Forward guidance – Modelling local and regional infrastructure supply

Technical paper

Final Report

Prepared by: **Motu Economic and Public Policy Reaserch**

Prepared for: **New Zealand Infrastructure Commission Te Waihanga**

Authors: **Stuart Donovan, Leon Stirk-Wang,**

January 15, 2026

Table of Contents

Executive Summary

PURPOSE OF THIS RESEARCH

The New Zealand Infrastructure Commission Te Waihanga (the "Commission") engaged Motu Research to assist with the provision of forward guidance for infrastructure. The purpose of this assistance was to develop methods to disaggregate national infrastructure estimates to the local level.

SCOPE OF OUR ANALYSIS

Table 1 summarizes the type, sectors, sub-sectors, supply measure, spatial scale, and coverage of infrastructure that were included in our analyses. For the horizontal sub-sectors, we assigned data on the length of infrastructure to a 1km grid produced by Statistics New Zealand. For education and health, data on capital values were provided to Motu Research at the SA1 level. We have national coverage for all sectors with the exception of water, where we have data for approximately 38.7% of the 1km grid cells that comprise New Zealand's land area. For all sub-sectors, we predict infrastructure supply in two scenarios: 2023 (i.e. the "current" scenario) and 2048 (i.e. the "future" scenario).

Table 1. Infrastructure types by type, sector, sub-sector, supply measure, spatial scale, and coverage

Type	Sector	Sub-sector	Supply measure	Spatial scale	Coverage
Horizontal	Roads	Local roads	Length	1km grid	100%
		State highways	Length	1km grid	100%
	Water	Supply	Length	1km grid	38.7%
		Waste	Length	1km grid	38.7%
		Storm	Length	1km grid	38.7%
	Electricity	Distribution (<11kV)	Length	1km grid	100%
Vertical	Education	Primary/Secondary	Capital value	SA1	100%
	Health	Hospitals	Capital value	SA1	100%

METHODOLOGY

Our approach to modelling infrastructure supply was informed by several considerations. First, since we want to predict the supply of local infrastructure both now and in the future, we adopt a parsimonious modelling approach and restrict ourselves to explanatory variables for which sub-national projections are readily available. Specifically, our models use a single causal explanatory variable: *population*.[1] For each sub-sector, we combine population data from Statistics New Zealand with infrastructure data from various sources.[2] Figure 1 illustrates infrastructure supply versus population (log) for the six horizontal sectors in Table 1 along with a non-linear, non-parametric trend line.[3] For local roads and three waters, we see that supply ramps-up above certain population thresholds. In contrast, we observe a more muted relationship between population and the supply of state highways.

Second, our methodology seeks to address three empirical challenges:

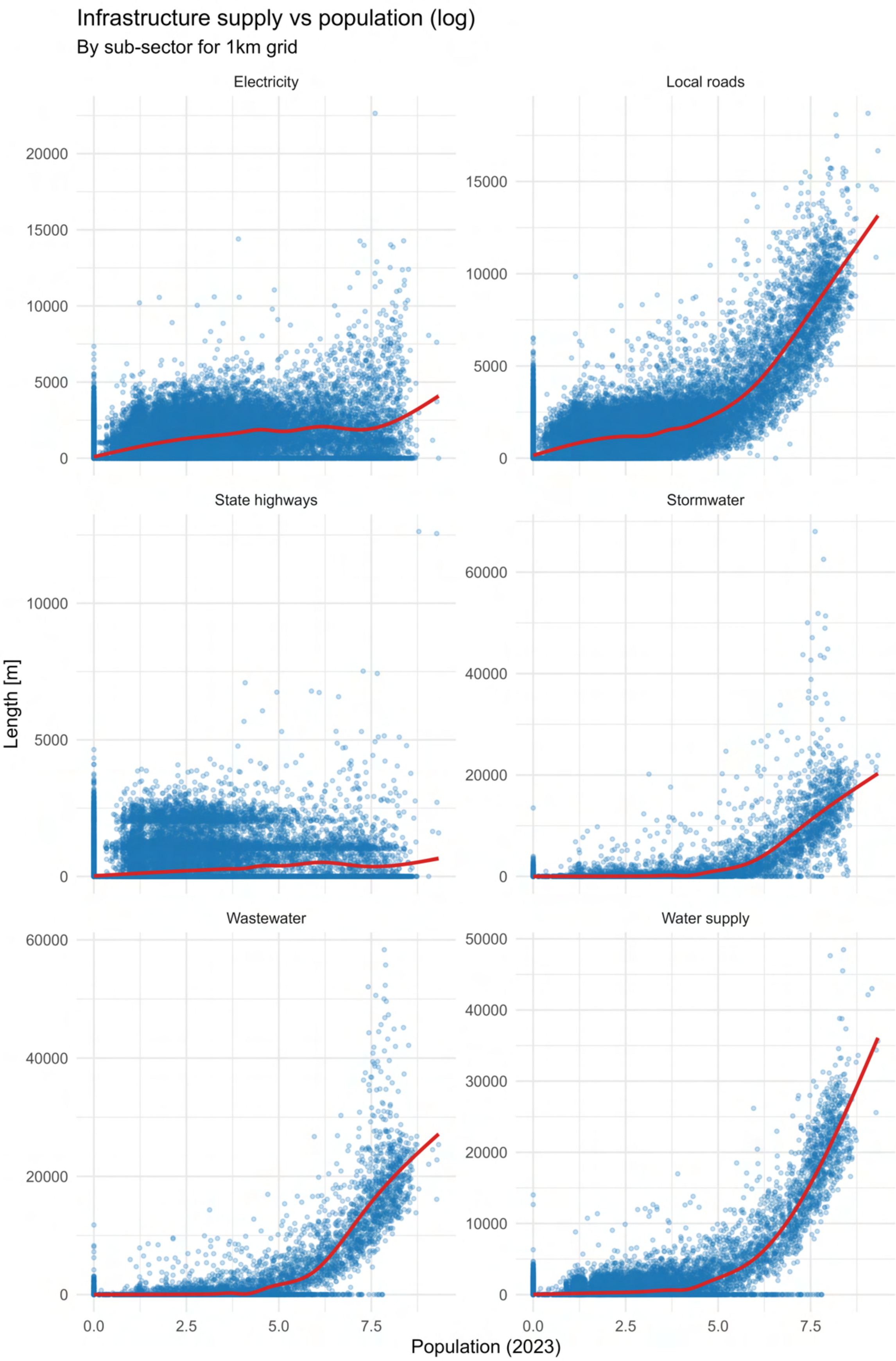
- First, infrastructure supply is highly *heterogeneous*: some places have high levels of infrastruc-

[1] Previous work by the Commission, including the Infrastructure Needs Analysis technical report[1] and Paying it Forward[2], found that population growth is a significant explanatory variable for country-level infrastructure.

[2] For the two models that use SA1s, that is, education and health, we adjust population using sector-specific age-based demand weights that were supplied by the Commission. Further details on these weights are provided in the body of the report.

[3] To avoid taking logs of zeros, we add 1 to the population in all grid cells.

Figure 1. Infrastructure supply versus population (log) for six horizontal infrastructure sectors



ture, whereas many have little to none. Indeed, most locations have zero infrastructure.

- Second, the supply of infrastructure exhibits spatial *spillovers*: The probability that infrastructure exists in one location is affected by the existence of infrastructure in adjacent locations.
- Third, infrastructure supply and population are *endogenously determined*: population affects the supply of infrastructure, and vice versa. That is, causality runs in both directions.

To address heterogeneity in infrastructure supply, we estimate "hurdle-Gamma" models. The hurdle component models whether infrastructure exists at all (a binary process), whereas the gamma component models the level of supply (a continuous process). The hurdle component is, in our view, a novel and useful aspect of our methodology because fixed costs mean that most locations have zero infrastructure and because some locations — such as those on the periphery of growing urban areas — might transition from zero infrastructure now to positive levels of supply in the future. We also address heterogeneity by including individual spatial effects, e.g. for SA1s, SA2s, and local authorities.

Although the inclusion of individual spatial effects also helps to address the challenge posed by spillovers, or "network effects", within spatial units, these effects can also operate over the boundaries between spatial units. To help control for spatial spillovers in local infrastructure supply that are not captured by the individual spatial effects, we also include the spatial lag of population in the hurdle component of the models. In most models, we find evidence that these spillovers — as proxied by lagged population — are important for explaining local infrastructure supply.

Finally, we address endogeneity via two methods. First, as noted above, we include individual spatial effects for SA1s, SA2s, and local authorities, which helps to control for omitted variables. Second, we use a control function to address residual endogeneity in infrastructure supply and population.[4]

MAIN RESULTS

Table 2 summarises the effects of (log) population on infrastructure supply for the two parameters that are of primary interest. First, we present the parameter of the hurdle sub-model, α^P, which defines the effect of population on the *existence of infrastructure*. Second, we present the parameter from the gamma sub-model, β^P, which defines the effect of population on *levels of infrastructure supply*. In Table 2, we also present results for a Baseline model and an Extended model, which are identical except that the latter includes an additional non-linear, non-parametric function to control for endogeneity.

As all models use a 1km grid *or* SA1s with controls for the area, we can interpret the estimates of α^P and β^P in Table 2 as the effect of population when the area is kept constant. Simply put, α^P and β^P measure the effect of changes in the *population density* of grid cells or SA1s. And because the gamma component uses a logarithmic link in which the population also enters in logarithmic units, the estimates of β^P can be interpreted as a conventional (constant) supply elasticity. In Table 2, we observe that estimates of $\beta^P < 1$ for all sub-sectors and in both models. This implies that the supply of local infrastructure changes less than proportionally to the population density.

For the vertical sectors of health and education where we have data on capital values, our results indicate economies of density in infrastructure supply.[5] Although our estimates of β^P for the horizontal sectors are also consistent with economies of density, these models measure supply based on the length of infrastructure rather than its capital value. As length will not capture the effects of population on infrastructure capacity, such as the width of roads or pipes, we cannot draw firm conclusions on whether these sub-sectors experience economies of density. To draw stronger inferences on this question, we would need data on the capital value of horizontal infrastructure at a local level.

[4] We instrument population with the crow-fly distance to regional centres and the coast, which are strong instruments.
[5] As we do not model operating costs, our results do not provide evidence of overall economies of density.

Table 2. Estimated parameters for population from the hurdle, α^P, and gamma, β^P components of the models. The latter can be interpreted as the constant elasticity of infrastructure supply with respect to population. The Baseline and Extended specifications are identical except the latter uses a control function to address endogeneity. Standard errors are shown in parentheses, which allow for heteroskedastic variance per SA1, SA2, and local authority.

Sector	Sub-sector	Baseline		Extended	
		Hurdle, α^P	Gamma, β^P	Hurdle, α^P	Gamma, β^P
Roads	Local roads	1.831 (0.013)	0.254 (0.002)	1.061 (0.029)	0.304 (0.003)
	State highways	0.669 (0.012)	0.050 (0.004)	0.291 (0.025)	0.021 (0.005)
Water	Supply	0.821 (0.022)	0.383 (0.006)	1.407 (0.066)	0.469 (0.006)
	Waste	0.849 (0.035)	0.476 (0.007)	1.206 (0.067)	0.533 (0.009)
	Storm	1.033 (0.026)	0.575 (0.010)	1.241 (0.052)	0.775 (0.013)
Electricity	Distribution	1.496 (0.011)	0.218 (0.002)	0.317 (0.030)	0.184 (0.003)
Education	Primary, Secondary	0.117 (0.023)	0.339 (0.025)	0.111 (0.025)	0.410 (0.027)
Health	Hospitals	0.437 (0.065)	0.506 (0.063)	0.417 (0.070)	0.525 (0.066)

Figure 2. The top and bottom panels compare observed and median predicted outcomes for local roads and state highways, whereas the left and right panels compare outcomes at the grid cell and regional levels. The diagonal lines denote where observed and predicted outcomes are equal.

