– Restrictions on the maximum (or forced min) use of resources such as biomass in biomass in cluster X in timeperiod Y, as the alternative is to give them unlimited availability at a given cost

– Restrictions on maximum release of certain products e.g. to reflect a CO_2 emissions cap

---

[51] As the final energy demand profiles constructed in the HOM, which must be supplied appropriately in the POM, are themselves based on operation of heating devices, which adequately meet the required comfort levels in the different building archetypes.

> – A security of supply de-rated capacity margin, above modelled peak demand, reflecting e.g. the ability to meet a 1-in-20 cold weather period. The user can define the severity of this constraint through the input data.

Adding user-defined constraints through the data is a relatively flexible process. The broad nature of the underlying formulations (e.g. restrictions on deployment or forced deployments) and the flexibility to target at one or a group of options, means that a wide range of real world issues can be represented such as area specific planning constraints (meaning that no build is possible) or delays in development.

8.6. Key POM process steps

8.6.1. Data pre-processing and simplification [POM-000]

This sub-module automatically generates all of the inputs necessary to undertake the POM optimisation outlined in section 8.4. Two key points at this stage are:

▶ To create all the data that is needed, the POM must in some cases combine data from multiple sources (SAM, NAM, HOM and exogenous user inputs). For example, the SAM provides information on what archetypes currently exist in each cluster, for both domestic and non-domestic buildings. This needs to be combined with data from the HOM to understand what the viable future pathways archetypes are for these existing archetypes and the associated pathway demand profiles and costs.

▶ The SAM / NAM have already undertaken the process of aggregating data to the cluster level, but the remaining data which varies by timeperiod (e.g. the costs of investment options which may change over time) and timeslice data related to energy demand profiles is still highly disaggregated (annual, half hour for each characteristic day respectively). This disaggregated data is then subject to pre-processing in the POM as based on the user's defined level of simplification – e.g. to run the POM pathway in 10-year steps.

The following sub-sections outline some of the specific logic requirements in this automated pre-processing sub-module in more detail.

8.6.1.1. Cluster-level energy demand diversity

The HOM creates *base* final energy demand profile(s) for each individual pathway archetype in isolation (covering gas, network hot water for district heating, and electricity). However, each cluster seen by the POM will then contain a number of existing archetypes and options for converting these into different archetypes, with implications for the resulting demand profiles (e.g. if a building with a gas boiler is then converted into one with a heat pump).

However, simply adding the *base* demand profiles for the same energy vector will overestimate peak demand in the cluster and as a consequence the necessary network reinforcement (or new build) costs, which is particularly problematic in the case of electricity.

In reality there are natural diversity effects at a local level, for example, in terms of electricity demand on a street level feeder, which affect the overall level of peak demand seen on the feeder. Adjustments need to be made to the base archetype profiles in POM to account for

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   197/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

diversity, however, it is important to separate out a number of different demand profile elements as they will be subject to different diversity adjustments which can be applied by the user:

1. Heat-related electricity consumption as variation is driven strongly by external temperature requirements and hence there is potentially more coincidence of demand on cold days

2. EV related consumption, as the extent of ownership and grouping of ownership in a cluster is a key exogenous assumption, and it will be important to easily adapt the diversity factor for this

3. Other electricity consumption covering appliances, lighting, etc which will again have some level of diversity, and is most closely related to historically observed diversity factors

There are two options for implementing these diversity effects in the current version of EPN

▶ The simpler (**Option a** in Figure 8-4) option is to pre-process the base profiles using a 'stretch factor' that relates to the number of archetypes in the cluster – i.e. a cluster with 100 buildings would see more smoothing *per building* than a cluster with fewer buildings. Separate factors would be used for each of the elements above and the diversity scaling factors themselves can vary by archetype and by timeslice (e.g. less diversity benefit on a very cold peak winter day compared to a characteristic winter).

  – The downside to this option is that for the heat-related electricity component it makes an ex-ante assumption about the number of buildings that will be converting to (or from) electrical heating over time, whereas this is a dynamic choice in the POM. This may overestimate the diversity benefit for heat-related electricity consumption, but not for the other element as the assumptions are known ex-ante.

▶ A more accurate approach (**Option b**) is to link dynamically the diversity adjustment for the number of archetypes using electricity as a primary heating device. The *incremental* effect on peak demand in the cluster per additional building using electricity for heating would decline reflecting the diversity effect more accurately.

  – This option relies on a MIP formulation of the optimisation problem (to create a piecewise linear curve), and creates a more complex optimisation problem, which by extension is significantly slower to solve.

Figure 8-4   Options for capturing cluster-level diversity effects
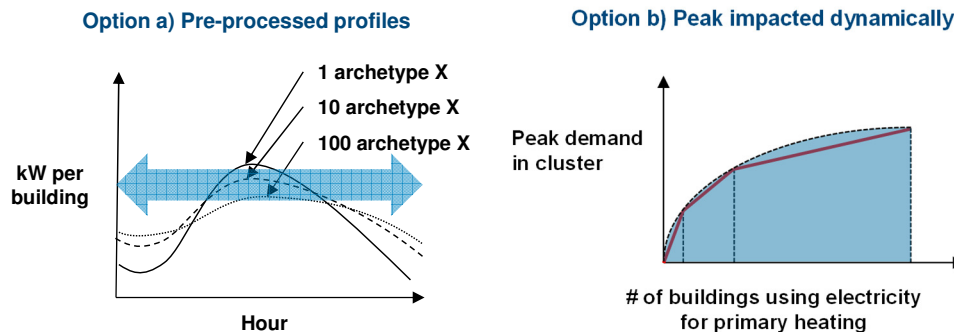

8.6.1.2.   *Problem simplification*

The ability to easily flex the complexity of the POM optimisation, to help improve performance, is a key design feature of the EnergyPath™ Networks Tool.   There are number of possible routes to simplification both within the POM and as part of the data processing in other modules which feeds into the POM. The range of routes is important as it is difficult to tell ex-ante which areas of detail are most important for understanding the pathway design (and may indeed vary by area):

▶ As a general point the POM aims to retain parallel LP and MIP structures as an LP problem is considerably quicker to solve than a MIP.

    – It is also possible to target the relaxation of integer variables which vary by timeperiod, for example keeping 'lumpy' MIP investment options in the near term when these cost are more important, but relaxing this towards the end of the pathway given the increasing levels of uncertainty about decision making at this point.

▶ POM pre-processing simplification

    – *Timeslicing* - building archetype energy demand profiles sent from the HOM are initially timesliced at their most granular level by typical characteristic day (e.g. average season, peak winter) and half-hourly within day. It will be possible to flexibly aggregate both the number of characteristic days (e.g. months to seasons) and / or the number of time periods within day (e.g. see Figure 8-5 below)

    – *Timeperiods* – in a similar manner to timeslices the POM receives timeperiod data at its most granular annual level initially (e.g. the cost of archetype conversion or network options). It will then be possible to aggregate this data to a smaller number of pathway steps e.g. 5/10-yearly periods or annually in the near-term with 5-year steps thereafter

▶ SAM / NAM pre-processing simplification – reducing the number of clusters

    – The number of clusters used to represent a local area is flexible and is defined at the SAM/NAM stage before being passed to the POM

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   199/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

▶ HOM / NAM pre-processing simplification – reducing the number of POM options

    – As described in section 5.5.3 part of the HOM process involves simplifying the number of archetype pathways used to represent a local area (both existing and the options for conversion). This process is flexible and hence further reducing the number of archetypes would by extension simplify the optimisation problem in the POM. Similarly, the number of non-domestic building pathways (which are specified directly by the user) in the SAM could be reduced. Applying building or cluster level retrofitting constraints to domestic archetypes would be another way to reduce the complexity in the POM.

    – In a similar manner the NAM would be able to aggregate similar network reinforcement options within a cluster where they are broadly similar in terms of cost and headroom provided. Also, network capacity can be restricted to a user-defined range (with min & max) so as to further simplify the problem.

Figure 8-5   **Example of POM timeslice pre-processing simplification**
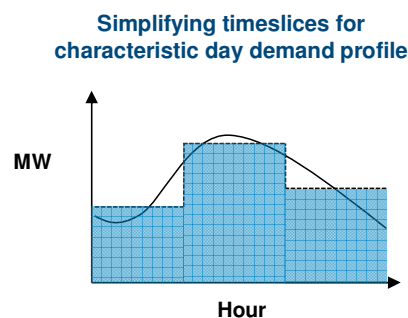
8.6.1.3. *National level boundary conditions*

To assess the optimal pathway design for the local area, over the pathway to 2050, it is important that relevant national level pathway conditions are consistent with those assumed in the POM.

A fundamental assumption is that the national level conditions are unaffected by decisions made by an individual local area, analogous to a "price-taker" rather than a "price-maker". Over time as the EnergyPath™ Networks Tool is used with multiple local authorities, insights from more accurate pathway design at a local could be used to refine the national level pathway (e.g. the costs and feasibility of district heating). An ongoing process would then be established such that insights from both the national and local area pathways are used to refine each other.

In the current version of EPN the UK ESME model provides this national-level information. ESME can undertake 5-yearly pathway optimisation steps from a 2010 base year to 2050 and is geographically disaggregated to the 12 political regions of the UK. Its within year timeslicing is also currently limited to 3 characteristic days (peak winter, typical winter and summer) and 5 within-day periods. For some boundary condition data some additional interpretation may be needed to bridge the gap between the resolution at the national pathway level and that needed by the POM within the EnergyPath™ Networks Tool - e.g. 5-yearly to annual timeperiods (e.g. via

ETI – SSH - EnergyPath™ Networks Tool – Functional Specification   200/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

linear interpolation) or moving from a regional spatial level (such as Scotland or London) to a specific Local Authority area within the region.

A range of national boundary conditions can be defined to inform the local area pathway design, from those which still provide all the original degrees of freedom in the POM, to those which tightly constrain certain aspects of the pathway. For example:

▶ The most flexible boundary conditions reflect price impacts, particularly the overarching fossil fuel prices consistent with the national pathway solution, as well as the shadow prices for carbon and other energy supply at the physical boundary of the local area (which are an output of the pathway) such as hydrogen and biomass[52]. For electricity the price is based on the LRMC (Long Run Marginal Cost) of generation in each ESME timeslice. The prices embed the cost of the carbon given the carbon intensity of production and shadow price of carbon from the ESME model.

  – The price of electricity at the boundary also provides a proxy for the value of export of surplus embedded generation within the local area energy system back onto the national system

▶ Similarly, the cost of specific technology options at the national level is mirrored to some extent at the national level. For example, the potential reduction in Solar PV or ASHP costs

  – However, a similar interpretation is needed to bridge the gap between the simpler technology representations at the national level, such as the cost of an ASHP, versus the wide variety of costs for different ASHP sub-types applied to different building archetypes in EPN. In general, the trend in building specific costs at the national level over time is used, rather than the absolute level, for the boundary condition,

▶ The most stringent boundary conditions involve forcing the deployment of particular options within the local area in line with the national pathway. For example, if 50% of all households in the national region have ASHPs this forced deployment could be imposed.