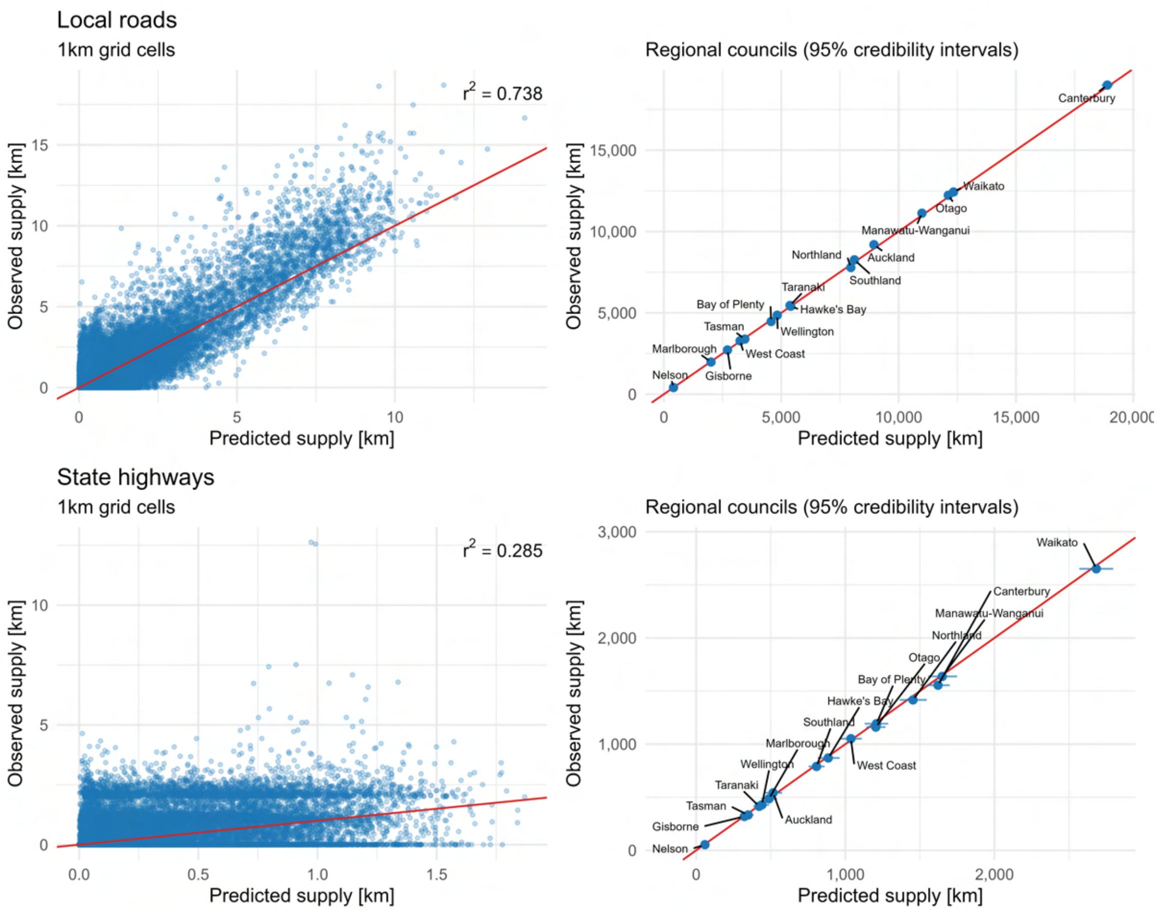
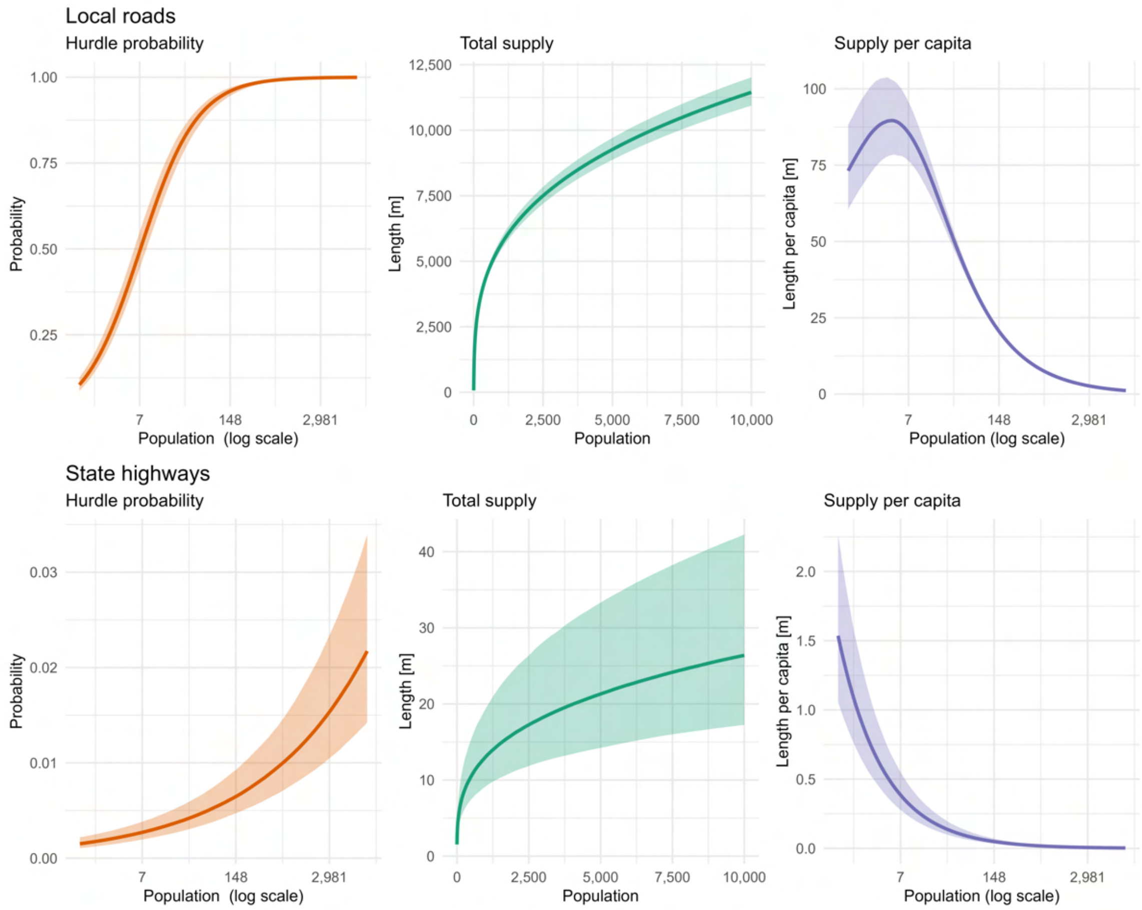


Figure 3. Implications of the Extended model for local roads (top) and state highways (bottom). The left panel shows the probability that infrastructure exists, the middle panel shows total infrastructure supply, and the right panel shows infrastructure supply per capita. Shaded areas denote 95% credibility intervals.



Regarding model performance, Figure 2 compares observed and predicted outcomes for local roads and state highways (results for other sub-sectors are presented in the main body of the report). Although we observe substantial unexplained variation in outcomes at the local level, especially for state highways, we find an extremely strong positive correlation at the regional level. The strong predictive power of these models at the regional level likely reflects the inclusion of spatial effects for SA1s, SA2s, and local authorities. We find the weakest correlation between observed and predicted outcomes for health, which likely reflects the concentrated and specialised nature of this sub-sector. In general, we are pleased to find that our models seem to be able to identify the causal effects of population on infrastructure supply *and* generate reasonably accurate predictions of supply at the regional level.

Finally, Figure 3 illustrates some implications of our models for local roads (top) and state highways (bottom). The left panel shows the hurdle probability that infrastructure exists; the middle panel shows the total supply of infrastructure; and the right panel shows the supply per capita. Several key implications of the models are evident in Figure 3. First, the probability that local roads exist in a grid cell approaches 100% at modest population levels, whereas the corresponding probability for state highways does not exceed approximately 1.5–3.5%. Second, we find that infrastructure supply increases with population but at a reducing rate, giving rise to a concave shape that is consistent with economies of density as discussed above. Third, the supply per capita tends to decrease with population, although the rate of decline decreases as the population increases.

DISCUSSION

We find hurdle-Gamma models provide a promising way to model the supply of infrastructure at a local level for eight relatively heterogeneous sub-sectors. These models appear able to combine robust causal identification with decent predictive power, at least at the regional level and with the possible exception of health. Estimated shares of infrastructure are relatively persistent over time, at least at the regional level. This likely reflects how several of the fastest-growing regions, such as Auckland and Canterbury, are already relatively populous and thus tend to benefit more from economies of density.

Notwithstanding these promising results, our approach is somewhat novel and would benefit from more ground-truthing. An obvious next step, for example, would be to perform a detailed systematic review of the literature on economies of density and scale in infrastructure supply. Although such a review was outside the scope of this research, our preliminary scan of the literature identified several studies that might provide external reference points to which our results can be compared [see, for example, 3, 4, 5, 6, 7]. However, drawing effective comparisons with the external literature will require care in reconciling differences in the spatial scale of different analyses. In this context, a detailed systematic review presents a useful albeit not necessarily straightforward direction for further research.

Other potential directions for further research include but are not necessarily limited to:

- For health and education, using a randomly-generated 1km grid rather than SA1s
- For horizontal sub-sectors, using data on capital values rather than infrastructure length
- Grounding our empirical models in theoretical economic models of infrastructure supply
- Investigating other plausibly exogenous instruments, such as vertical elevation
- Including additional explanatory variables, such as topographical complexity
- Incorporating updated population projections from Statistics NZ
- Modelling the change in infrastructure supply over time
- Applying models to other infrastructure sectors

In our view, the first two items on this list are a high priority. Specifically, implementing these items would allow us to estimate models that use consistent spatial units and outcome measures across all sub-sectors. This consistency would reduce the vulnerability of our results to empirical problems, such as the modifiable area unit problem ("MAUP") and measurement error in the supply variable, both of which could be a potential source of bias. In general, we suggest that our results are interpreted judiciously and applied carefully, at least until further research can confirm the merits of our approach. Nonetheless, we consider that this work provides a useful starting point for further research.

Although the primary purpose of our work was to help the Commission provide forward guidance, the potential exists for our findings to be applied in other contexts, including but not limited to:

- *Land use planning*, such as that undertaken by local authorities
- *Infrastructure planning*, such as the setting of development contributions or levies
- *Economic analysis*, for example, by embedding these models within quantitative spatial models

Further research may well wish to consider potential applications like these.

Finally, we see our resulting estimates of local infrastructure shares as a useful complement to, rather than a replacement for, those presented in sector-level strategies. Whereas our approach has the benefit of being grounded in empirical evidence on the effects of population, sectoral strategies may be better positioned to respond to wider considerations, such as the effects of new technologies.

Acknowledgement and disclaimer

Motu Research acknowledges funding from the New Zealand Infrastructure Commission, which made this research possible. The contents of this report, including but not limited to any errors, remain the responsibility of the authors alone and not that of the Commission.

1 Introduction

1.1 Background

The New Zealand Infrastructure Commission Te Waihanga (the "Commission") engaged Motu Economic and Public Policy Research ("Motu Research") to assist with the provision of "Forward Guidance" as part of the National Infrastructure Plan. The purpose of this assistance was to help the Commission disaggregate national estimates of infrastructure investment to the local level in current and future scenarios. Further information on how the results of this work have been used by the Commission is provided in the latter's technical report on modelling and forecasts for the overall Forward Guidance, while the resulting estimates are detailed in the National Infrastructure Plan itself.

1.2 Scope of our analyses

Table 3 summarizes the type, sectors, sub-sectors, supply measure, spatial scale, and coverage of the infrastructure data and analysis addressed by our research. For the horizontal sub-sectors, we assigned data on the length of infrastructure to a 1km grid produced by Statistics New Zealand. For education and health, data on capital values were provided to Motu Research at the SA1 level. We have national coverage for all sectors with the exception of water, where we have data for approximately 38.7% of the 1km grid cells that comprise New Zealand's land area. For all sub-sectors, we predict infrastructure supply in two scenarios: 2023 (the "current" scenario) and 2048 (the "future" scenario).

Table 3. Infrastructure types by type, sector, sub-sector, supply measure, spatial scale, and coverage

Type	Sector	Sub-sector	Supply measure	Spatial scale	Coverage
Horizontal	Roads	Local roads	Length	1km grid	100%
		State highways	Length	1km grid	100%
	Water	Supply	Length	1km grid	38.7%
		Waste	Length	1km grid	38.7%
		Storm	Length	1km grid	38.7%
	Electricity	Distribution (11kV)	Length	1km grid	100%
Vertical	Education	Primary/Secondary	Capital value	SA1	100%
	Health	Hospitals	Capital value	SA1	100%

More information on the data that we use is provided in Section 2.1 and Appendix A.

1.3 Limitations

We note three important limitations of our analyses prior to discussing our methodology and approach.

First, our measure of supply for horizontal sub-sectors is the length of infrastructure. However, in many places infrastructure investment appears to be increasingly focused on addressing capacity constraints, rather than expanding network length. As such, our estimates should be interpreted less as where infrastructure investment is needed and more as where growth pressures may arise.

Second, we emphasise that our future forecasts are subject to several sources of uncertainty. There is uncertainty, for example, in the models and associated parameters that we use to estimate local infrastructure supply. Additional uncertainty is also introduced by Statistics NZ's population projections, which are periodically revised — often substantially — over time.

Third, our data has some notable limitations. This is especially relevant to the water sub-sectors, where we must use our models to estimate local infrastructure supply in the 61.3% of grid cells where we do not observe infrastructure supply. Similarly, the data for water are sourced from multiple different entities, increasing the risk of inconsistencies in the raw data.

2 Methodology

This section outlines our methodology for modelling local infrastructure supply. First, we summarise the data used in our analyses; second, we present the empirical models that we use to predict infrastructure supply; and, finally, we explain how we use the models to post-process infrastructure shares.

2.1 Data

Table 4 summarises the data sources employed in this analysis. Detailed descriptions of individual sources are provided as citations, while Appendix A focuses on documenting the spatial processing, allocation, and transformation steps used to construct the datasets used in the analyses. This separation reflects the fact that most complexity in the data arises not from the raw inputs themselves, but from the spatial and temporal harmonisation that is required to make them suitable for modelling.

Table 4. Summary of datasets used in our analyses and their sources

Category	Base year	Dataset	Source
Geographic Boundaries	2018	SA1 boundaries	StatsNZ [8]
		SA2 boundaries	StatsNZ [9]
		TA boundaries	StatsNZ [10]
		RC boundaries	StatsNZ [11]
	2023	250 m grid	StatsNZ [12]
		500 m grid	StatsNZ [13]
		1 km grid	StatsNZ [14]
Infrastructure	NA	State Highways	LINZ [15]
		Local roads	LINZ [15]
		Electricity distribution	Infrageomatics [16]
		3-waters networks	Wellington Water [17, 18, 19]
			KCDC [20, 21, 22]
			SWDC [23, 24, 25, 26, 27, 28]
			Watercare [29, 30, 31]
			Canterbury RC [32, 33, 34]
			Southland [35, 36, 37]
			Tasman [38, 39, 40]
			Waikato [41, 42, 43]
		Hospitals	ESNZ [44]
		Schools	ESNZ [44]
Population	2018	Census 2018	StatsNZ[45]
		Subnational population projections	StatsNZ [46]
	2023	Grid population estimates (250 m)	StatsNZ [47]
		Grid population estimates (500 m)	StatsNZ [48]
		Grid population estimates (1 km)	StatsNZ [49]
		Census 2018 population by age	StatsNZ [45]
		Coastlines polygons	LINZ [50]

2.1.1 Sources of data

The analyses makes use of three main classes of data, specifically:

1. **Boundary geometries,** which define the spatial units of analysis.

- *Grid boundaries*: 250 m, 500 m, and 1 km grid layers obtained from Stats NZ (Datafinder).
- *Administrative boundaries*: SA1, SA2, Territorial Authority (TA), and Regional Council (RC) layers from Stats NZ (Datafinder). The 2018 boundaries are used throughout, primarily because they align with the 2018 Census population data used as explanatory variables.

2. **Infrastructure supply data,** which represents horizontal networks and vertical infrastructure.

   - *Horizontal networks*: roads, electricity distribution lines, and three-waters networks (water supply, wastewater, stormwater). Sources and coverage vary by infrastructure type.
   - *Vertical infrastructure*: education and health infrastructure capital values per SA1 were obtained from Earth Sciences New Zealand.[6]

3. **Explanatory, control, and instrument variables,** which are used for demand modelling.

   - Stats NZ grid population estimates (250 m, 500 m, and 1 km grids).
   - 2018 Census population data (SA1). Including usually resident population counts disaggregated by 5 year age bands (for modelling age-weighted demand for vertical infrastructure like schools and hospitals) and totals (used for horizontal networks)
   - Stats NZ subnational population projections (SA2, 2018 base). These provide high, medium, and low scenario projections of age-disaggregated population at five-year intervals from 2018 to 2048 at the SA2 geography.
   - Crow-fly distances to largest regional centre and coast (instrument for population)

2.1.2 Exploratory analyses

Tables 5 and 6 provide descriptive statistics for the infrastructure and 2023 population variables for the 1km grid and SA1 geometries, where the mean is calculated over all spatial units, e.g. total number of grid cells or SA1s.[7] For all sub-sectors, zeros are the most common outcome.

Table 5. Summary statistics for grid-level variables (1 km grid)

Variable	Mean	Max	Zero share (%)
Local roads (km)	0.409	18.7	67.9
State highways (km)	0.055	12.6	95.1
Water supply (km)	0.308	48.5	91.8
Stormwater (km)	0.162	68.0	95.1
Wastewater (km)	0.221	58.3	97.1
Electricity distribution (km)	0.313	22.6	74.2
Population (2023)	18.915	11,146.7	79.2

Table 6. Summary statistics for SA1-level variables

Variable	Mean	Max	Zero share (%)
Education capital (NZD $000s)	989.120	351000.0	88.0
Hospital capital (NZD $000s)	338.061	914000.0	97.3
Population (unweighted)	172.858	1665.7	2.6

[6] ESNZ mapped data on building footprint and envelope information for a related report for the Commission on natural hazard risk. See “Estimating National Scale Losses to Infrastructure from Natural Hazards” report by GNS Science [44].

[7] For example, the mean hospital value is $338,000 per SA1. However, because 97.3% of SA1 units contain 0 hospital capital, the value of the average hospital will be much higher than the mean spending per SA1.

Figure 4. Infrastructure supply versus population (log) for six horizontal infrastructure sectors

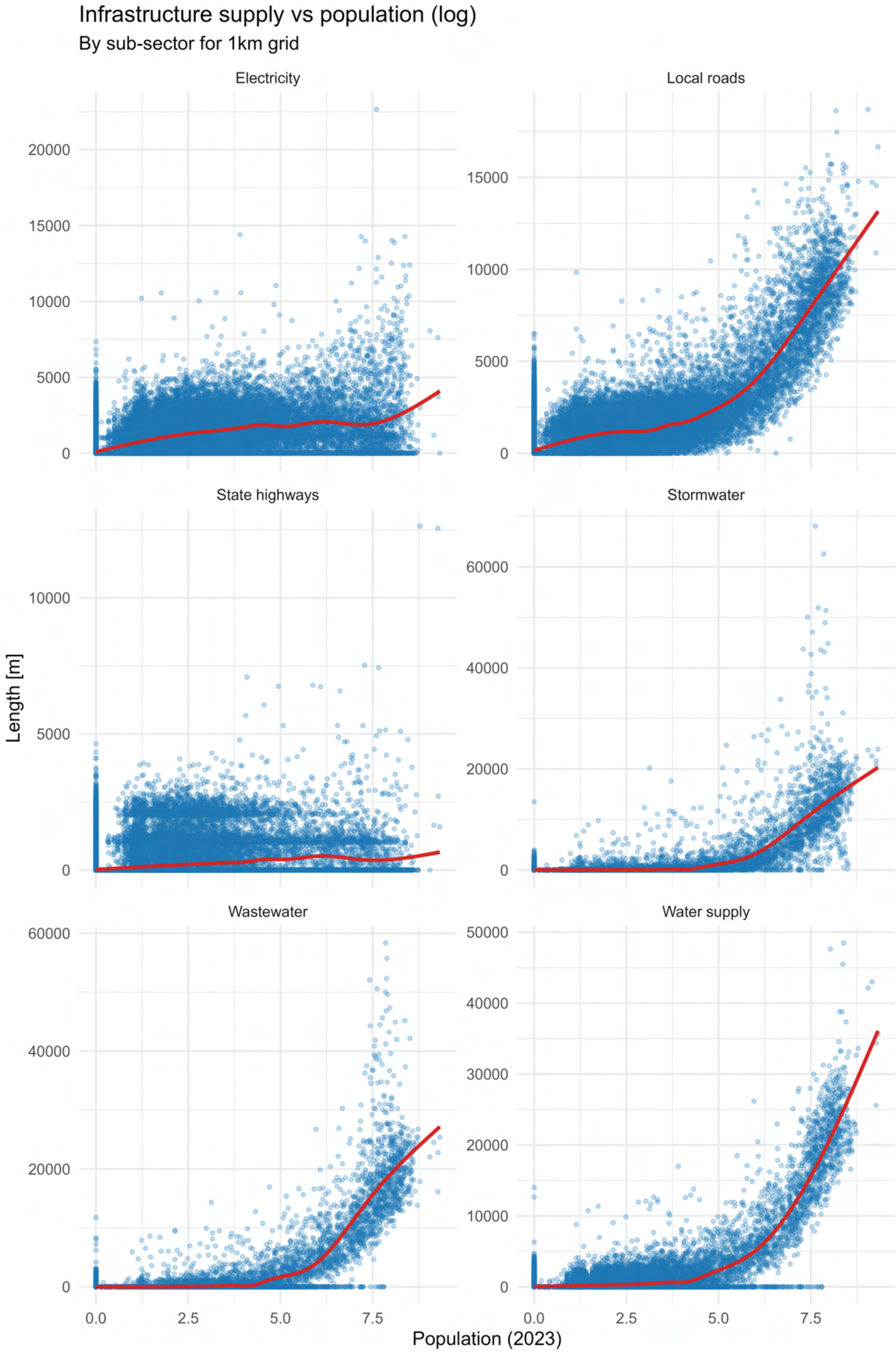
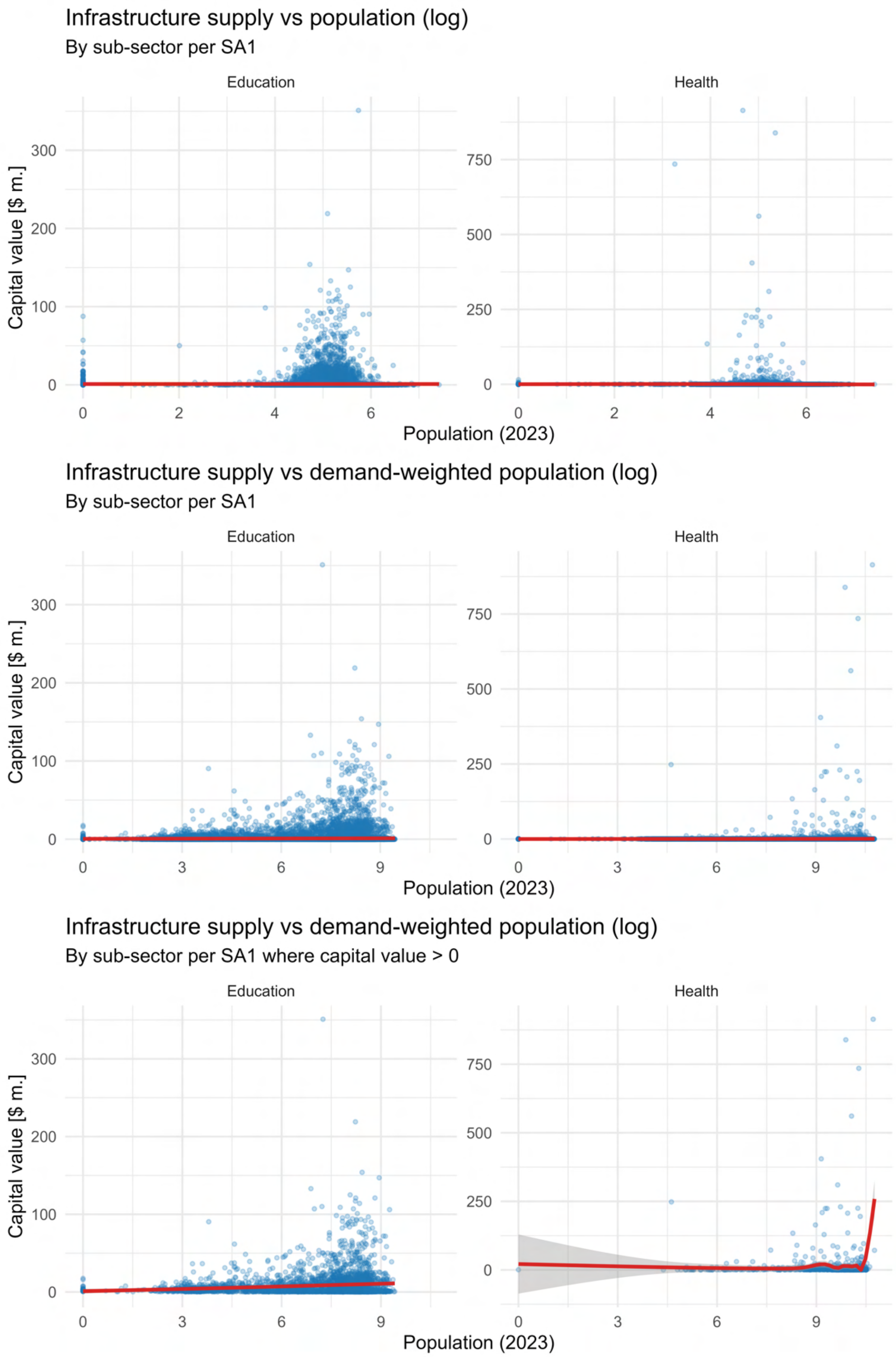


Figure 5. Infrastructure supply versus population (log) for education (left panels) and health (right panels). The top panel shows capital value versus population; the middle panel show capital value versus demand-weighted population; and the bottom panel shows capital value > 0 versus demand-weighted population.



Figures 4 and 5 present scatter plots of infrastructure supply (vertical axes) versus population (horizontal axes) for each of the horizontal and vertical infrastructure sub-sectors that we analyse. For each panel, we also add a non-linear, non-parametric trend line.

The relationship between infrastructure supply and population generally appears to be stronger for the horizontal sectors shown in Figure 4 compared to the vertical sectors shown in Figure 5. In Figure 4, we observe relatively muted but consistent relationships between population and the supply of electricity distribution and state highways. This contrasts strongly with the relationship observed for local roads and the three waters, where supply increases strongly above approximately 150 people per square kilometre. Notwithstanding some similarities between horizontal sub-sectors, we also observe notable differences. At low levels of population, for example, the supply of local roads is often positive whereas the supply of three waters is usually non-existent. We expect the latter empirical outcome is likely to reflect higher fixed costs in the provision of water infrastructure vis-à-vis local roads.

For the two vertical sectors in Figure 5, we plot three variants. First, the top panels show the value of the infrastructure versus the population within 2 km of the centroid of the SA1. Second, the middle panels show the infrastructure value versus the demand-weighted population within 2 km of the centroid of SA1, where 5-year population age bands are weighted based on usage.[8] Third, the bottom panels are the same as the middle panels except they only plot data for SA1s where capital values > 0. Interestingly, only the latter set of panels indicates a positive relationship between infrastructure supply and population. For education, the positive relationship with the demand-weighted population is fairly consistent across the range of population that we observe. In contrast, for health, we observe a relatively flat relationship below relatively high populations. The latter relationship may reflect the tendency to concentrate capital investment in fewer, more specialised hospital facilities, whereas education uses a more distributed service delivery model, especially for primary school facilities.

The summary statistics in Tables 5 and 6 as well as the exploratory analyses in Figures 4 and 5 support the adoption of empirical models that can accommodate heterogeneity in supply. In all sectors, for example, zero infrastructure is by far the most likely outcome. For this reason, we consider it important to adopt an empirical methodology that can accommodate the existence or otherwise of infrastructure in addition to predicting levels of supply. The following section outlines an empirical model in this spirit.

2.2 Empirical model

2.2.1 Baseline specification

HORIZONTAL INFRASTRUCTURE

We begin by presenting our Baseline specification for horizontal sectors, such as roads, three waters, and electricity. Let $Y_i \geq 0$ denote the length of infrastructure in grid cell i. We model Y_i as the outcome of two statistical processes, which we distinguish between using an indicator variable Z_i:

$$Z_i = \begin{cases} 0, & \text{if } Y_i = 0, \\ 1, & \text{if } Y_i > 0. \end{cases}$$

$Z_i = 0$ denotes infrastructure does not exist in grid cell i, that is, $Y_i = 0$ whereas $Z_i = 1$ denotes grid cells with strictly positive levels of infrastructure, that is, $Y_i > 0$. We assume that Z_i follows a Bernoulli

[8] For example, on average children aged 5-14 years will use education infrastructure more intensively than those aged 15-19 years, as some people in the latter age band will have finished secondary school. For this reason, people aged 5-14 years have a weight of 98.7% whereas those aged 15-19 years have a weight of 56%. These weights were provided by the Commission.

distribution, $Z_i \sim$ Bernoulli(π_i), where $\pi_i = \Pr(Y_i > 0)$ defines the probability of non-zero infrastructure. In these grid cells, we assume $Y_i > 0$ follows a Gamma distribution, $Y_i \sim$ Gamma(μ_i, κ_i).

Pulling these two statistical processes together yields the full hurdle–Gamma model:

$$Y_i \sim \begin{cases} 0, & \text{with probability } 1 - \pi_i, \\ \text{Gamma}(\mu_i, \kappa_i), & \text{with probability } \pi_i. \end{cases}$$

In this model, $1 - \pi_i \in (0, 1)$ defines the hurdle probability, whereas $\mu_i > 0$ and $\kappa_i > 0$ define the mean and shape parameters of the Gamma distribution, respectively. We then specify three linear models:

$$\operatorname{logit}(\pi_i) = \alpha_0 + \alpha^P \log P_i + \alpha^{P(1)} \log P_i^{(1)} + u^{\text{hu}}_{\text{TA}(i)} + u^{\text{hu}}_{\text{TA:SA2}(i)} + u^{\text{hu}}_{\text{TA:SA2:SA1}(i)}, \tag{1}$$

$$\log(\mu_i) = \beta_0 + \beta^P \log P_i + u^{\text{ga}}_{\text{TA}(i)} + u^{\text{ga}}_{\text{TA:SA2}(i)} + u^{\text{ga}}_{\text{TA:SA2:SA1}(i)}, \tag{2}$$

$$\log(\kappa_i) = \zeta_0 + u^{\text{sh}}_{\text{TA}(i)} + u^{\text{sh}}_{\text{TA:SA2}(i)} + u^{\text{sh}}_{\text{TA:SA2:SA1}(i)}. \tag{3}$$

Here, we use a logit link for the Bernoulli distribution and a log link for the mean and shape of the Gamma distribution. The terms $\log P_i$ and $\log P_i^{(1)}$ denote the log of population and the log of the first spatial lag of population, respectively. The former term is of primary interest, whereas the latter term seeks to control for spatial spillovers in infrastructure supply. Because we use a Gamma distribution with a logarithmic link and population enters the linear model in logarithmic form, the parameter β^P can be interpreted as a constant elasticity of infrastructure supply with respect to population.