There are a wide range of national pathway conditions that could be imposed at the local level, particularly at the more stringent level. The overarching design of key constraints in the POM (see section 8.5.3) is such that the user can easily implement different variations of these depending on the operation of the tool. More stringent boundary conditions should be tested as part of wider sensitivity analysis, and changing national boundary conditions (such as the level of transmission connected electricity decarbonisation) should form a key part of resilience testing of the local area pathway.

V2.1 of the EPN tool has been designed to automatically extract the required boundary condition inputs from v3.4 of the UK ESME databased. If the underlying database structure of the UK

---

[52] I.e. it cannot automatically be assumed that biomass is used in the local area it is produced, as the value of this limited resource must be considered within the context of national decarbonisation options such as CCS

ESME model changes in future releases it may then be necessary to update the EPN tool boundary processing.

8.6.1.4. *Additional user-defined inputs*

Here the user can define all other scenario inputs which have not already been entered or have come from other modules such as design standards (e.g. 'peak winter' reserve capacity margins) or other constraints, such as restrictions on the deployment rate of solar PV.

8.6.2. Pathway Optimisation process [POM-001]

At this stage the final optimisation process is effectively automated and would undertake a repeated set of optimisation for all *sets* of input data generated by the Monte Carlo simulation process.

8.6.3. Export POM results [POM-002] and ArcGIS spatial results [SAM-009]

Following successful completion of the POM optimisation there exists automated processes to:

▶ Export a set of pre-defined POM results into the master EPN model database so that the user can interrogate these.

▶ Create results layers to view in ArcGIS where these are differentiated by cluster. The user can configure which sets of results are passed to ArcGIS (see section 6.5.10 for further details)

8.7. Output validation

To make the POM optimisation problem tractable, detailed analysis of the existing and future energy system is first undertaken in the HOM / NAM / SAM, but this representation is then simplified for the POM primarily through spatial and temporal aggregation.

Whilst the design of the EnergyPath^{TM} Networks Tool allows for the level of 'detail' seen by the POM is easily flexed, it is important to be able to undertake validation of the POM outputs via additional routes such as comparison with separate, area-specific engineering studies.

In addition, it would in theory be possible to take the more aggregated cluster level results from the POM pathway and test their 'feasibility' and 'appropriateness' via the more detailed modelling process used in the NAM and the HOM. Where there is a divergence between the two sets of results, the inputs sent to the POM could be refined to better calibrate a more suitable outcome from the POM itself, or alternatively additional constraints could be added to the POM to drive the more appropriate outcome[53].

---

[53] In a similar manner to the way the UK ESME model contains a number of more detailed structures for analysing building heating and peak supply / demand balancing. These structures are 'turned off' by default to enable faster solving time, but the insights from solutions using the more detailed structures are calibrated into the default setup via custom constraints.

It should be noted that the ability to undertake this 'feedback' analysis is not automatically configurable in v2.1 of EPN, however, for subsequent version of EPN it could be implemented as follows in the:

▶ **NAM**: the network solutions chosen for electricity and heat at a cluster and inter-cluster level from the POM could be tested within the more detailed spatial load flow models to:

- Ensure the solution is feasible - i.e. the demand conditions do not violate the network constraints[54]. If this does occur it may imply an underestimate in reinforcement in the POM and this could be remedied (for example) by adjusting the level of peak demand that triggers reinforcement in the final network options that are sent to the POM

- Re-examine and refine the more detailed network deployment options, which would be chosen at the NAM-level as part of a least cost-solution given the boundary conditions of increasing load over time from the POM solution. Where this implied a significantly different set of network deployment choices than that seen in the POM; the options, or costs of the options, passed to the POM could be refined.

- Finally if the detailed network analysis in the NAM shows the solution is feasible and the wider POM analysis does not need to be updated, the information used for the individual project briefs for network options could draw more directly on the more detailed cost and other information from the NAM

▶ **HOM**: in a similar manner, the POM applies demand profile as calculated by EnergyPlus for each characteristic day that meets a desired target temperature. Where there is optionality in heating system operation (e.g. with storage) EnergyPlus aims ex-ante to minimise the peak demand requirement for the building to both minimise the size of the heating system and likely peak load requirements on the wider energy networks. However, the conditions in the POM may suggest that this is not necessarily the optimum strategy, for example, given different electricity/gas prices for hybrid systems or the shape of aggregate load on the local area system, which may suggest shifting load to a pre-defined period rather than trying to minimise the peak seen across all days.

- By passing these key drivers back into the dynamic energy simulation component of the HOM it would be possible to understand whether different operational strategies should be considered and whether, additional profiles could be created in the HOM and passed to the POM for a further iteration of the pathway optimisation, as this may indirectly wider pathway choices such as the level of peak demand and network reinforcement.

---

[54] In a similar manner to the way the ESME solution for the electricity system is tested in a PLEXOS hourly resolution dispatch model to better understand whether the system operation is feasible from a security of supply and flexibility perspective.

9. Modelling approach to uncertainty

9.1. Overview

A key design requirement for the EnergyPath™ Networks Tool is to allow the user to efficiently tackle 'uncertainty' associated with the inputs and interpretation of the outputs. This 'uncertainty' can be grouped into two broad categories:

▶ The more technical elements of tackling uncertainty within the tool itself *which is the purpose of this section*; e.g. what is the modelling approach to dealing with uncertainty in inputs and associated design issues, what are the most material input parameters which exhibit uncertainty, what other elements of the modelling approach introduce uncertainty and how, etc

▶ The overarching use of the tool itself (and the features within it) as part of a business process that aids real-world decision making, in light of the inherent uncertainty of the problem the tool is trying to tackle – i.e. a strategic pathway design for the next 30+ years. *This is discussed further in section 4*

The main technical approaches to tackling uncertainty in the tool will be two-fold

▶ Scenario analysis (i.e. manually changing one or more discrete input parameters) to enable what-if testing or resilience analysis, such as target internal temperature levels for domestic buildings. The benefit of this approach is that it allows the user to easily understand the directional impact of changing parameters (within the context of other fixed inputs)

▶ Monte Carlo simulation of key input parameters. The benefit of this approach is that it enables the user to more efficiently assess a wider range of parameter uncertainty on pathway design; both in terms of the number of states for an individual parameter (e.g. 100 of samples from a triangular distribution of fuel prices) and across multiple, and potentially correlated parameters (e.g. 100 of simulated sets of coal, gas, oil and biomass prices)

The types of uncertainty within the tool can be categorised broadly as follows:

▶ Data driven uncertainty:

  – Uncertainty in input data that can only sensibly be parameterised in a discrete sense (e.g. low, medium, high values) and which lends itself more to specific scenario values or discrete sensitivity testing. Examples might include the level of industrial energy demand that could exist in LA area over a 30 year pathway, which is highly uncertain, but has significant implications for wider energy system design

  – Uncertainty in input data which can sensibly be parameterised as a continuous variable (e.g. normally distributed with a mean of X and standard deviation of Y)

and which lends itself more to Monte Carlo simulation. Examples may include the distribution of technology costs[55]

▶ Modelling driven uncertainty (i.e. that driven by the approach or use of the data as opposed to uncertainty in the data itself).

– Examples of this include anywhere the more detailed representation of the LA is simplified to make the optimisation problem in the POM tractable. For example, the need to aggregate the spatial information from more detailed zone- to cluster-level. This means that spatial information on precisely where options are deployed within the cluster is lost for the purposes of decision making.

– Mitigation strategies to overcome this modelling introduced uncertainty are important and in this example are two-fold; the first is to use assumptions which minimise any distortion in decision making as this spatial information is lost[56], the second is to have the ability to easily flex the level of spatial granularity and re-run the POM to understand how this impacts the solution.

9.2. Summary of module-specific uncertainty

The tables below summarise the uncertainty introduced by both data and modelling approach in the different parts of the EnergyPath™ Networks Tool and how is treated and mitigated, respectively.

---

[55] Note that when information on the distribution of a parameter is limited, a default approach is to parameterise the min, mode, max of a triangular distribution.   However, for some parameters, such as the industry example above, it may still be better to understand the impact of changing this factor via discrete scenario testing rather than including this within a wider set of simulated inputs; as it then becomes difficult to unpick what is the key driver of the results.

[56] For example, as outlined in section **Error! Reference source not found.** the approach for aggregating load in a cluster for the purposes of assessing network reinforcements

Table 9-1   Key data driven uncertainty

Module	Uncertain parameters	Propagation across tool	Treatment
HOM	Costs for converting one archetype into another driven by uncertainty in component costs for insulation, heating systems, etc	• Output passed to POM directly for use in pathway optimisation	• **[Monte Carlo]** The user can define a distribution of input costs for heating systems and efficiency measures, which is then be used as part of Monte Carlo Sampling – The central pathway costs generated by the HOM-004 sub-module are combined with the sampled data to create a series of sampled pathway costs within POM-000

HOM	Archetype final energy demands and profiles. This is driven by underlying uncertainty around a number of components:
	
	• Heat related
	  – External temperature profiles and comfort requirements
	  – Building heating system performance
	  – Building shell performance (e.g. U-values of insulation measures)
	• Other lighting / appliances
	  – Driven by usage patterns
		• Output passed to SAM (used indirectly as part of calibration of archetype matching using historic demand by area see 6.5.1)
		• Output passed to NAM to help determine likely maximum demand range to be tested by network modelling
		• Output passed to POM for direct use in pathway optimisation
		• Output passed to POM to help define cluster diversity factors
			The treatment of uncertainty is different for each underlying component used to create the final demand profile for each archetype
			
			• Heat related
			  – **[Scenario]** Uncertainty in external temperatures and comfort requirements/behaviour is captured by the creation of typical day and edge-case or peak days and so these are treated deterministically in the POM, with no additional Monte Carlo simulation (however, the number of characteristic days, definition of edge cases, is flexible)
			  – **[Monte Carlo]** the user can define a distribution of domestic heating system efficiencies to sample from. As an approximation the energy demand profiles created by the HOM-003 submodule are scaled linearly (in POM-000) based on the relative change in efficiency of the central versus simulated value, to avoid the time-consuming step of running EnergyPlus for each set of simulated efficiencies
			  – *[Monte Carlo] the user can simulate a distribution of u-values for insulation measures that are used to define the Base Archetypes (CWI, LI, SWI glazing. These feed into the HOM-001 sub-module to generate X simulated sets of SAP model inputs and subsequently propagate through the subsequent Base archetype simplification process in HOM-002 (see section 5.5.3) as they change the imputed HLRs of the based archetypes, and ultimately create updated sets of demand profiles that feed into the POM (with*

			potential further adjustments due to the simulated heating system efficiencies as mentioned above) • **[Scenario]** Uncertainty in other lighting / appliances is captured via changing exogenous assumptions

SAM	• EV demand and profiles driven by
	   – Number of EVs
	   – Charging profile
		• Output passed to NAM to help determine likely maximum demand range to be tested by network modelling
		• Output passed to POM for direct use in pathway optimisation
			• **[Scenario]** For EVs uncertainty in charging profiles would be reflected by exogenous assumptions, consistent with the National Pathway data

SAM	• Non-domestic energy benchmarks by activity class	• Output passed to POM for direct use in pathway optimisation	• *[Monte Carlo] the user can specify an uncertainty distribution around one or more of the benchmarks. The simulated sets of benchmarks then ultimately propagate through into the POM in the form of multiple sets of demand profiles for each non-domestic pathway option.*
SAM	• Non-domestic retrofitting costs	• Output passed to POM for direct use in pathway optimisation	• *[Monte Carlo] the user can specify an uncertainty distribution around one or more of the retrofitting options. The simulated sets of retrofitting costs then ultimately propagate through into the POM in the form of multiple sets of retrofitting costs for each non-domestic pathway option.*

NAM	Costs context of network reinforcement or new build options. As discussed in section 7.3.4 the different parameters that lead to the uncertainty in cost will be treated differently, but can be grouped broadly as:
	• Base component costs of equipment and installation or economies of scale with increasing size of reinforcement
	• General site (urban, rural, etc)
	• Specific ground conditions (e.g. excavation difficulty)
		• Output passed to POM directly for use in pathway optimisation	• **[Monte Carlo]** The user can specify a distribution of upgrade costs for each network option based on the base component costs, which could be used by the POM for Monte Carlo sampling – e.g. within a triangular distribution.
			   – The POM-000 sub-module takes the simulated parameters and scales the cost curve outputs of the NAM-007 sub-module, based on the relative change in costs of the central versus simulated value, to avoid re-running the earlier time consuming steps in the NAM
			   – On top of the simulated costs, location specific scalars could be used to represent the further mark-up/down of costs related to general site context and specific ground conditions
			   – Base component costs across similar network options would have to be correlated

POM	Cluster diversity demand scalars separated by • Heat-related electricity demands • EV electricity demand • Other electricity demand • Heat network demand	• N/a	• *[Scenario]* These are explored by discrete scenario testing, however, the option to undertake Monte Carlo simulation alongside other outputs could still be included
POM	Other non-building / non-network energy system data for new options – e.g. CHP, other district heat sources, network storage or other network embedded generation. The uncertainty is most likely to cover cost components (CAPEX, FOM, VOM), efficiency and availability or load factor		• *[Monte Carlo]* POM contains cost data necessary to define distribution for Monte Carlo sampling
POM	National pathway boundary conditions. This can be divided into • Carbon shadow price • Variable costs, availability and carbon intensity of transmission connected electricity and heat at boundary of local area • Cost of increasing available peak supply at boundary • Fossil fuel and biomass prices • EV ownership • Potentially also relative cost trends by timeperiod for key technologies (e.g. Solar PV)	• Taken from ESME outputs and used directly in POM	• *[Scenario]* ESME outputs are themselves a set of simulation results reflecting uncertainty. The averages of the simulated inputs/outputs are used as inputs to EPN
POM	Other user defined scenario assumptions – e.g. location and level of industry in each time period	• Used in POM pathway optimisation	• *[Scenario]* These are explored by discrete scenario testing.

Table 9-2   Key approach driven uncertainty

Module	Uncertain parameters	Propagation across tool	Mitigation strategy
HOM	Number of building archetypes used to describe area after filtering/simplification process affects the accuracy of the representation	• Final 'menu' of archetypes passed to SAM for archetype matching on a spatial basis • POM options available for upgrading buildings by cluster	Archetype filtering process is designed to be flexible such that the user can relatively easily tailor the number of archetypes. This more granular information is then automatically propagated through to the rest of the EnergyPath™ Networks Tool via the standard mechanisms.
SAM	Synthetic network generation (see 6.5.5) where real data is limited initially.	• Network topology data is passed to the NAM and will impact load flow modelling and the generation of reinforcement options used in the POM	The base process for generating artificial network topology is flexible in that it can accommodate actual connections data, where available, or can be calibrated to real data at a more aggregate level, which is likely to be available in most cases. E.g. total number of customers connected to a substation (even if it is not known exactly which final feeder each customer is connected to). The base process for generating an artificial topology was informed by our DNO partner WPD and is judged to be the most appropriate approach in the absence of real world information.
SAM	Building archetype matching (see section 6.5.1) – uncertainty in understanding what the best match is for each unique building in the area given limited data	• Archetype matching affects options seen by POM within each cluster as part of pathway optimisation	Two key mitigation options: • Process is flexible to accommodate better data as it becomes available at different levels of granularity and improving matching process • Validation exercise to compare estimated annual demand for electricity and gas from matching of the existing stock with historic spatial estimates. Key issue for calibration exercise is that this is compound problem of uncertainty in HOM energy demand data (as described in table Table 9-1) and archetype allocation data
NAM	Accuracy of the derived cost functions for zones, clusters and inter-clusters • This includes how accurately network reinforcement functions are collated; • It also includes how accurately additional load is assumed to be distributed within a cluster;	• Network cost functions used by POM within each cluster as part of pathway optimisation	As described in section 7.5.4.2 there exists a spectrum of possible options for representing the accuracy of the network cost functions, we have implemented the penultimate option 4b. In addition the key mitigation option described below also applies in the case of the accuracy of the cost functions.
NAM / SAM	Cluster definition, both number and shape see sections 6.5.8 and 6.5.9), the driving principles of which are to:	• Central to defining the network cost functions which are used in POM	The core cluster generation process aims to balance the representation of heat and electricity networks in particular. However, the process of defining the clusters is flexible such that

	• Create the maximum number of clusters which is computational feasible in the POM to improve granularity • Create clusters which contain network reinforcement / new build options which are as representative as possible for all energy vectors so as not to unduly impact the cost-optimisation process in the POM			the user can both automatically generate different cluster definitions (e.g. subject to different threshold and network constraints) as well as manually impose cluster definitions, and the final cluster data will automatically be processed in the required format for the POM
POM	Data simplification steps (see section 8.6.1.2) to minimize the computational complexity of the POM optimisation • Timeslice aggregation • Timeperiod aggregation • LP / MIP structure (a more tailored linearization of integer variables in some timeperiods only – i.e. later ones – is also possible as an intermediate option)		• Central to the POM's ability to accurately trade-off different energy system design options	It is difficult to say ex-ante where granularity matters most for the defining the energy system pathways and may even be different in each local area. Therefore the process of data simplification is designed to be flexible, such that the user can easily change granularity in one area and automatically reprocess the data in the necessary format for the POM optimisation

In the case of data driven uncertainty, the vast majority of this is propagated through the tool in what is effectively a 1-step process to the POM. A range of uncertainty on the inputs in preceding modules (such as the costs for pathway archetypes in the HOM or network options in the SAM) is either passed directly or captured in the intermediate outputs and passed to POM for use in Monte Carlo simulation, i.e. running the POM optimization X times depending on how many X samples' worth of Monte Carlo inputs have been generated. By default the POM would use the central/mean values in the deterministic case where the POM is run once.

There are two more complicated areas of uncertainty propagation across the EnergyPath[TM] Networks Tool that also need to be considered:

▶ Where ESME is used to generate boundary conditions for the POM it is possible to undertake simulations reflecting national level uncertainties in ESME first. The current version of EPN takes the average of any ESME simulated inputs/outputs before passing it into EPN (and prior to any Monte Carlo simulation in EPN itself). However, if the ESME data was passed to EPN on a simulation-by-simulation basis, to coincide with the simulations in EPN, then greater care would need to be taken to ensure that data is consistent in each simulation in each model.

  – For example, if heat pump costs were being simulated in both ESME and EPN, it would be inconsistent if ESME's sim 1 costs were higher than the central case, whilst EPN's sim 1 costs were lower. At present there is no automated functionality to manage this simulation-by-simulation consistency.

▶ The 2-stage uncertainty propagation from potential Monte Carlo simulation of some elements in the HOM that drive the final energy demand profiles seen by the POM for the domestic archetype pathways. The current design principle for EPN is that where the HOM-related uncertainty is driven primarily by behaviour (e.g. target comfort requirements) this is explored by different deterministic scenarios, whereas technology-related uncertainty (e.g. performance of heating systems or insulation) can potentially be explored via Monte Carlo simulation of parameters that ultimately lead to multiple sets of simulated demand profiles in the POM.

9.3. Monte Carlo input generation sub-modules

As outlined in section 8 the approach to Monte Carlo simulation in the POM has significant user flexibility with respect to how it is configured in term of:

▶ Which parameters to simulate - such as technology costs, technology performance, technology deployment constraints, resource availability, etc (see Table 9-1). Any non-simulated parameters use their default deterministic values in the optimisation

▶ The distributions used[57] – such as triangular (generally used by default when there is limited knowledge of the uncertainty surrounding a parameter as it is easier conceptually to define minimum / mode / maximum values)

---

[57] Normal, triangular, two-point discrete, uniform

▶ The parameters used to define the uncertainty distributions in each time period, as in many cases the spread of uncertainty increases to a zenith over the medium term (e.g. the next 10-20 years), but beyond this does not increase further. In the EPN Monte Carlo process only the end point (e.g. the 2050 value) is technically simulated and the deterministic profile of costs over the pathway is adjusted up or down accordingly. However, unlike the ESME simulation it is possible for the user to adjust the shape of this uncertainty spread over the pathway as shown below[58].