To account for unobserved heterogeneity while avoiding overfitting, the linear models in Equations (1), (2), and (3) include random intercepts for SA1s, SA2s, and TAs, which we specify as follows.[9]

$$\begin{aligned}
u^{\text{hu}}_{\text{TA}(i)} &\sim \mathcal{N}\big(0,\ \sigma^2_{\text{TA,hu}}\big), \quad u^{\text{hu}}_{\text{TA:SA2}(i)} \sim \mathcal{N}\big(0,\ \sigma^2_{\text{SA2,hu}}\big), \quad u^{\text{hu}}_{\text{TA:SA2:SA1}(i)} \sim \mathcal{N}\big(0,\ \sigma^2_{\text{SA1,hu}}\big), \\
u^{\text{ga}}_{\text{TA}(i)} &\sim \mathcal{N}\big(0,\ \sigma^2_{\text{TA,ga}}\big), \quad u^{\text{ga}}_{\text{TA:SA2}(i)} \sim \mathcal{N}\big(0,\ \sigma^2_{\text{SA2,ga}}\big), \quad u^{\text{ga}}_{\text{TA:SA2:SA1}(i)} \sim \mathcal{N}\big(0,\ \sigma^2_{\text{SA1,ga}}\big), \\
u^{\text{sh}}_{\text{TA}(i)} &\sim \mathcal{N}\big(0,\ \sigma^2_{\text{TA,sh}}\big), \quad u^{\text{sh}}_{\text{TA:SA2}(i)} \sim \mathcal{N}\big(0,\ \sigma^2_{\text{SA2,sh}}\big), \quad u^{\text{sh}}_{\text{TA:SA2:SA1}(i)} \sim \mathcal{N}\big(0,\ \sigma^2_{\text{SA1,sh}}\big).
\end{aligned}$$

This specification of random intercepts explicitly allows for the hierarchical nested structure of spatial units, whereby SA1s are nested within SA2s and SA2s are nested within TAs. More formally, we have $\text{TA} \supset \text{SA2} \supset \text{SA1}$. It is straightforward to model the hierarchical nested structure of these different spatial units. Specifically, instead of including random intercepts for each individual type of spatial unit, we instead include individual random intercepts for the highest level (TAs) and then further random intercepts for interaction terms, specifically $\text{TA} : \text{SA2}(i)$ and $\text{TA} : \text{SA2} : \text{SA1}(i)$. Modelling the hierarchical nested structure of spatial units in this way ensures that spatial variation in infrastructure supply is attributed to the appropriate level, which reduces the risk of confounding between levels and increases the interpretability of estimates of group-level variance, or heterogeneity.

Random intercepts, u, help to address the empirical challenges posed by heterogeneity, spillovers, and omitted variables. That said, we are not especially interested in estimates of their associated parameters. Instead, we are primarily interested in the estimated parameters for log population, $\log P_i$, in the hurdle and gamma models, indicated by α^P and β^P, respectively. Together, these two parameters govern the extent to which the population affects the supply of local infrastructure. In all models, α^P and β^P can be interpreted as the effects of population holding constant the area of the 1 km grid cell.

[9] Random intercepts are sometimes referred to as random effects, varying effects, or group effects. In this paper, we treat these terms as synonymous and interchangeable.

VERTICAL INFRASTRUCTURE

Data on infrastructure supply, Y_j for vertical sub-sectors — namely, health and education — was supplied at SA1 level, j, rather than 1km grid cells, i. Due to differences in these spatial units, we made several subtle but important changes to the linear models presented in Equations (1), (2), and (3).

First, unlike 1km grid cells, the area A_j of SA1s varies. For this reason, we include a non-linear and non-parametric control for the area, $s(A_j)$, in the linear models for $\text{logit}(\pi_j)$ and $\log(\mu_j)$. Second, rather than using the population of an SA1, we instead calculate the population within 2km of each SA1 centroid. This ensures that P_j is calculated over a constant area. Third, whereas horizontal sectors use the total population, for vertical sectors we use the age-weighted population, $\log P_j^{(d)}$, which reflects the contribution to infrastructure supply from people of different ages within the population.[10] Fourth, as the population catchment for each SA1 is approximately 12.5km² versus 1km² for grid cells, we drop the spatial lag of population.[11] Finally, because we are now modelling infrastructure supply, Y_j, at the level of SA1s, we include hierarchical random intercepts, u, only for TAs and SA2s.

These changes result in the following specifications for the linear models for vertical sub-sectors:

$$\text{logit}(\pi_j) = \alpha_0 + \alpha^{P(d)} \log P_j^{(d)} + s^{hu}(A_j) + u^{\text{hu}}_{\text{TA}(j)} + u^{\text{hu}}_{\text{TA:SA2}(j)}, \tag{4}$$

$$\log(\mu_j) = \beta_0 + \beta^{P(d)} \log P_j^{(d)} + s^{ga}(A_j) + u^{\text{ga}}_{\text{TA}(j)} + u^{\text{ga}}_{\text{TA:SA2}(j)}, \tag{5}$$

$$\log(\kappa_j) = \zeta_0 + u^{\text{sh}}_{\text{TA}(j)} + u^{\text{sh}}_{\text{TA:SA2}(j)}. \tag{6}$$

Here, we are primarily interested in parameter estimates for age-weighted population, $\alpha^{P(d)}$ and $\beta^{P(d)}$.

2.2.2 Extended specification

The Baseline model specifications are vulnerable to two forms of endogeneity. First, they implicitly assume that causality runs only from population, $\log P$, to infrastructure supply, Y. In practice, the relationship may be bidirectional: infrastructure may respond to population, and population may in turn respond to infrastructure. Second, although the inclusion of spatial effects, u, helps mitigate omitted-variable bias, it does not eliminate it. Unobserved factors may still vary within spatial units in ways that correlate with infrastructure supply, thereby biasing the estimated parameters for population.

For this reason, we also estimate an Extended specification that uses a control function to address endogeneity. In the first stage, we regress our potentially endogenous variable, $\log P$, against a set of exogenous instruments, Z; a set of spatial effects, u^z; and interactions between the two, $Z \circ u^z$. We use two instruments for population: The crow-fly distance to the nearest regional centre and the crow-fly distance to the coast. These instruments are plausibly exogenous because they affect the population but not infrastructure supply directly except via their effects on population. Statistical tests indicate that our two instruments are highly relevant. The residuals of this first stage model, ϵ, capture the *endogenous component* of $\log P$ for each observation, which we wish to control for in our model.

In the second stage, we extend the linear models in Equations (1), (2), (4), and (5) to include the residuals, ϵ, from the first stage of the control function. Compared to instrumented variables, control functions have the advantage of allowing us to allow for non-linear endogenous effects. For this reason, we include the endogenous residuals, ϵ, as the argument to a non-linear, non-parametric function, $s(\epsilon)$.

[10] Demand weights for education and health by five-year age bands were sourced from the Commission.

[11] Intuitively, we expect "network effects" that give rise to spatial correlations in the supply of infrastructure will be stronger for horizontal sectors than vertical sectors.

The Extended specifications also provide us with a simple test of endogeneity. Specifically, if we find that the parameters for $s(\epsilon)$ are statistically significant, then we have evidence of endogeneity. We can also complement this conventional test of endogeneity with an additional test that compares the out-of-sample performance of the two specifications. If the Extended specification has better out-of-sample performance than the Baseline specification, then we have evidence that controlling for endogeneity via $s(\epsilon)$ improves the external validity of the model and strengthens causal inferences.

2.3 Post-processing

Having identified our preferred models, we use them to estimate infrastructure shares in 2023 and 2048. The latter year aligns with those used in Statistics NZ's population projections and is close to the end of the period covered by the Commission's forward guidance.

To do so, first we predict the supply for individual spatial units k in 2023 and 2048, which we denote by $\hat{Y}_k^{2023}$ and $\hat{Y}_k^{2048}$, respectively. Here, k can be grid cells or SA1s. The predicted change in estimated infrastructure supply over this period can then be readily calculated as $\Delta Y_k = \hat{Y}_k^{2048} - \hat{Y}_k^{2023}$.

The second post-processing step varied slightly by sub-sector. For sub-sectors where our data has national coverage, we take as given observed supply in 2023, Y_k^{2023}. Our forecast then simply adds the change in supply, ΔY_k, such that $Y_k^{2048} = Y_k^{2023} + \Delta Y_k$. This approach uses observed outcomes in 2023 as the base to which we add the change in supply between 2023 and 2048 predicted by our models. In contrast, for three waters we do not have national coverage. Here, we instead use the predicted supply in both years, $\hat{Y}_k^{2023}$ and Y_k^{2048}. Simply put, for three waters we use our model to predict both the existing *level* of infrastructure and the *change* in infrastructure from 2023 to 2048.

Finally, infrastructure shares in 2023 and 2048, S_k^{2023} and S_k^{2048} are calculated by dividing by the total amount of infrastructure, $\overline{Y}^{2023} = \sum_k Y_k^{2023}$ and $\overline{Y}^{2048} = \sum_k Y_k^{2048}$. That is, $S_k^{2023} = Y_k^{2023}/\overline{Y}^{2023}$ and $S_k^{2048} = Y_k^{2048}/\overline{Y}^{2048}$. To generate estimated shares, we use the median predictions from our models, which were found to be more stable. For three waters, we apply an additional post-processing step, which combined the results from the three individual sub-models — that is, water supply, waste water, and storm water — into a single composite share for three waters as a whole. To do so, we weight the individual shares for water supply, waste water, and storm water by 0.33, 0.45, and 0.22, respectively, where these weights were derived from asset and expenditure data for Auckland [51], Wellington [52], and Christchurch [53]. Further information on the sources of these weights is provided in Section A.9.

3 Results

We now present our results. First, we summarise the results of our regression models and, second, we present information on model performance, such as observed versus predicted outcomes.

3.1 Regression results

Table 7 presents the estimated parameters for population (log) in the hurdle, α^P, and gamma, β^P models, where the associated standard errors are in parentheses. The Baseline and Extended specifications are identical, except the latter also controls for endogeneity per Section 2.2.2. The parameter in the gamma model, β^P, is equivalent to a (constant) elasticity of supply with respect to population.

Table 7. Estimated parameters for population from the hurdle, α^P, and gamma, β^P components of the models. The latter can be interpreted as the constant elasticity of infrastructure supply with respect to population. The Baseline and Extended specifications are identical except the latter uses a control function to address endogeneity. Standard errors are shown in parentheses, which allow for heteroskedastic variance per SA1, SA2, and local authority.

Sector	**Sub-sector**	**Baseline**		**Extended**	
		Hurdle, α^P	Gamma, β^P	Hurdle, α^P	Gamma, β^P
Roads	Local roads	1.831 (0.013)	0.254 (0.002)	1.061 (0.029)	0.304 (0.003)
	State highways	0.669 (0.012)	0.050 (0.004)	0.291 (0.025)	0.021 (0.005)
Water	Supply	0.821 (0.022)	0.383 (0.006)	1.407 (0.066)	0.469 (0.006)
	Waste	0.849 (0.035)	0.476 (0.007)	1.206 (0.067)	0.533 (0.009)
	Storm	1.033 (0.026)	0.575 (0.010)	1.241 (0.052)	0.775 (0.013)
Electricity	Distribution	1.496 (0.011)	0.218 (0.002)	0.317 (0.030)	0.184 (0.003)
Education	Primary, Secondary	0.117 (0.023)	0.339 (0.025)	0.111 (0.025)	0.410 (0.027)
Health	Hospitals	0.437 (0.065)	0.506 (0.063)	0.417 (0.070)	0.525 (0.066)

We find strong and consistent evidence that population affects infrastructure supply, with all estimates of α^P and β^P positive and statistically significant at conventional levels.

In terms of endogeneity, tests indicate the Extended specification has better out-of-sample performance for all sub-sectors and the estimated parameters for the control function are almost always precisely estimated. As such, we have evidence that endogeneity is present and good reason to choose the Extended specification as our preferred model. Nevertheless, both specifications produce broadly similar parameter estimates and predictions for most sub-sectors. Taken together, this suggests that endogeneity is present, although controlling for it does not significantly affect our results.

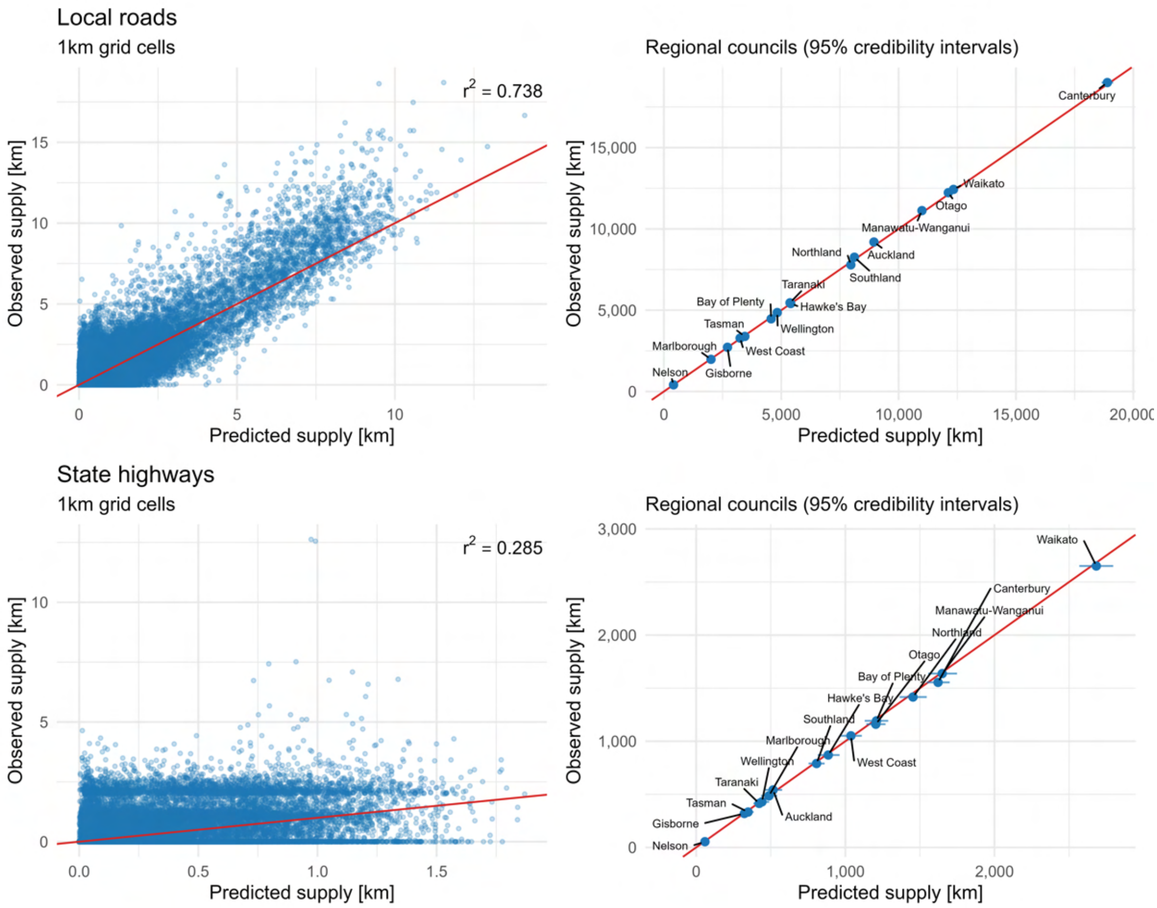
3.2 Predictive performance

The following sub-sections present figures comparing observed versus predicted outcomes for each infrastructure sector at both the local (grid cell or SA1) and regional level.

3.2.1 Roads

In Figure 6, the top and bottom panels compare observed and predicted outcomes for local roads and state highways, respectively, whereas the left and right panels compare observed and predicted outcomes for grid cells and regions, respectively. For local roads, we observe a positive correlation ($r^2 = 0.738$) between outcomes at the local level, and an almost perfect alignment at the regional level.

Figure 6. Roads. The top and bottom panels compare observed and predicted outcomes for local roads and state highways, respectively, whereas the left and right panels compare outcomes at the grid cell and regional levels, respectively. The diagonal lines denotes where observed and predicted outcomes are equal.



Although we find a smaller positive correlation ($r^2 = 0.285$) for state highways at the local level, we again observe an almost perfect alignment between observed and predicted outcomes at the regional level. These results suggest that the supply of local roads and state highways is relatively predictable at the regional level, even if there exists more unexplained variation at the local level.

3.2.2 Water

In Figure 10, the top, middle, and bottom panels compare observed and predicted outcomes for water supply, wastewater, and stormwater, respectively, whereas the left and right panels compare outcomes for grid cells and regions, respectively. Turning first to the grid cells shown in the left panel, we observe strong positive correlations for all three water sub-sectors ($r^2 = 0.929$, $r^2 = 0.942$, and $r^2 = 0.914$, respectively). High values likely reflect the relatively distributed nature of water infrastructure.

At the regional level, again we observe a strong alignment between observed and predicted outcomes. The predicted supply of stormwater is somewhat higher than observed outcomes for larger regions, like Auckland and Canterbury. This may suggest that — compared to the other two water sub-sectors — the supply of stormwater infrastructure responds more strongly to other factors, such as geography, topography, and climate or, alternatively, it might reflect the non-linear effects of population.

Figure 7. Water. The top, middle, and bottom panels compare observed and predicted outcomes for water supply, wastewater, and stormwater, respectively, whereas the left and right panels compare outcomes at the grid cell and regional levels, respectively. The diagonal lines denotes where observed and predicted outcomes are equal.

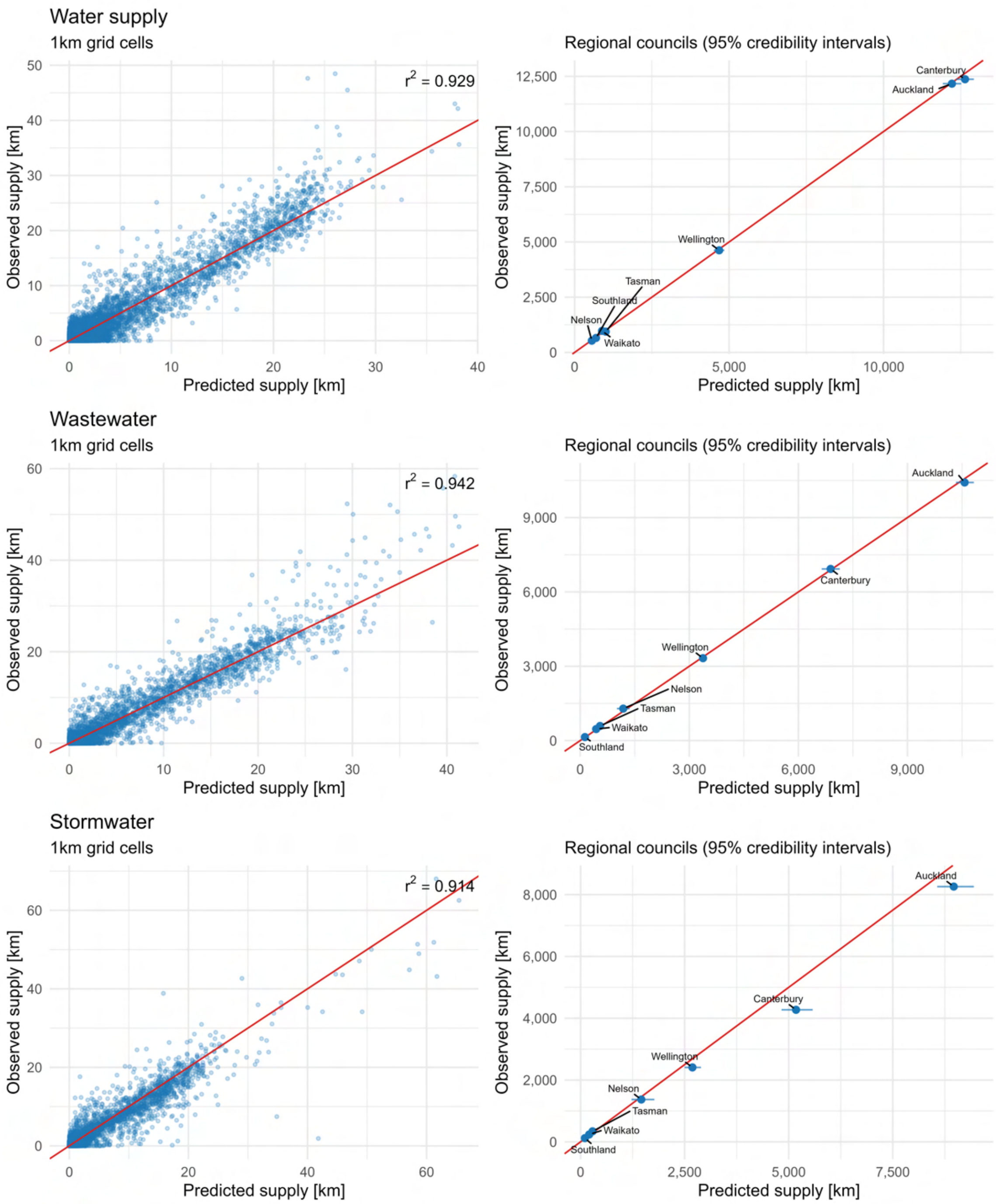
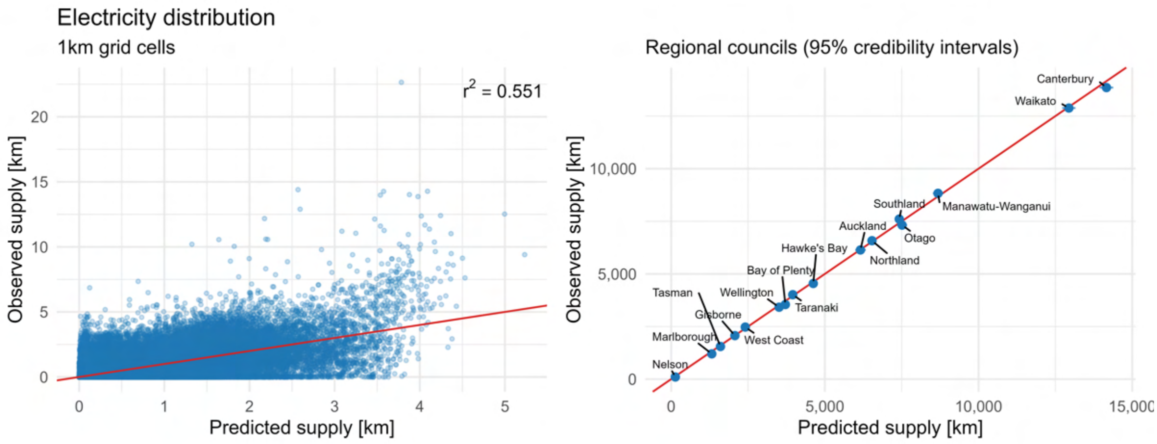


Figure 8. Electricity distribution. The left and right panels compare observed and predicted outcomes at the local and regional levels, respectively. The diagonal line is where observed and predicted outcomes are equal.



3.2.3 Electricity

In Figure 8, the left and right panels compare observed versus predicted outcomes for electricity distribution at the level of grid cells and regions, respectively. When comparing outcomes at the grid cells, we observe a positive correlation ($r^2 = 0.551$). This is lower than those found for water infrastructure and local roads but higher than that for state highways. At the regional level, we again observe strong alignment between observed and predicted outcomes.

3.2.4 Education and health

In Figure 9, the top and bottom panels compare observed versus predicted outcomes for education and health, respectively, for SA1s (left panels) and regions (right panels).

For education, we find a weak correlation ($r^2 = 0.094$) between observed and predicted outcomes at the SA1 level, which is the lowest of the infrastructure sub-sectors that we consider. This implies there is considerable unexplained variation in the local supply of education infrastructure. Nevertheless, at the regional level, we again observe a strong alignment between observed and predicted outcomes.

In contrast, for health we observe a much stronger correlation ($r^2 = 0.36$) at the local level but a relatively weak correlation at the regional level. Indeed, the observed value of health infrastructure in Auckland, Waikato, and the Bay of Plenty is significantly higher than that predicted by the model. We suspect this discrepancy reflects the relatively concentrated and specialised nature of health services that are delivered by hospitals in these regions, which tend to serve a regional if not national role.

Figure 9. Education. The left and right panels compare observed and predicted outcomes at the local and regional levels, respectively. The diagonal line is where observed and predicted outcomes are equal.

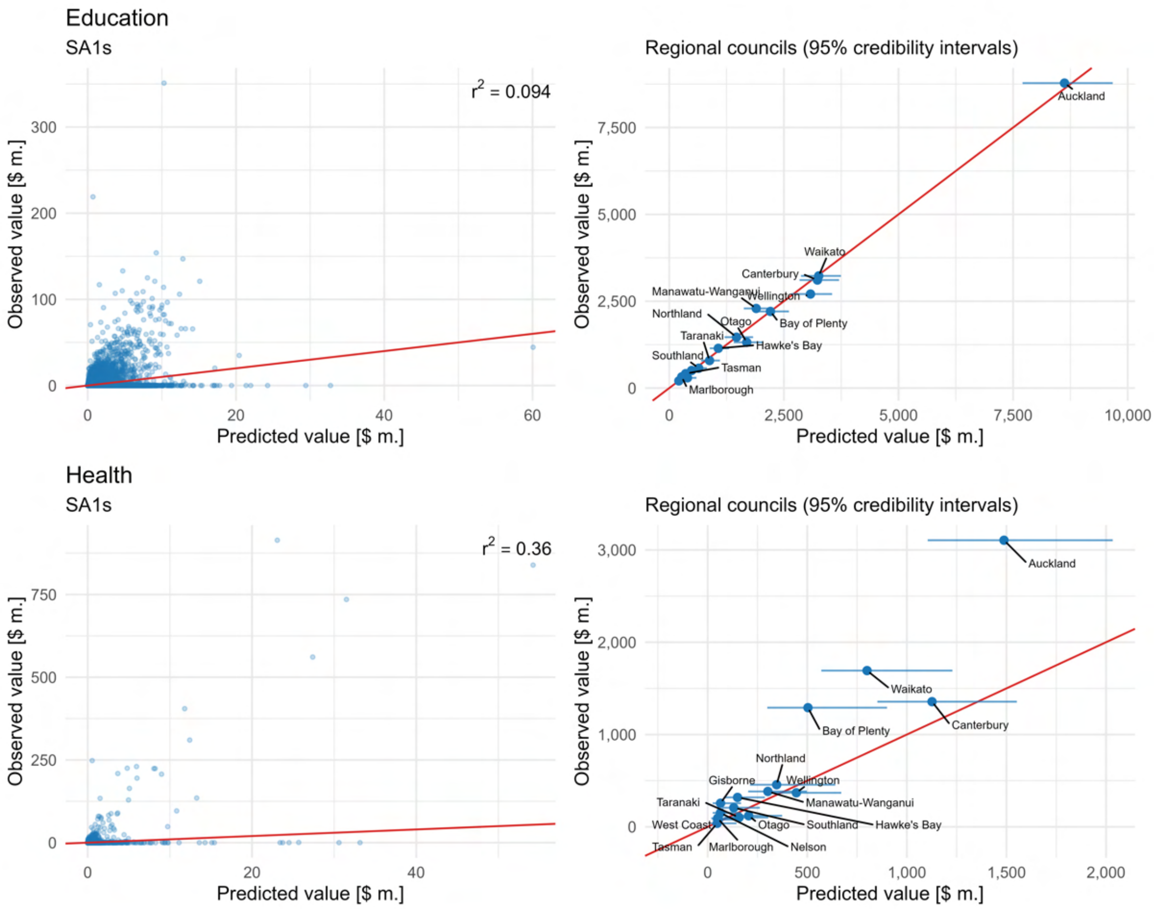
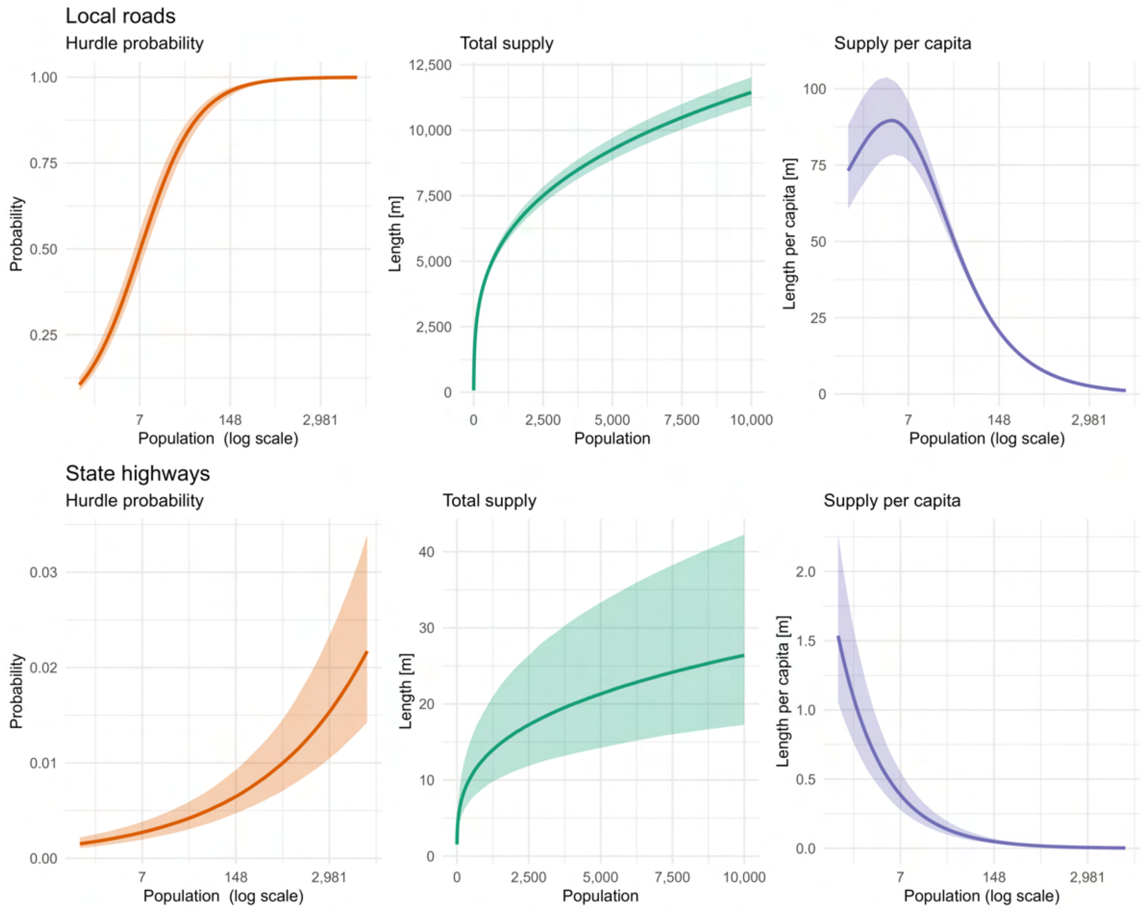


Figure 10. Implications of the Extended model for local roads (top) and state highways (bottom). The left panel shows the probability that infrastructure exists, the middle panel shows total infrastructure supply, and the right panel shows infrastructure supply per capita. Shaded areas denote 95% credibility intervals.