Figure 9-1   Example of how Monte Carlo simulated values are applied
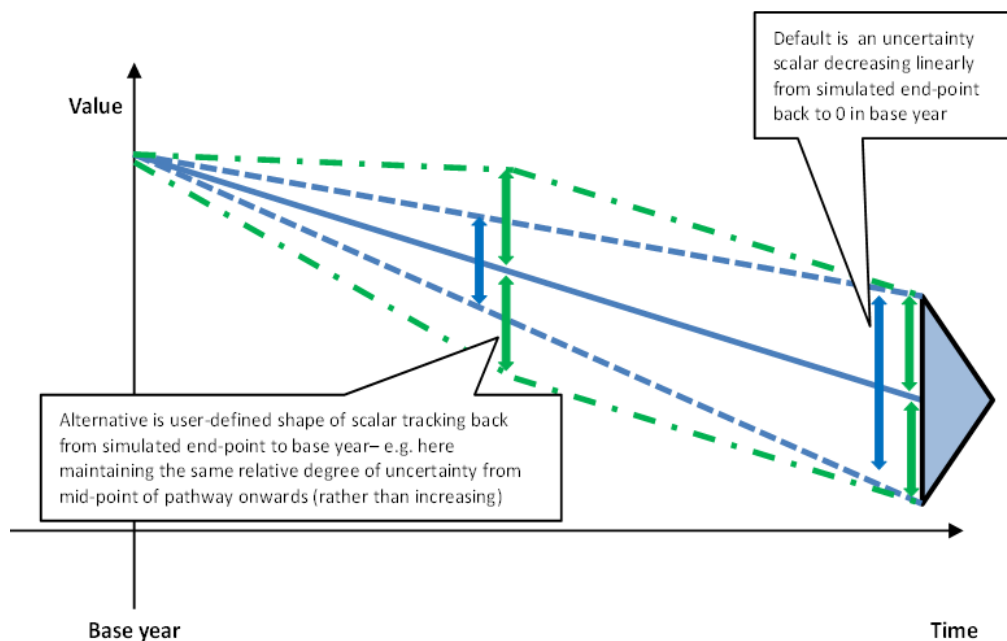
An additional complexity involved with the simulation of input parameters is accounting for potential used-defined correlation in parameters. For example, the correlation between similar technology options to ensure that the costs for ASHP variants which share similar core components, but which are associated with different building archetypes, move in a similar direction between simulations.

This requires the application of complex correlation matrices (more so than the approach currently used in ESME). As a result the current EPN design does not allow flexibility in the simulation of different uncertainty parameters for each cluster in the local area (although this could be included in future). This is primarily because there is limited rationale for a key input (e.g. the cost of a technology or supply chain constraints) to have a radically different uncertainty distribution by virtue of being a small geographic distance away; but it avoids further complexity in the specification of correlation factors.

---

[58] This is in contrast to the approach taken in ESME whereby only the data parameters in 2050 are simulated and the deterministic profile of data values from 2010 to 2050 are scaled up or down to meet the simulated 2050 value. I.e. uncertainty always becomes greater over time and it is not possible to simulate and increasing range of uncertainty to e.g. 2020/2030 and a fixed range of uncertainty beyond this point.

9.4. Multi-scenario data management

As part of facilitating the exploration of uncertainty through scenario analysis (as opposed to those parameters which can be considered within the Monte Carlo sampling process) the EPN tool and database structure has ability to manage multiple scenarios simultaneously for some aspects of the tool, which are time consuming to run. This allows the user to more easily switch between multiple sets of input data, and save relevant intermediate data rather than re-producing them each time or switching between multiple databases for different scenarios. Other processing steps are still calculated 'on-demand'.

The multi-scenario functionality applies to:

▶ **Cluster definition** – allowing the user to define multiple sets of clusters via ArcGIS and also save the resulting NAM cost curves associated with different cluster definitions

▶ **Domestic building archetype pathways** for heat and insulation – allowing the user to set up multiple sets of constraints (forcing or restricting certain options)

– As these pathways are linked to the cluster definitions they are also defined as a function of the clusters above

▶ **Domestic demand profiles** – allowing the user to define multiple sets of target temperature profiles and saving the resulting EnergyPlus demand profiles

▶ **Non-domestic building demand** – allowing the user to define multiple sets of energy benchmarks

10. Concluding remarks

The intent of this document has been to describe the required purpose, behaviour, data and components of the EnergyPath^{TM} Networks Tool v2.1 to a sufficient level of detail that facilitates modelling experts understanding and review of the current model implementation. The functional specification should be read in conjunction with the other key Stage 1 deliverables, in particular the Design and Data Architecture (deliverables D2 and D3, respectively) to understand the overall picture for the design and implementation of the tool, although it should be noted that only the Functional Specification and accompanying model user guide have been updated to v2.1.

11. Appendix – Full logic / process flow diagram

The diagram below shows the full logic / process flow diagram for the tool. This should be viewed in conjunction with the (D2) Design Architecture which highlights how these flows are organised from a technical design perspective – e.g. how databases are assigned and coordinated across the tool.

Figure 11-1　Full logic / process flow diagram

12. Appendix - Base archetype Simplification

12.1. Overview

Mathematical clustering has been used as a key means of reducing the total number of building base archetypes that are ultimately used for dynamic thermal simulation in EnergyPlus (n.b. this is distinct from the concept of spatial 'clustering' within the wider EnergyPath™ Tool).

This involves the translation of **base archetypes**, which are defined in terms of construction attributes (window type, wall types etc.) into a smaller number, which are defined by their physical performance attributes. These have been defined broadly as their heat loss and their thermal mass, although the physical size of the building will also be material.

The key questions explored in this testing are:

1. What type of mathematical clustering approach is the most suitable for application in EnergyPath™ (centroid, hierarchical, density), given the type of data inputs that are to form the basis of the analysis?
2. What settings give the best performance? I.e. what settings with which clustering approach arrive most rapidly at an optimum number of clusters with the best level of separation between cluster types?
3. Are the dimensional parameters outlined in section 5.5.1 suitable for differentiating between different archetypes, or do additional metrics need to be included in the clustering analysis process?

12.2. Test Data

The test data used is from the English Housing Survey (EHS). This comprises the full statistical release, processed in the ETI's ISO 13790 SAP/BREDEM model from the Optimising Thermal Efficiency of Existing Housing (OTEoEH) project.

Thermal mass and heat loss rate are calculated for each recorded dwelling in the EHS via use of the SAP model, giving the data spread shown below in Figure 12-1. The floor area of the dwelling i.e. it's size, is also a material characteristic, as shown in Figure 12-2. As can be observed in these figures, English Housing Survey data are for the most part, tightly grouped with no obvious separation between data points at the lower end of the scale range.

Figure 12-1   EHS Dwelling Data Expressed as a Function of Heat Loss, Thermal Mass
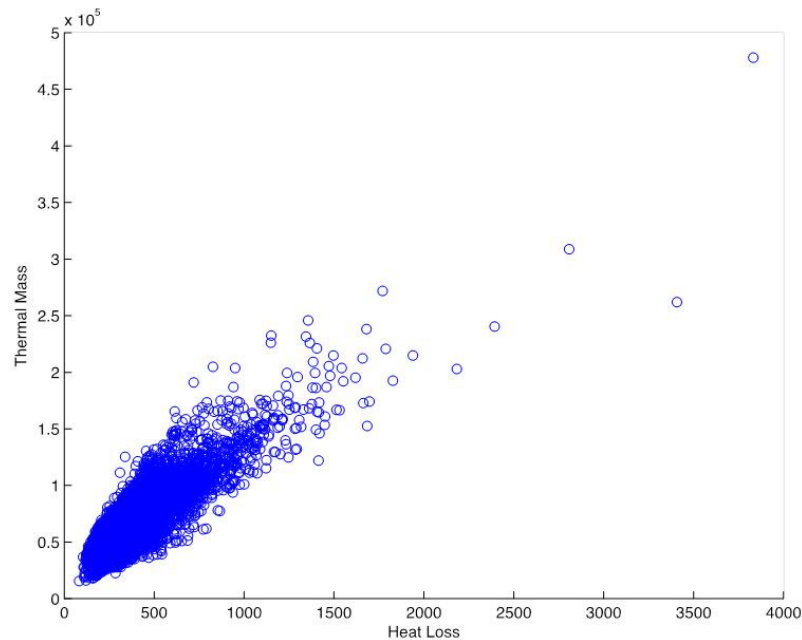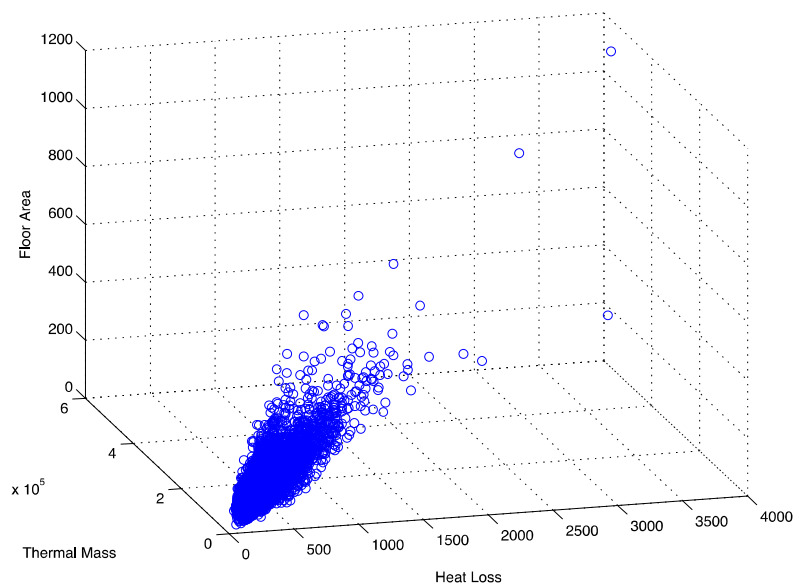

Figure 12-2   EHS Dwelling Data Expressed as Functions of Heat Loss, Thermal Mass, and Floor Area

It is expected that the data that will be mathematically clustered in EnergyPath™ will comprise the heat loss / thermal mass / size characteristics that emerge as the output of a master list of archetype combinations being processed through the ETI OTEoEH SAP model, repurposed for EnergyPath™. By the time this data reaches the clustering stage, the total combinatorial list of possible archetypes will have:

1. Had various logic rules applied to it to eliminate "impossible" combinations of archetype characteristics. For example, pre-20th century housing with cavity walls, multi-storey bungalows etc.

2. Been statistically filtered by a series of user control decisions relating to the statistical prevalence of different archetypes in the real-world building stock. For example, the user may be offered the option to eliminate combinations that make up "X%" of the stock, where X is a user input value. The use of this step will be limited to avoid excluding archetypes, which although statistically rare at an aggregate level, may be prevalent in a particular spatial area of a local authority.

12.3. Analytical Approach

As a first step, all input data (heat loss, thermal mass, floor area) are normalised to prevent numerical differences in units influencing the data clustering process. A number of clustering methods are then evaluated using the test dataset identified above:

▶ Hierarchical clustering – agglomerative linkage groups data points based on relative distance from one another in a tree structure (also known as a dendrogram)

▶ Centroid-based clustering – k-means groups all input data into a user-determined number of voroni cells, optimises the position of their centres, and groups all objects by their distances from the cluster centre

▶ Density-based clustering – groups objects in the data set by defining the minimum number of objects that are considered to be in a cluster and the search distance

In addition, and in the interest of exploring approaches for automated clustering (i.e. with minimal end-user input), a number of evaluation methods are tested. These generally assess the degree of separation between clusters and select the number of clusters with the highest separation as the 'optimal' solution. Methods tested include:

▶ Calinski Harabasz Index (1) – higher is better

▶ Davies Bouldin Index (2) – lower is better

▶ Average Silhouette Index (3) – higher is better

▶ Gap/Elbow Method (4) – compares variance between different cluster groups to establish the point at which adding additional clusters yields no new information, represented by a steep change in gradient

12.4. Hierarchical Clustering

12.4.1. Optimum Number of Clusters

Hierarchical clustering using an agglomerative linkage approach was trialled for different numbers of clusters up to 50. The upper limit of 50 using in testing is informed by the computational constraints of the optimisation approach used in the EnergyPath[TM] Pathway Optimisation Module (POM). For reasonable run-times this requires that the number of clustered archetypes does not significantly exceed a 10^3 order of magnitude (i.e. a few thousand or so).