3.3 Model implications

3.3.1 Roads

Figure 10 presents implications of our models for local roads (top) and state highways (bottom). We make three observations. First, the probability that local roads exist in a grid cell approaches 100% at low population levels, whereas the corresponding probability for state highways never exceeds approximately 1.5–3.5%. Second, we find that infrastructure supply increases with population but at a reducing rate, giving rise to a concave shape that is consistent with economies of density. Third, supply per capita decreases with population fairly consistently.

3.3.2 Water

Figure 11 presents implications for water supply (top), waste water (middle), and storm water (bottom). First, the probability that water supply and storm water infrastructure exists approaches 100% at moderate population levels. Second, total supply increases with population but is somewhat less concave than for roads. Third, supply per capita follows a non-monotonic pattern: Initially increasing with population and then flattening off if not falling. This may reflect high fixed costs in supply.

Figure 11. Implications of the Extended model for water supply (top), waste water (middle), and storm water (bottom). The left panel shows the probability that infrastructure exists, the middle panel shows total infrastructure supply, and the right panel shows infrastructure supply per capita. Shaded areas denote 95% credibility intervals.

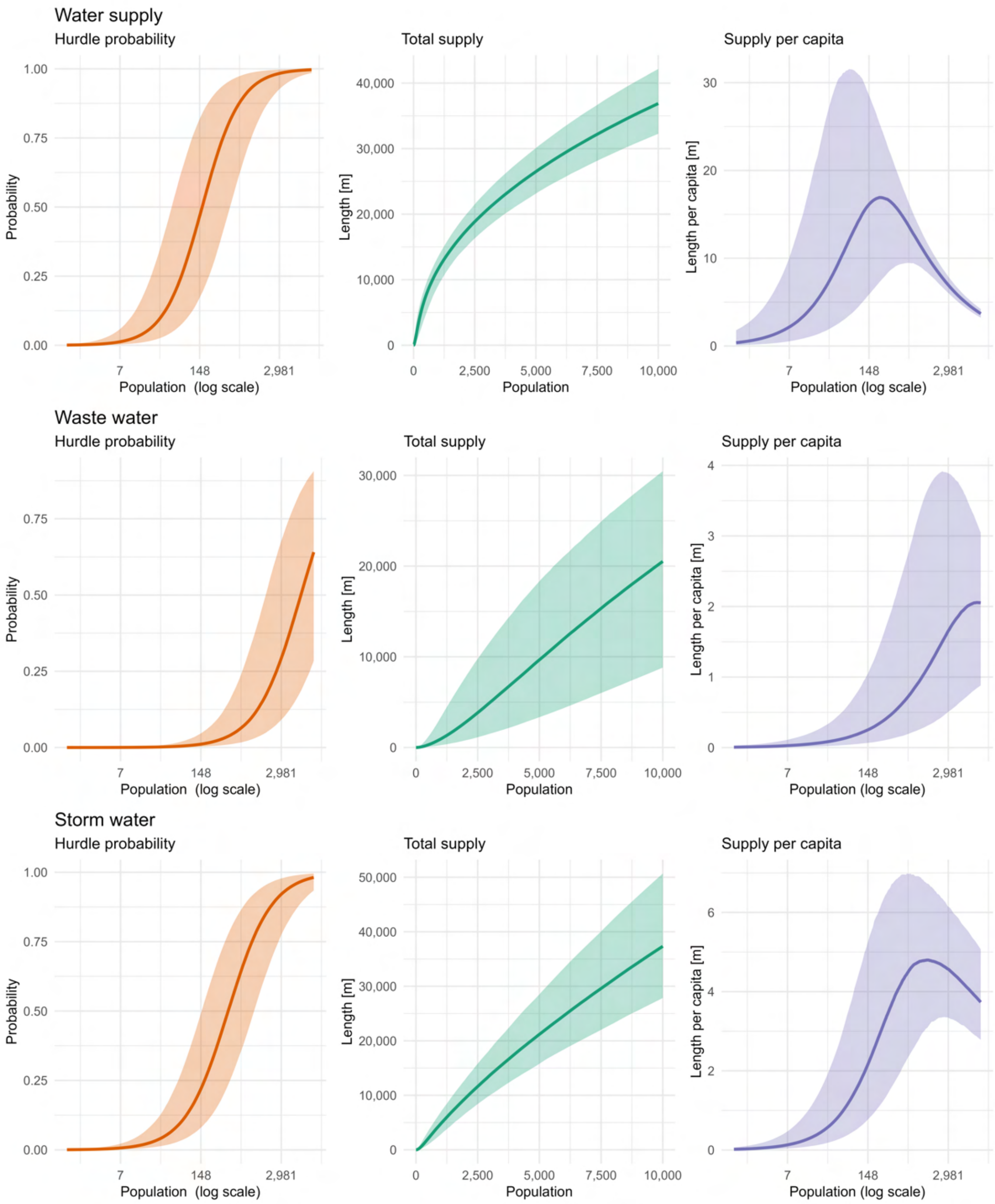
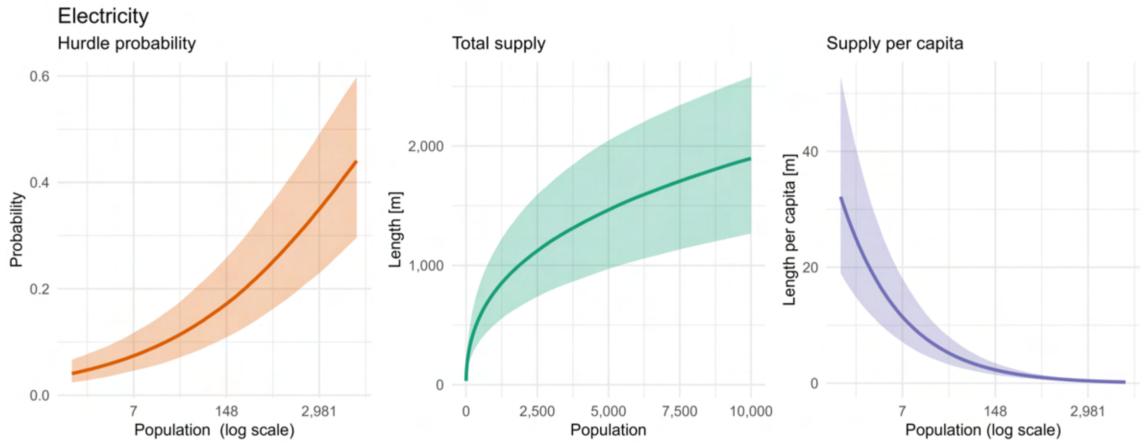


Figure 12. Implications of the Extended model for electricity distribution. The left panel shows the probability that infrastructure exists, the middle panel shows total infrastructure supply, and the right panel shows infrastructure supply per capita. Shaded areas denote 95% credibility intervals.



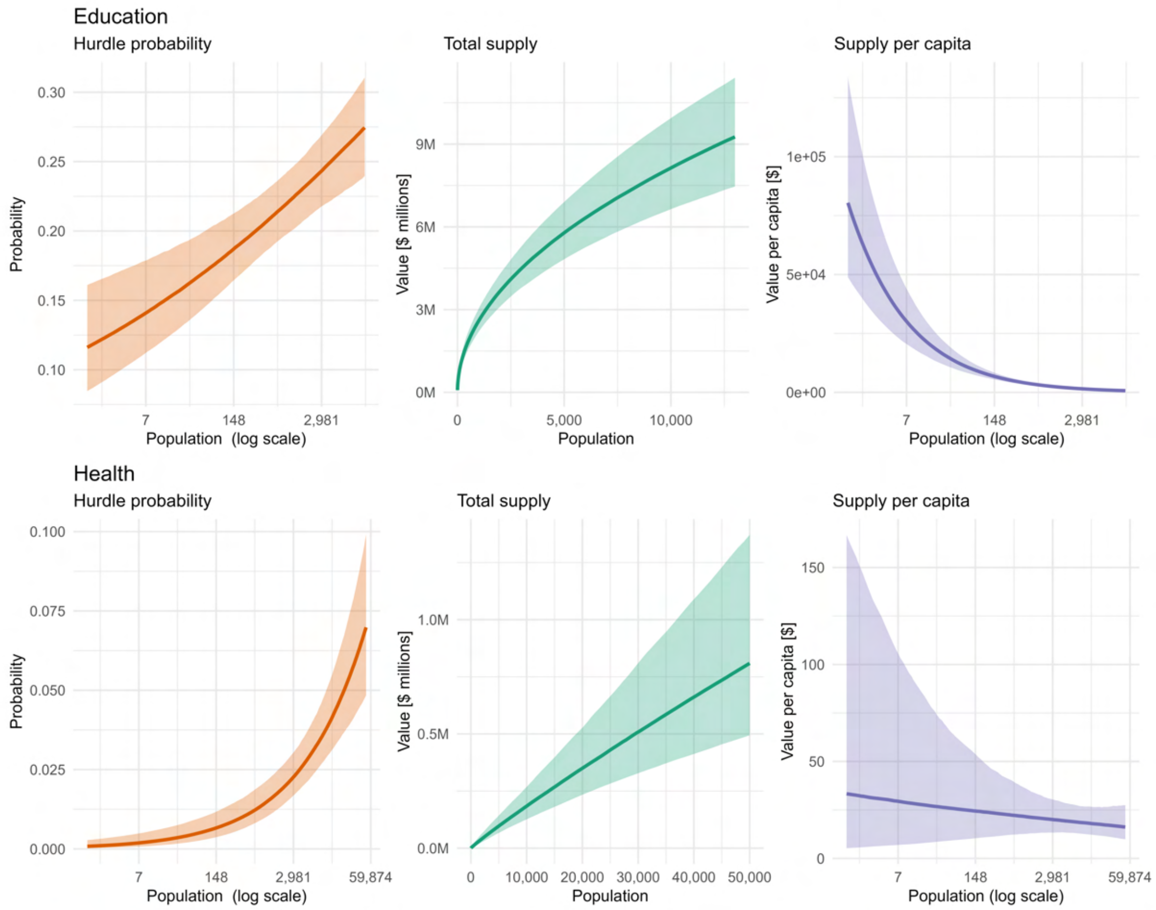
3.3.3 Electricity

Figure 12 presents implications for electricity. First, the probability electricity distribution infrastructure exists in a grid cell approaches 30–60% at high population levels. Second, total supply increases with population and observes a strongly concave shape. Third, supply per capita falls monotonically with population, which is consistent with the relatively low value of the β^p parameter in Table 7 and is consistent with strong economies of density in the supply of infrastructure for electricity distribution.

3.3.4 Education and health

Figure 13 presents implications for education (top) and health (bottom). First, the probability education infrastructure exists in an SA1 approaches 25–30% in SA1s with large population catchments versus only 5–10% for health. Second, total supply increases with the population and follows a concave shape for education but less so for health. Third, supply per capita falls steeply with population for education but not for health, which falls only slowly as the population increases. For health, we also observe considerable uncertainty in the estimates, which is consistent with our earlier results in Figure 8.

Figure 13. Implications of the Extended model for education (top) and health (bottom). The left panel shows the probability that infrastructure exists, the middle panel shows total infrastructure supply, and the right panel shows infrastructure supply per capita. Shaded areas denote 95% credibility intervals.



3.4 Sensitivity tests

We tested various alternatives to the Baseline and Extended specifications discussed above, including:

- *Additional explanatory variables*, such as employment and income. We did not find evidence that including income improved model performance, although we did find evidence for employment. However, we decided not to include employment in the models because we could not locate detailed sub-national projections. Hence, we opted for the more parsimonious approach.
- *Finer spatial resolutions*, specifically 250m and 500m grid cells (for horizontal sub-sectors) and SA2s (for vertical sub-sectors. We decided to use 1km grid cells for horizontal sub-sectors because it provided national coverage and was computationally efficient. For vertical sub-sectors, we used SA1s as they provided the most detailed spatial resolution.
- *Different spatial effects*, for the individual spatial effects, u, we estimated models that used Student's-t rather than Gaussian distributions. The former distribution has fatter tails and implies that random intercepts, u, will shrink less towards the mean, which may be beneficial in situations where extreme values are present.
- *Non-linear population effects*, in our model for health, we tested a variant that allowed population to have non-linear effects on infrastructure supply. Results suggested this model performed better than the simpler model with a constant elasticity, although the predictive performance of the model was similae to the basic model.
- *Linear control functions*, rather than the non-parametric, non-linear form, $s(\epsilon)$, that we adopt in the Extended specification. We also tested alternative instruments for population.

In general, few of the alternative specifications that we tested performed significantly better than our chosen models. The only alternative that showed considerable promise was models that included employment. We suggest further research could seek to identify detailed sub-national projections for employment such that the latter could then be included in our models.

3.5 Regional shares

We used our models to predict shares for each infrastructure sub-sector in 2023 and 2048, per the methodology previously described in Section 2.3. Appendix B presents maps of our results at the regional level, where we show shares in 2023 and 2048 as well as the change in share in this period.

The most notable feature of these maps is the relatively high degree of persistence in the shares of infrastructure for all sub-sectors, which do not change much over the 25-year period that we analyse. The reason that differences in population growth between regions do not necessarily flow through to differences in shares is likely explained by the fact that the fastest-growing regions, such as Auckland and Canterbury, are already some of the most populous. In these regions, a decent proportion of population growth is projected to happen in areas where infrastructure already exists. As such, growth is more able to leverage existing infrastructure networks, giving rise to economies of density.

Of course, persistency in the estimated shares does not imply that infrastructure networks at the regional level are static and will not change. Instead, even persistent shares of investment can lead to significant changes in the quality and coverage of the underlying infrastructure networks over time. Further information on how the results of this project have been used by the Commission is provided in the latter's technical report on modelling and forecasts for the overall Forward Guidance, while the resulting estimates are detailed in the National Infrastructure Plan itself.