Figure 12-3 illustrates how the different numbers of cluster trials performed under various evaluation criteria.

Figure 12-3   Evaluation of Agglomerative Linkage Clustering Under Four Methodologies
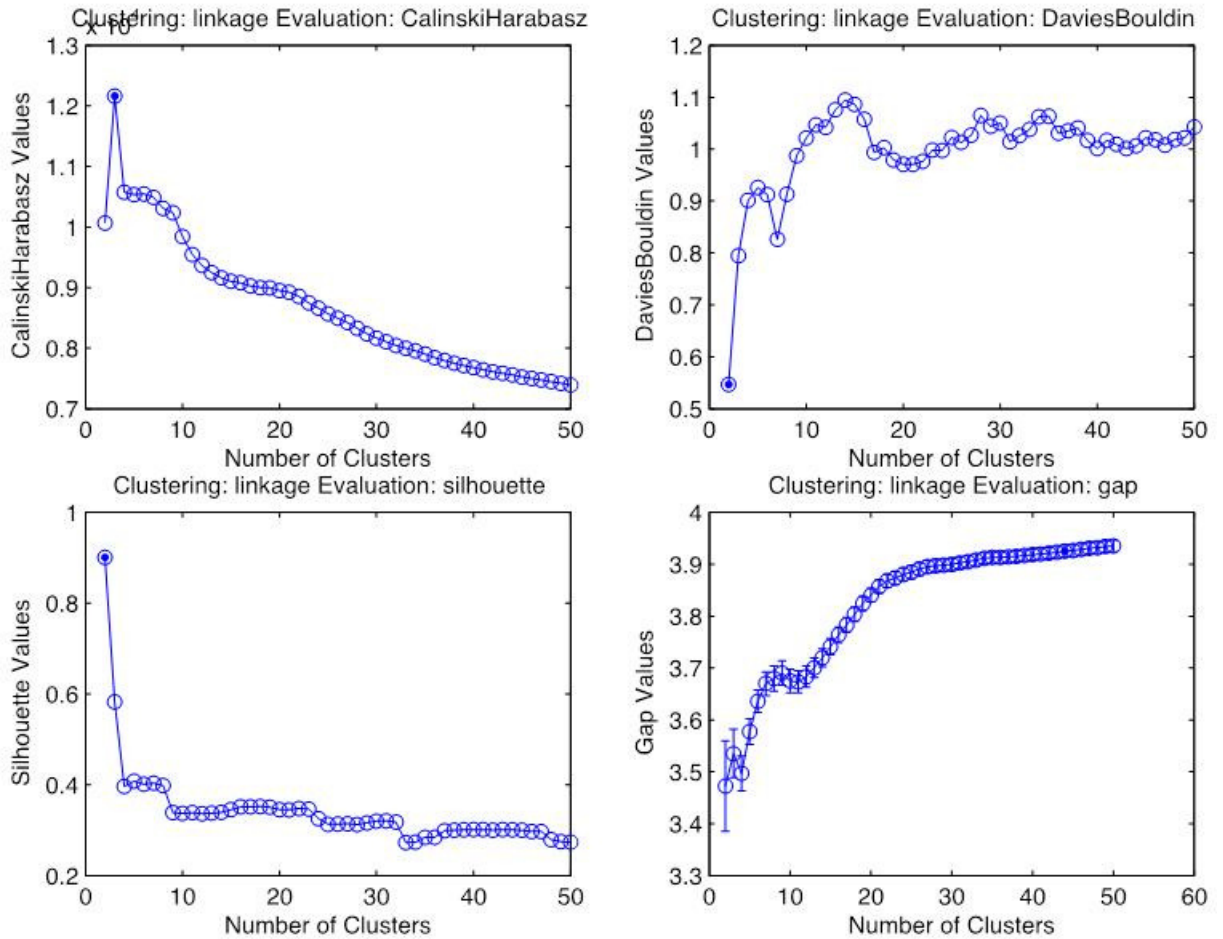

Figure 12-4 displays the distribution of the mathematically optimum number of clusters determined with each method. It can be immediately observed that there is variation in the optimum number of clusters determined using each method, with Calinski Harabasz recommending 3 clusters, Davies Bouldin and the silhouette average metrics recommending 2 clusters, and the gap method showing that beyond 44 clusters no signfiicant additional variance is explained between the variables.

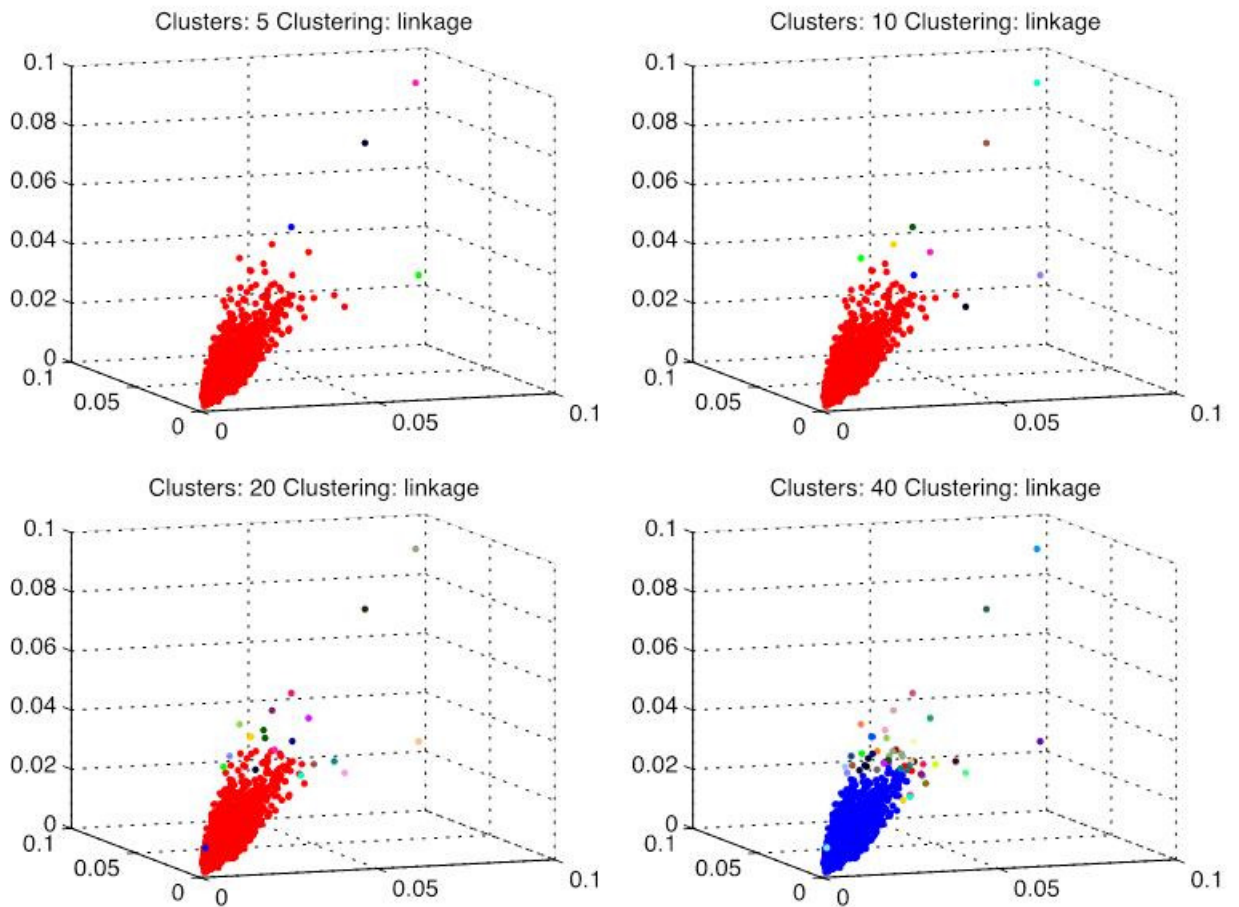**Figure 12-4   Optimised Agglomerative Linkage Clustering Under Four Different Evaluation Criteria**
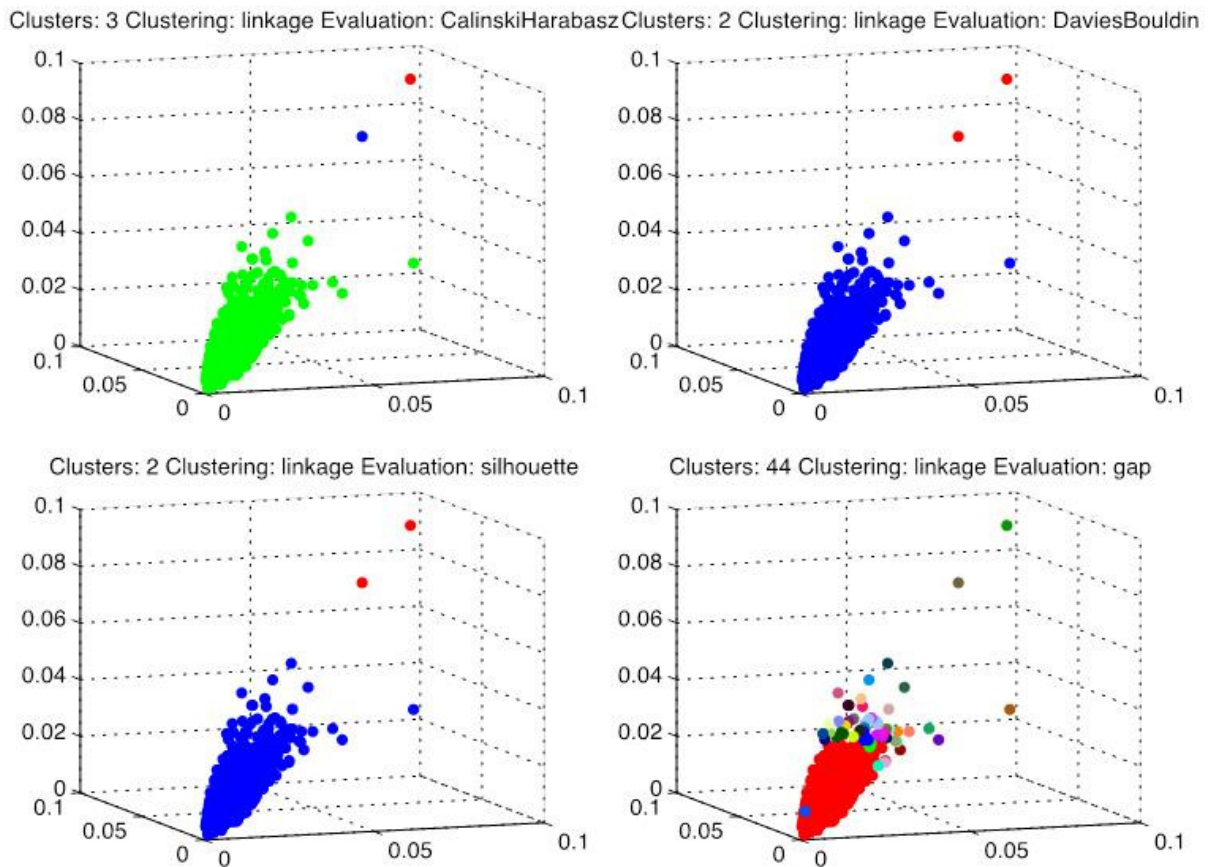
The difference in 'optimum' cluster numbers found using each approach can be explained by the different underlying evaluation methodologies. All are effectively measuring the separation of different cluster groups from one another but the separation is being calculated in different ways. The analysis here shows that there is no single 'best' cluster grouping. The 'optimum' number of clusters for the EnergyPath[TM] tool itself as a whole is also likely to be driven by criteria such as the total number of variables that can be employed in the Pathway Options Module (POM) optimiser and the materiality of the impact on the POM solution from having more or less granularity in the building representation. For this reason, manual clustering is explored below in Section 12.4.2. The defined number of clusters comes from a direct user input, i.e. trading off greater precision with more clusters versus model solving time.

12.4.2. Manually Defined Clusters

Figure 12-5 shows how the same dataset becomes partitioned in response to different target cluster values, using euclidean distances. Qualitatively, it can be seen that, regardless of the number of clusters used, the hierarchical method has difficulty distinguishing between tightly grouped data points. The structure of the parititioning strategy best explained as a single large "trunk" cluster close to the zero point of the thermal mass, heat loss and size axes, with the more remote points from the trunk becoming their own groups.

ETI – SSH - EnergyPath[TM] Networks Tool – Functional Specification   221/231

Baringa Partners LLP is a Limited Liability Partnership registered in England and Wales with registration number OC303471 and with registered offices at 3rd Floor, Dominican Court, 17 Hatfields, London SE1 8DJ UK.

12.5. Centroid-based Clustering

12.5.1. Optimum Number of Clusters

Centroid-based clustering using the k-means method was trialed for different numbers of clusters, up to 50 in total. The upper limit of 50 was chosen for the same reasons as described above in Section 12.4.1. Figure 12-6 shows the resulting plots from different trial methods.

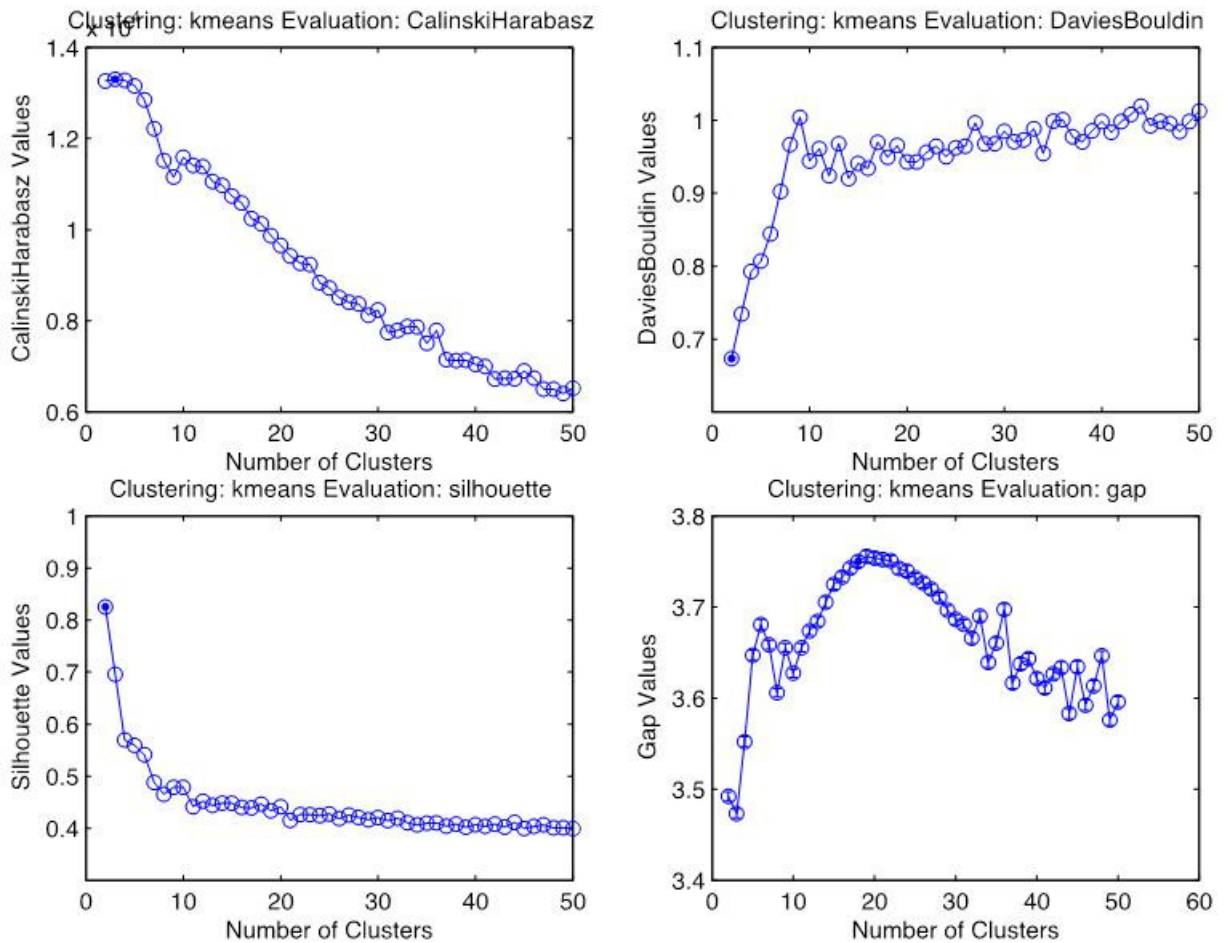**Figure 12-6   Evaluation of k-means Clustering Under Four Methodologies**

Figure 12-7 displays the distribution of the mathematically optimum number of clusters determined with each method. Under Calinski Harabasz, Davies Bouldin, and silhouette average methods, the k-means approach yields an identical number of clusters to the agglomerative linkage approach. However, by comparing the qualitative distribution of points between Figure 12-6 and Figure 12-7, one can see that the allocation of data points differs substantially between the two methods. This will be discussed in more detail later in the document. Gap evaluation also yields a different optima, with no significant variance found beyond a grouping using 18 clusters as opposed to 44 under this method.

Figure 12-7   Optimised k-means Clustering Under Four Different Evaluation Criteria
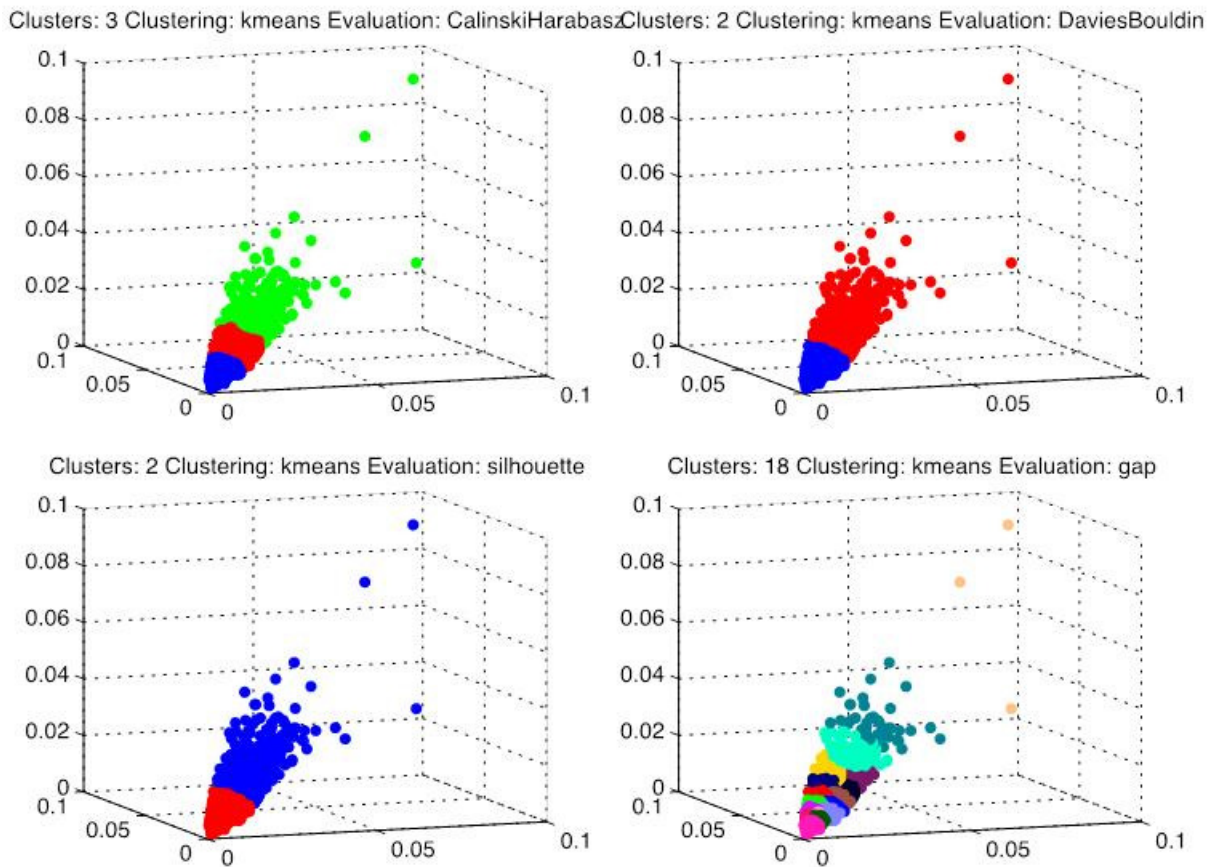

As discussed in Section 12.4.1, the variation between the 'optimum' number of clusters achieved when evaluating solutions with different methodologies is an indicator of the complexity of the problem. However, the optimum number of clusters is more likely to be driven by the design of EnergyPath™ as a whole, particularly the decision variable limits imposed by the Pathway Optimisation Module (POM), rather than mathematical metrics derived in isolation (such as these). For this reason, manual clustering, which could be determined by direct user input or computed by another part of the program, is explored below in Section 12.5.2.