4 Concluding comments

4.1 Discussion

We find that our approach offers a relatively simple way to predict the supply of infrastructure at the local level. More specifically, hurdle-Gamma models seem able to identify the causal effect of population on infrastructure supply for eight relatively heterogeneous sub-sectors and generate reasonably accurate predictions, at least at the regional level and with the possible exception of health. In simple terms, these empirical models appear to combine robust causal identification with decent predictive power, which is a somewhat uncommon but useful combination of attributes.

The low predictive power of our model of health infrastructure is unfortunate but unsurprising. Compared to other sub-sectors, the spatial distribution of health infrastructure is highly heterogeneous, with facilities ranging from small local medical centres to large hospitals. The latter often deliver highly specialised services and play a regional, if not national, function. In this context, we are unsurprised that our models struggle to explain variation in health infrastructure as a function of local population. However, we still identify a robust causal effect of the population on the supply of health infrastructure.

On the other hand, the models for the other seven infrastructure sub-sectors appear to perform extremely well. Our confidence in the models for local roads, state highways, electricity distribution, and education is further enhanced by the national coverage of our data. Complete coverage means these models only operate "out-of-sample" in a temporal sense, in that we use them to predict supply in the future. Provided that we identify a robust causal effect of population on supply, we can be confident that our models will capture changes in infrastructure supply due to changes in the population.

In contrast, our data for three waters do not have national coverage. As such, we must apply the models out-of-sample in both a temporal and a spatial sense. That is, we are using the models to predict supply in the future *and* in areas where we do not observe outcomes. For this reason, we are somewhat less confident in estimates for three waters compared to those sub-sectors for which we have national coverage. A more cautious approach to interpreting and applying our results for three waters is also warranted due to local differences in the way data on infrastructure supply are recorded.

Finally, we see our estimates of local infrastructure shares as a useful complement to — rather than a substitute for — more conventional sector-level strategies developed by government agencies. Whereas our approach has the benefit of being based on empirical evidence, sector-level strategies may be better positioned to respond to other considerations, such as the effects of new technologies.

4.2 Other applications

The primary purpose of this work was to develop methods for estimating the supply of infrastructure at the local level to support the work of the Commission. That said, the potential exists for these models to be usefully applied in other contexts, including but not limited to:

- *Land use planning*, such as that undertaken by local authorities. These models could, for example, be used to understand the relative scale of infrastructure supply in different land use scenarios.
- *Infrastructure planning*, such as costs for new development. These models could be used, for example, to inform estimates of the share of infrastructure costs that are attributable to growth.
- *Economic analysis*, for example, by embedding these infrastructure supply models within quantitative spatial models that can then be used to evaluate the economic effects of policies.

Although we consider these models to have the potential to be usefully applied elsewhere, they are subject to some limitations and possible extensions, as noted in Sections 1.3 and 4.3, respectively.

4.3 Possible extensions

Notwithstanding our promising results, we consider it important to emphasise that our approach is relatively novel. As such, we suggest that the results are interpreted judiciously and applied carefully, at least until further research can confirm the merits and implications of our approach.

An obvious next step, for example, would be to perform a detailed systematic review of the literature on economies of density and scale in infrastructure supply. Although such a review was outside the scope of this research, our preliminary review of the literature identified several studies that might act as useful external reference points to which our results can be compared [see, for example, 3, 4, 5, 6, 7]. However, drawing effective comparisons with the external literature will require care in reconciling differences in the spatial scale of analyses. Whereas we analyse infrastructure supply at the level of grid cells ($n = 270,845$) or SA1s ($n = 29,603$), for example, most studies appear to use much larger spatial units, such as whole networks or entire regions. In this context, a detailed systematic review seems to present a useful albeit not necessarily straightforward extension of this work.

Other potential directions for further research include, but are not necessarily limited to:

- For health and education, adopting a randomly-generated grid rather than SA1s
- For horizontal sub-sectors, using data on capital values rather than infrastructure length
- Grounding our empirical models in theoretical economic models of infrastructure supply
- Investigating other plausibly exogenous instruments, such as vertical elevation
- Including additional explanatory variables, such as topographical complexity
- Incorporating updated population projections from Statistics NZ
- Modelling the change in infrastructure supply over time
- Applying models to other infrastructure sectors

In our view, the first two items on this list are perhaps the highest priority. More specifically, undertaking further work to address these two items would allow us to estimate models that use consistent spatial units and supply measures across all sub-sectors. This consistency would reduce the vulnerability of our results to empirical problems, such as the modifiable area unit problem ("MAUP") and measurement error in the outcome variable, both of which could be potential sources of bias and/or imprecision. Although we do use a control function in an effort to address potential sources of endogeneity, we suggest that it is better to adopt methods that avoid these problems more directly.

Looking further down the list, the fourth and fifth items seem to present straightforward ways to strengthen our identification of causal effects and increase confidence in these results. Although item seven on the list would also strengthen identification by enabling us to estimate the models using panel data, collecting changes in infrastructure supply over time may be more resource intensive. Nevertheless, we suspect that these data could be useful for many aspects of the Commission's work, and recommend that it be considered as an option for further work. The experience in Canterbury following the 2010 earthquake, for example, may provide a useful quasi-experimental context in which to analyse the effects of population on infrastructure supply over time.

4.4 Conclusions

Notwithstanding these opportunities for possible extensions, we consider that our approach, even as it currently stands, provides a useful starting point for thinking about the future evolution of infrastructure supply at the local level. To the best of our knowledge, the detailed approach used in this work is relatively novel. And, as is clear from our results, it is possible to estimate models a detailed spatial scales that nonetheless generate accurate aggregate predictions, for example at the regional level.

A Data appendix

This appendix documents the datasets, spatial transformations, and allocation methods used to construct the dependent, explanatory, and instrumental variables used in the analysis. All spatial operations follow a hierarchy linking SA1, SA2, and regular grid geometries. Population, infrastructure supply, and ancillary variables are processed separately and reconciled into a common geometry-specific master dataset.

A.1 Data sources

Primary data sources are described in the report body and summarised in Table 4. This appendix focuses on data processing and variable construction.

A.2 Data processing

The heterogeneity of data sources and the spatial and temporal structure required for estimation and projection necessitate a multi-stage preprocessing pipeline. Each master dataset is constructed deterministically from a configuration file that defines its base geometry and associated processing rules.

The pipeline follows the structure below:

1. Configuration and geometry hierarchy
2. Infrastructure supply variables
3. Instrument construction
4. Population variables
5. Projected explanatory variables
6. Spatial lags and catchments

Design principles The pipeline is designed around two principles:

- **Spatial consistency:** variables are generally not transferred between incompatible geometries.
- **Reproducibility:** all processing steps are controlled by explicit configuration files. Each master dataset has a single base geometry, which defines the unit of observation and governs all subsequent aggregation and allocation steps.

A.3 Configuration files and geometry hierarchy

Each model instance is defined by a configuration file specifying its base geometry (e.g. 250 m grid, SA1, SA2). The base geometry remains fixed throughout estimation and prediction.

Two geometry families are used:

- **Grid geometries**: 250 m, 500 m, and 1 km grids, which nest exactly.
- **Administrative geometries**: SA1, SA2, Territorial Authority, and Regional Council boundaries, which also nest exactly (with some minor exceptions).

Each master dataset is based on a single geometry from one family. Where necessary, higher-level geometries are attached via deterministic joins. For grid-based datasets, administrative boundaries are attached using centroid assignment.

The base geometry determines both the spatial aggregation of variables and the appropriate data sources. For example, grid-based models use grid population estimates, while administrative-boundary models use Census population data.

A.4 Infrastructure supply variables

Infrastructure supply variables are harmonised to the base geometry and fall into two categories: horizontal network infrastructure and vertical social infrastructure.

A.4.1 Horizontal infrastructure

Horizontal infrastructure supply is measured as total network length (metres) within each spatial unit. Although some datasets include capacity or condition information, these variables are incomplete or inconsistent across regions and are not used.

Roads The LINZ Road Addressing dataset provides complete national coverage of road geometries. Two supply measures are derived:

- *Local roads*: all segments not identified as State Highways.
- *State Highways*: extracted using a name-based filter.

Electricity distribution Electricity line data are obtained from OpenInfrastructureMap/InfraGeomatics. To approximate distribution networks, the analysis restricts attention to lines below approximately 11 kV to exclude transmission infrastructure. Supply is measured using line length only.

Three-waters Three-waters networks data are available only where councils provide public API access and coverage is nationally incomplete. Coverage includes Auckland, Wellington, Canterbury regions, and some rural districts, but some mid-sized centres remain unavailable.

3-waters network datasets differ in:

- the set of recorded variables (diameter, material, condition, etc.),
- geometric detail (e.g. whether property laterals are included),
- the extent and type of assets recorded.

Networks are processed separately for water supply, stormwater, and wastewater. To ensure national comparability, only line length is used. Regions without coverage are assigned missing values, and regional random effects absorb cross-area data quality differences.

A.4.2 Vertical infrastructure

Vertical infrastructure comprises education (primary and secondary) and health (hospital) facilities. Capital value estimates are supplied by ESNZ[44] and are available at both SA1 and grid resolutions.

Tertiary education assets are excluded by identifying spatial units intersecting known tertiary footprints obtained from the OpenStreetMaps[54] and setting corresponding values to zero. Capital value is used as the supply measure, as it provides a consistent proxy for physical capacity.

A.5 Population variables

Population is used as the primary explanatory variable.

A.5.1 Population variables: scope and geometry alignment

Population variables are generally used at their native spatial resolution and are not transferred between grid-based and administrative geometries. The exception to this is the calculation of projected grid population values which are adjusted using the same additive allocation method derived from the SA2 population projection datasets.

When the base geometry is a grid (250 m, 500 m, or 1 km), population is taken directly from Stats NZ grid population estimates for the current period (2023). These values are joined one-to-one and are used primarily in horizontal infrastructure models. Grid population estimates are not age-disaggregated and cannot be demand weighted.

When the base geometry is administrative (SA1 or SA2), population is taken directly from the 2018 Census at SA1 resolution, including full five-year age-band detail. These values are used for vertical infrastructure modelling. Grid population estimates are not used in administrative-boundary models.

A.5.2 Population projections and spatial allocation

Population projections are constructed using an additive allocation method that preserves projected totals at the SA2 level. For each SA2 k, age band j, year t, and scenario s, the projected change is

$$\Delta_{k,j,t,s} = \mathrm{pop}^{\mathrm{proj}}_{k,j,t,s} - \mathrm{pop}^{\mathrm{base}}_{k,j}.$$

When modelling at SA1 resolution, this change is allocated proportionally using base-year Census shares. For SA1 $i \in k$,

$$s_{i,j} = \frac{\mathrm{pop}^{\mathrm{base}}_{i,j}}{\sum_{i' \in k} \mathrm{pop}^{\mathrm{base}}_{i',j}}, \qquad \Delta_{i,j,t,s} = s_{i,j} \cdot \Delta_{k,j,t,s}.$$

Projected SA1 population is then

$$\mathrm{pop}^{\mathrm{proj}}_{i,j,t,s} = \mathrm{pop}^{\mathrm{base}}_{i,j} + \Delta_{i,j,t,s}.$$

When modelling at SA2 resolution, projected changes are applied directly without further spatial disaggregation. Grid-based projections use a simplified version of this method applied to total population only, as grid data lack age-band detail.

A.5.3 Demand weighting

Demand-weighted population is constructed for vertical infrastructure models by applying age-specific weights to projected population counts. Education and health infrastructure use distinct weight vectors reflecting differing age profiles of demand. Horizontal infrastructure models do not use age weighting. Age weights for health and education demand were obtained from the national forward guidance model.

A.6 Instruments

A.6.1 Distance-based instruments

Two distance-based instruments are constructed.

First, distances to major regional population centres are calculated. A set of 16 point locations representing the dominant urban centre in each region is defined internally in the configuration files. Points are chosen to approximate population-weighted urban centres rather than administrative centroids. For each spatial unit, Euclidean distances from its centroid to each centre are computed.

Second, a distance-to-coast instrument is defined as the shortest distance from the centroid of each spatial unit to the New Zealand coastline boundary. Units with centroids outside the landmass polygon are assigned a distance of zero.

All distances are computed in a projected coordinate reference system (NZTM2000) and measured in metres.

A.7 Spatial lags and catchments

Spatial aggregation is performed only after all base and projected demand variables are constructed. Spatial lags and catchments aggregate existing values; they do not reallocate population.

A.7.1 Neighbour-based spatial aggregation

For administrative geometries, spatial lags are constructed using first-order contiguity. For unit i with neighbours $\mathcal{N}(i)$,

$$X_k^{\text{lag}} = \sum_{j \in \mathcal{N}(i)} X_j,$$

and an equivalent area-weighted mean. This avoids imposing artificial grid structure on irregular polygons.

Grid geometries are treated similarly and have spatial lags of first and second order. Neighbour contiguity in regular grids is equivalent to the Moore neighborhood of the order of the lag degree.

A.7.2 Distance-based catchments

For vertical infrastructure models, demand is aggregated within circular catchments defined by centroid-to-centroid distances. For radius r,

$$X_k^{(r)} = \sum_{j: d(i,j) \leq r} X_j.$$

Catchments are computed for multiple radii specified in configuration files (e.g. 2 km, 5 km, 10 km, 25 km, 50 km) and are evaluated separately for each projection year and scenario.

A.8 Final master datasets

Each master dataset contains a single observation per base spatial unit, with columns for infrastructure supply, population demand, instruments, spatial lags, and catchments. Different model specifications subset from a common master file, supporting consistent estimation, prediction, and robustness checks.

A.9 Sources for 3-waters allocation

Section 2.3 describes a post-processing step that recombines the 3-waters sub-models into a single value for predicted 3-waters expenditures. The weights of each sub-model were determined using data obtained from the financial statements for Auckland Council[51], Wellington City Council[52], and Christchurch City Council[53].

Financial reports were used for the following date ranges for each Council:

- Auckland Council: 2016 - 2024
- Wellington City Council: 2014 - 2024
- Christchurch City Council: 2018 - 2024

Where possible, both the share of asset book value and annual CAPEX of each infrastructure class (water, stormwater, wastewater) were calculated.