12.5.2. Manually Defined Clusters

Figure 12-8 shows how the same dataset becomes partitioned in response to different target cluster values, using euclidean distances. Qualitatively, it can be seen that as the number of clusters is increased, the k-means method partitions the "trunk" area described for the hierarchical clustering approach into many more groups.

Figure 12-8   Variation in k-means Cluster Allocation Under Manual Control
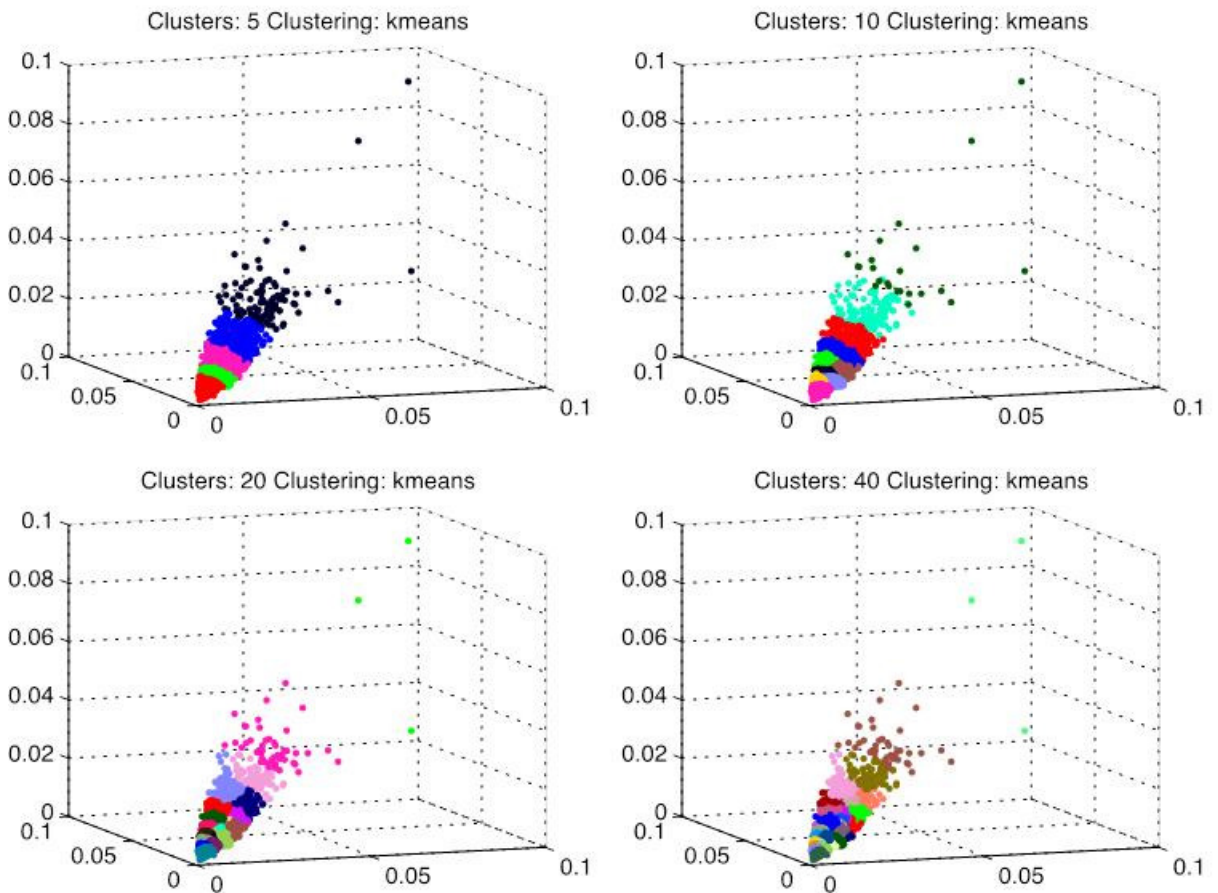
12.6. Density-based Clustering

Density based clustering methods such as DBSCAN, OPTICS, and their variants, allow some data points to be ignored entirely and considered as 'noise' in the data, as opposed to hierarchical and k-means methods, where every data point must be assigned to a group. Figure 12-9 and Figure 12-10 below illustrates how this concept can be applied to the EHS test data described above in Section 12.2, using the density-based spatial clustering of applications with noise (DBSCAN) algorithm (5). Here, the minimum number of data points required to form a cluster has been set to a value of 3, while the search radius is varied. From the graphical plots it can be seen that:

▶   As the search radius is reduced, an increasing proportion of the original data points are deemed to be "noise" and discarded from the analysis.

▶   At smaller search radii, the model defines an increased number of clusters

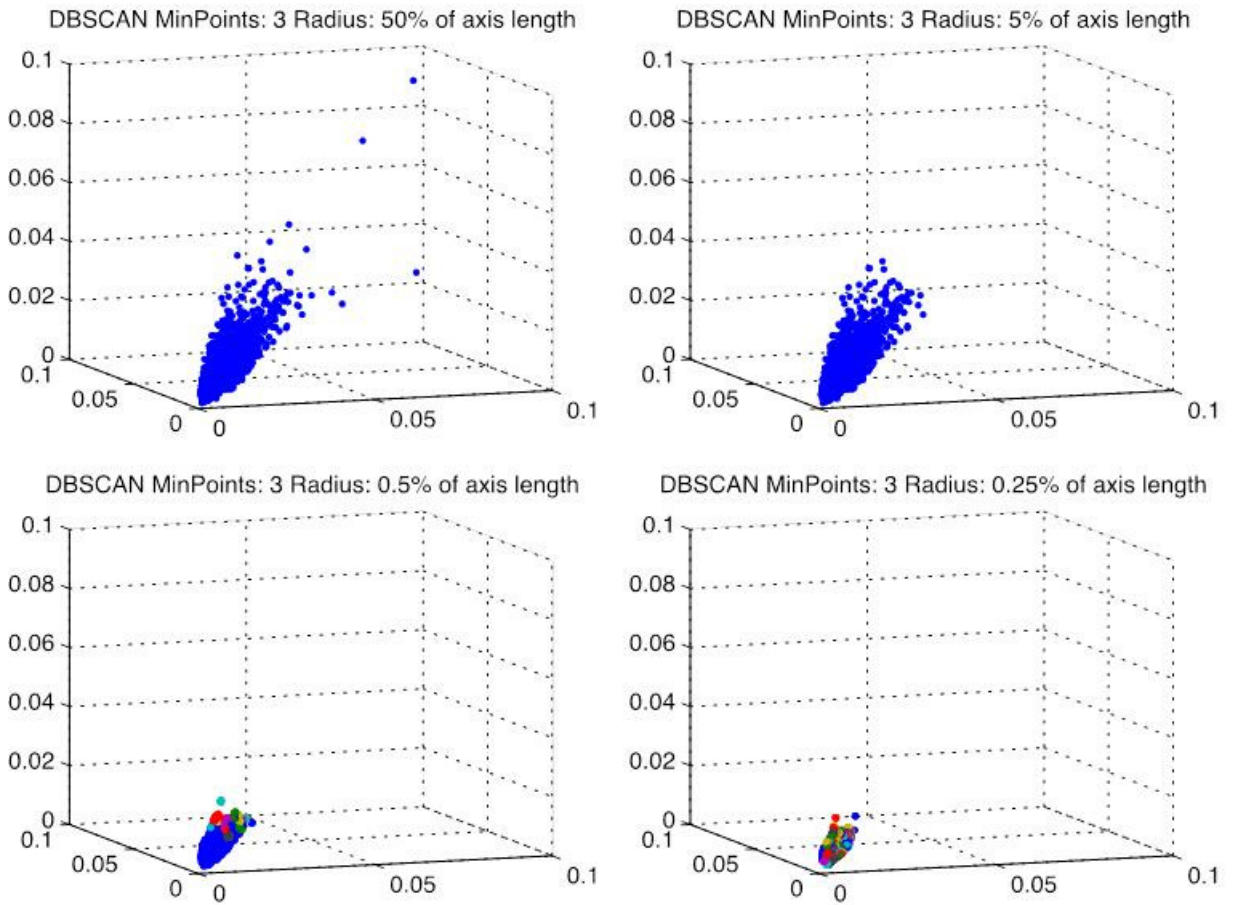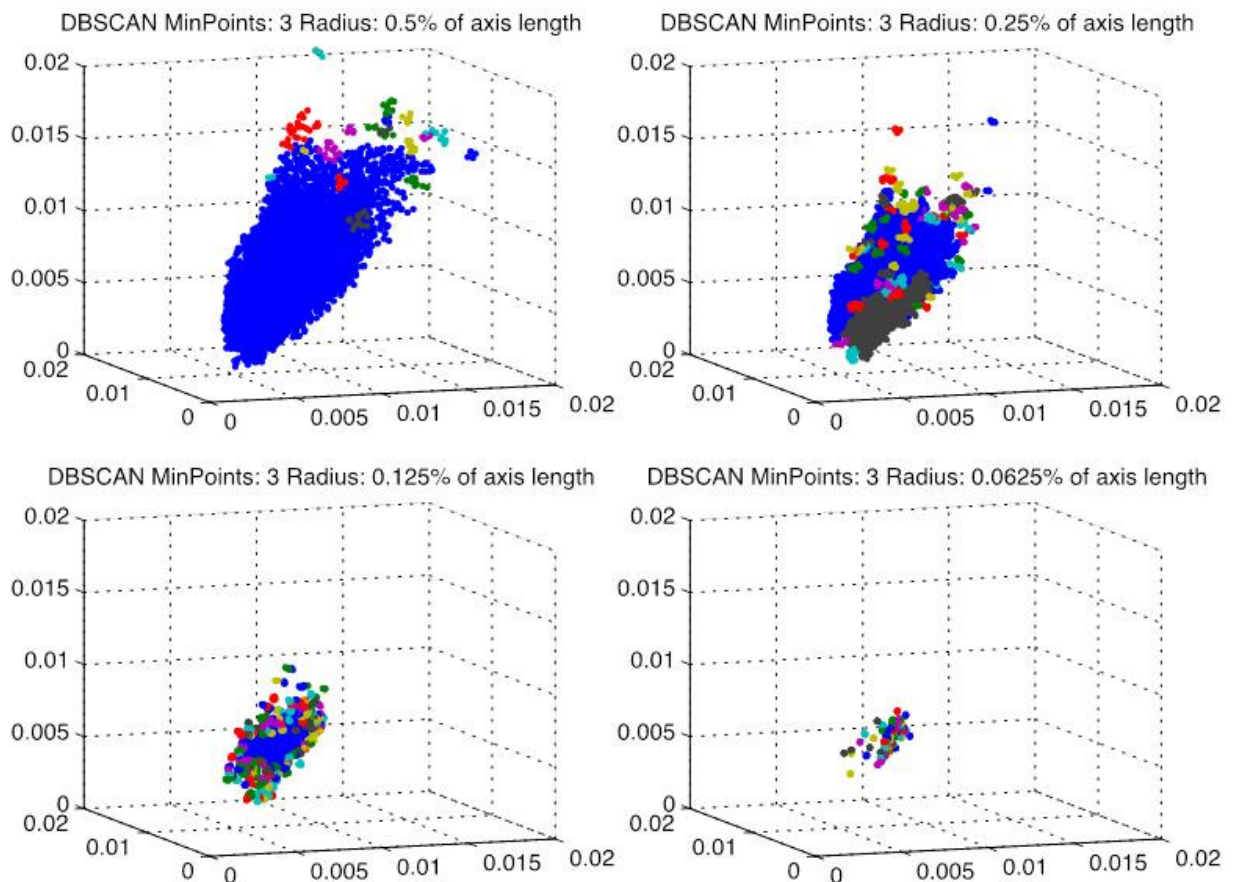**Figure 12-9   Variation in DBSCAN Cluster Allocation Under Manual Control (1 of 2)**

12.7. Discussion

12.7.1. Choice of Algorithm

The clustering approach taken for EnergyPath^{TM} Networks needs to balance the production of useful outputs for dynamic energy simulation in the HOM against the needs of the end user. It is therefore first worth revisiting assumptions about the conceptual purpose of the mathematical clustering exercise, the format of the required outputs, and the inputs from the end user:

- ► Conceptually, the purpose of the mathematical clustering stage is to collapse large volumes of data on different dwelling types into a manageable number of representative classes, allowing the level of detail to be matched to the computational power available for simulation and/or optimisation in other EPN modules.

  - It is envisaged that the primary reason for adjusting the clustering process in the tool for a given local authority area would be to increase the granularity (i.e. number) of information that is reflected in the pathway optimisation

- ► The outputs from the clustering exercise are intended to feed directly into dynamic building thermal simulation in EnergyPlus. This is currently envisaged as a single set

of thermal mass/heat loss/size values that are representative of each cluster, as opposed to a range or distribution of values.

- ▶ The shape and distribution of the clusters produced should be sufficient to pick up when a significant change to a building's properties have occurred, such as after an energy efficient retrofit.

- ▶ Finally, the operating assumption regarding inputs is that end users will have a variety of datasets from which they seek to build a representative picture of the local housing stock, and that the solution used for grouping similar dwelling types together into a proxy representation will need to be automated as much as possible i.e. target end users are not mathematicians or data mining experts and will not habitually seek to vary algorithmic parameters in order to get the "best" grouping of the data.

With these requirements in mind, it is possible to narrow down the choice of approach. The DBSCAN algorithm is not appropriate for a number of reasons:

- ▶ The algorithmic formulation requires the end-user to decide for themselves what forms a 'cluster' i.e. the number of objects and a search radius. This is difficult to conceptualise and it is not obvious what the optimum settings are in the problem space being considered, nor how this will change with different data or different densities.

- ▶ The algorithm is also designed to eliminate data that is not relevant to cluster formation as "noise". However, in the context of the EPN tool it is likely that the only data that will be input to the model by the end user will be real world observed or collected data, and therefore data that are relevant for inclusion in cluster formation.

- ▶ Clusters established using DBSCAN have no obvious 'centre', so an additional step is required to select which values are representative and should be used as outputs.

- ▶ The number of clusters cannot be easily pre-specified, and although an additional optimisation step could be used to achieve density-based clustering subject to an upper limit on the number of classes available in the solution, there are more straightforward ways of meeting the required outcome.

Both hierarchical and centroid-based approaches allow the number of target clusters used for data grouping to be easily pre-specified, which facilitates the flexing of detail to match available computer performance (this could potentially be automated, as illustrated above in Section 12.4.1 and Section 12.5.1).

K-means clustering, unlike hierarchical clustering, is a non-deterministic search optimisation method and solutions found are not always replicable. The method requires seeding of start points, and the outcome is sensitive to the initial guess for where these are located in the problem space[59]. However, in the context of the EnergyPath™ project, this factor alone is of

---

[59] Employing the same start points will however, yield the same outcome.

second order importance compared to the quality of the hierarchical clustering approach, and it can be mitigated against by running the algorithm multiple times to avoid local minima[60].

More important is the qualitative shape of the clusters and the allocation of data points to different groups. This can be illustrated by a visual comparison of the cluster arrangements produced for a 50-cluster group using k-means and agglomerative linkage, as shown below in Figure 12-11. This illustrates that:

1. Agglomerative linkage essentially identifies a single monolithic structure that encompasses the lion's share of the data points, with the other clusters effectively being ~49 outliers, each comprised of a single point
2. K-means clustering produces 50 cellular regions that are fairly evenly distributed throughout the problem space

For the purposes of mathematical clustering on densely grouped data in EnergyPath[TM], the latter result is more desirable. The k-means solution will enable significant changes in building performance to be represented by allocating a dwelling to different cluster groups, whereas the agglomerative linkage result would have buildings with extreme differences in performance being considered as part of the same group.

Figure 12-11   **Qualitative Comparison of 50 Clusters Defined Using Agglomerative Linkage and k-means Methods**
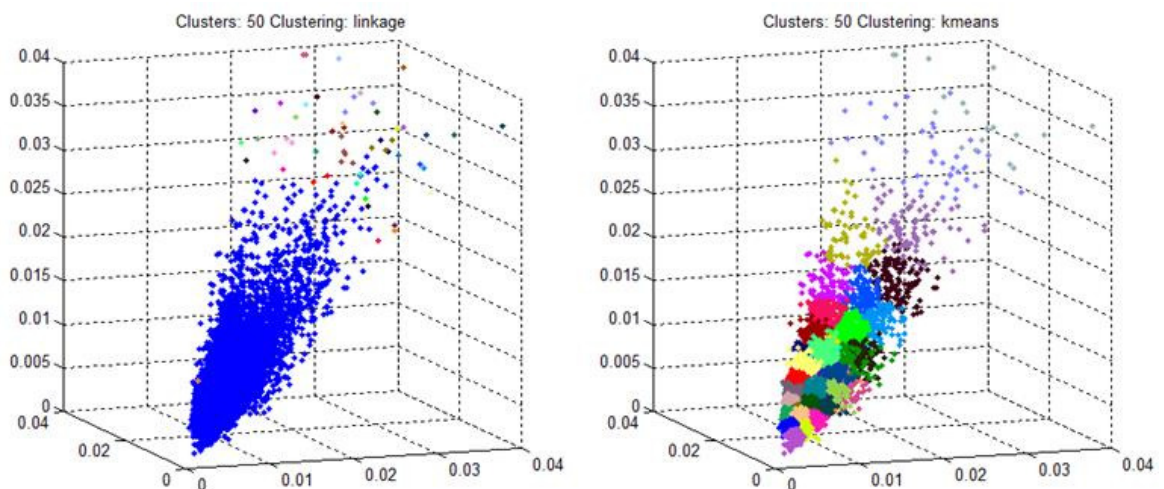
12.7.2. Choice of Evaluation Method

For automated clustering in k-means, the gap evaluation method is found to give the best performance in terms of the distribution of clusters. Evaluation using Calinski Harabasz, Davies Bouldin, and Silhouette methods were found to recommend low numbers of clusters i.e. 2-3 only, which is likely to be insufficient to capture the underlying degree of variation in the problem space being considered.

---

[60] DBSCAN is also non-deterministic (core points and noise points are always the same but border points are sometimes allocated differently between core and noise areas)

12.8. Conclusions

Having assessed a variety of clustering methodologies and evaluation approaches, the objectives of the study set out in Section 12 can now be reviewed:

1. *What type of mathematical clustering approach is the most suitable for application in EnergyPath^TM (centroid, hierarchical, density), given the type of data inputs that are to form the basis of the analysis?*

   It is recommended that design proceeds on the basis of a centroid-based method such as k-means clustering.

2. *What settings give the best performance? I.e. what settings with which clustering approach arrive most rapidly at an optimum number of clusters with the best level of separation between cluster types?*

   Initial testing discussed in this report revealed that the gap evaluation approach appeared to give cluster arrangements with the most useful separation for investigating the type of problem explored in EnergyPath^TM. However, the decision on the most appropriate number of clusters ultimately will need to be determined by considering program-wide parameters such as the limit on the total number of decision variables for evaluation in the Pathway Optimisation Module (POM) and the materiality of impact on the POM solution. The k-means methodology illustrated above takes the number of clusters as an input argument, so implementing this in code is straightforward.

3. *Are the dimensional parameters outlined in the functional specification suitable for differentiating between different archetypes, or do additional metrics need to be included in the clustering analysis process?*

   The functional specification identifies thermal mass and heat loss parameters as key performance metrics. The results of this study also show that the total size of each dwelling, for example expressed in units of floor area, are also material for differentiating between different building archetypes.

12.9. References

1.   Calinski T, Harabasz J. A dendrite method for cluster analysis. Commun Stat - Theory Methods [Internet]. 1974 [cited 2014 Jul 17];3(1):1–27. Available from: http://www.tandfonline.com/doi/abs/10.1080/03610927408827101

2.   Davies DL, Bouldin DW. A Cluster Separation Measure. IEEE Trans Pattern Anal Mach Intell [Internet]. 1979 Apr [cited 2014 Jul 14];PAMI-1(2):224–7. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4766909

3.   Kaufman L, Rousseeuw PJ. Finding Groups in Data [Internet]. Kaufman L, Rousseeuw PJ, editors. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1990 [cited 2014 Jul 17]. Available from: http://doi.wiley.com/10.1002/9780470316801

4.   Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B (Statistical Methodology [Internet]. 2001 May

[cited 2014 Jul 17];63(2):411–23. Available from:
http://doi.wiley.com/10.1111/1467-9868.00293

5.   Ester M, Kriegel H-P, Sander J, Xu X. A Density-Based Algorithm for Discovering
        Clusters in Large Spatial Databases with Noise. In: Simoudis E, Han J, Fayyad UM,
        editors. Proceedings of the Second International Conference on Knowledge
        Discovery and Data Mining (KDD-96) [Internet]. AAAI Press; 1996. p. 226–31.
        Available from: http://www.aaai.org/Library/KDD/kdd96contents.php