These values are summarised in table 8

Table 8. Shares of book value and annual CAPEX by network type and council

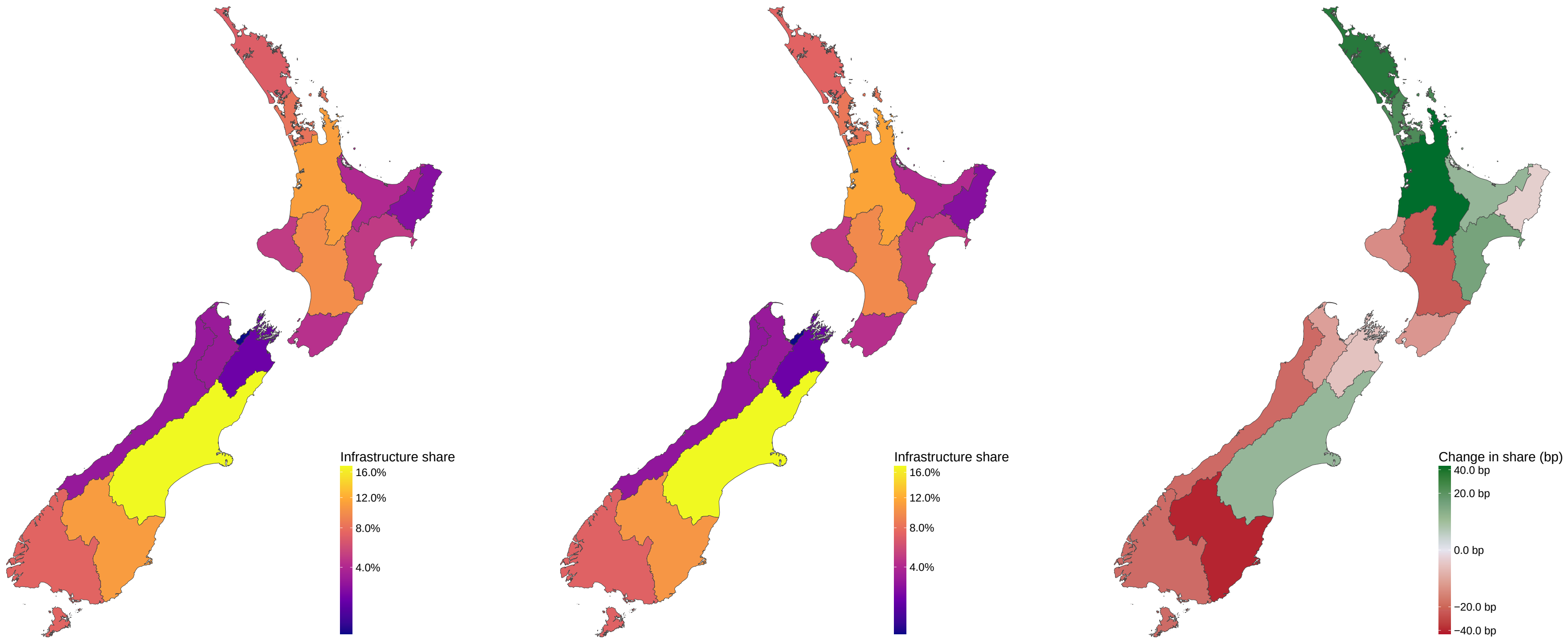
<table>
<tr><th>Council</th><th>Date range</th><th>Measure</th><th>Water</th><th>Wastewater</th><th>Stormwater</th></tr>
<tr><td rowspan="2">Auckland</td><td rowspan="2">2016–2024</td><td>Share book value</td><td colspan="2">62%</td><td>38%</td></tr>
<tr><td>Share annual CAPEX</td><td>29%</td><td>53%</td><td>18%</td></tr>
<tr><td rowspan="2">Wellington</td><td rowspan="2">2014–2024</td><td>Share book value</td><td>–</td><td>–</td><td>–</td></tr>
<tr><td>Share annual CAPEX</td><td>40%</td><td>43%</td><td>16%</td></tr>
<tr><td rowspan="2">Christchurch</td><td rowspan="2">2018–2024</td><td>Share book value</td><td>29%</td><td>47%</td><td>24%</td></tr>
<tr><td>Share annual CAPEX</td><td>33%</td><td>50%</td><td>17%</td></tr>
<tr><td rowspan="2">Overall</td><td rowspan="2">–</td><td>Share book value</td><td>30%</td><td>39%</td><td>31%</td></tr>
<tr><td>Share annual CAPEX</td><td>34%</td><td>49%</td><td>17%</td></tr>
</table>

Note that Wellington City Council did not report shares of book value by water infrastructure class. And that Auckland Council did not differentiate between water and waste water classes when reporting book value shares.

The final weights of 0.33 water, 0.45 waste water, and 0.22 storm water given in section 2.3 reflect an approximation of these shares. They are weighted towards the annual CAPEX shares rather than the book value shares to account for differing depreciation rates across classes of water infrastructure.

B Results appendix

B.1 Local roads



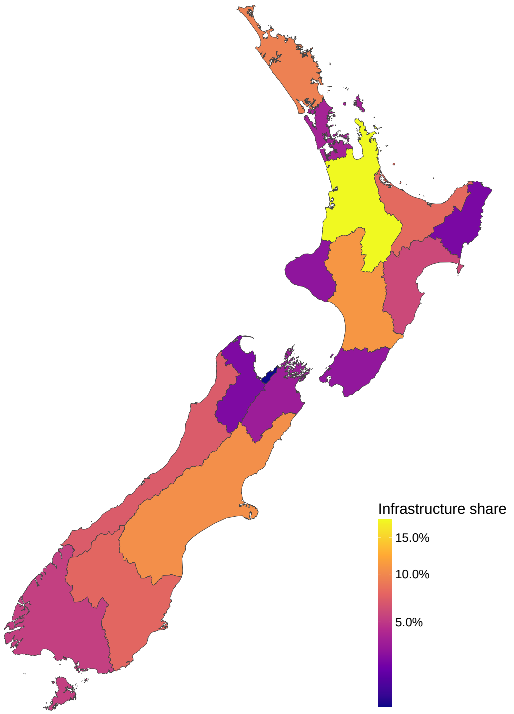
(a) 2023 share

(b) 2048 share

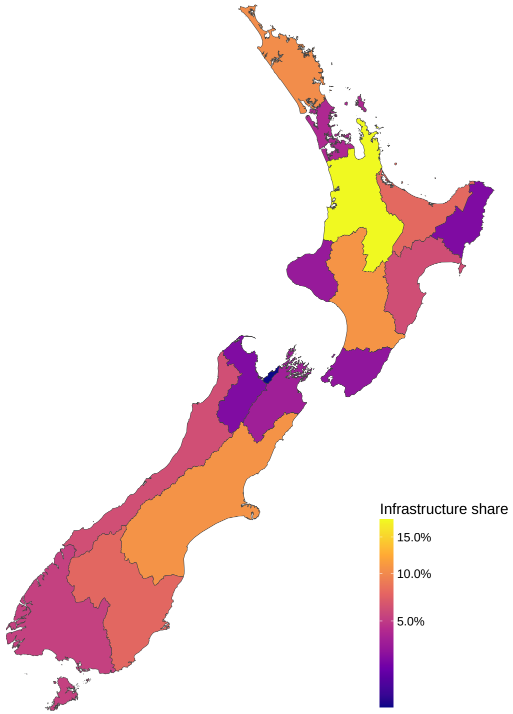
(c) Change, 2023–2048

Figure 14. Maps showing the share of modelled local road infrastructure expenditure for 2023, 2048, and the change in shares between 2023 and 2048.

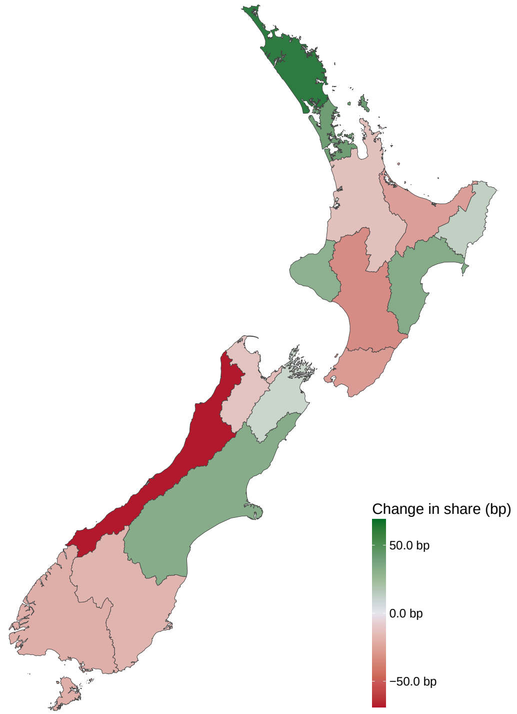
B.2 State highways



(a) 2023 share



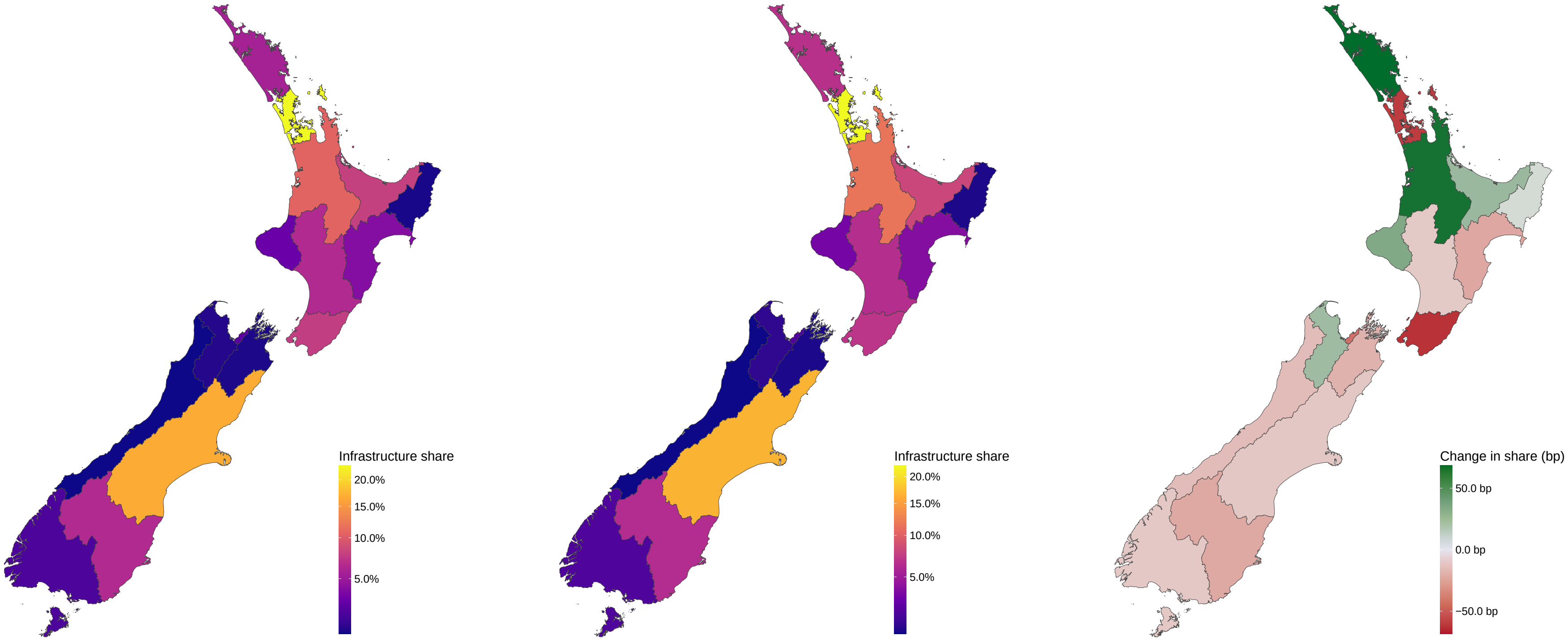
(b) 2048 share



(c) Change, 2023–2048

Figure 15. Maps showing the share of modelled state highway infrastructure expenditure for 2023, 2048, and the change in shares between 2023 and 2048.

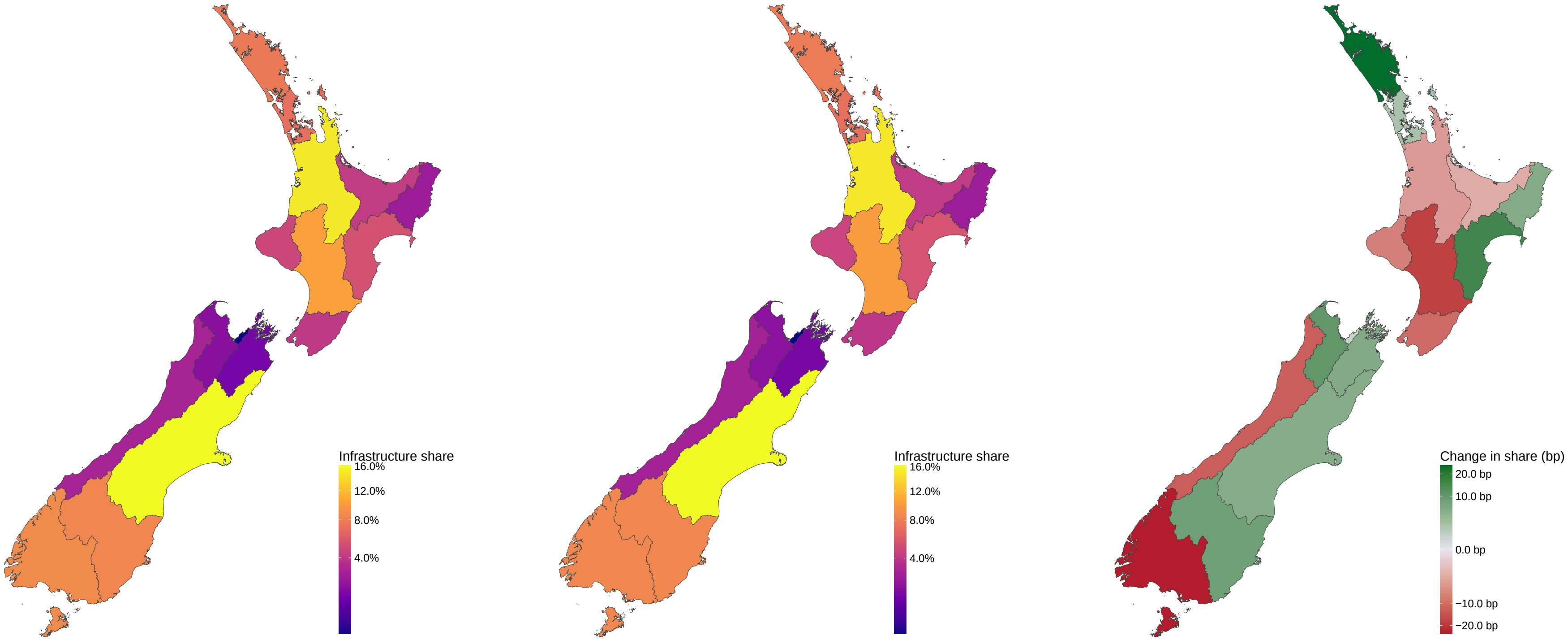
B.3 3-waters



(a) 2023 share *(b) 2048 share* *(c) Change, 2023–2048*

Figure 16. Maps showing the share of modelled 3-waters infrastructure expenditure for 2023, 2048, and the change in shares between 2023 and 2048.

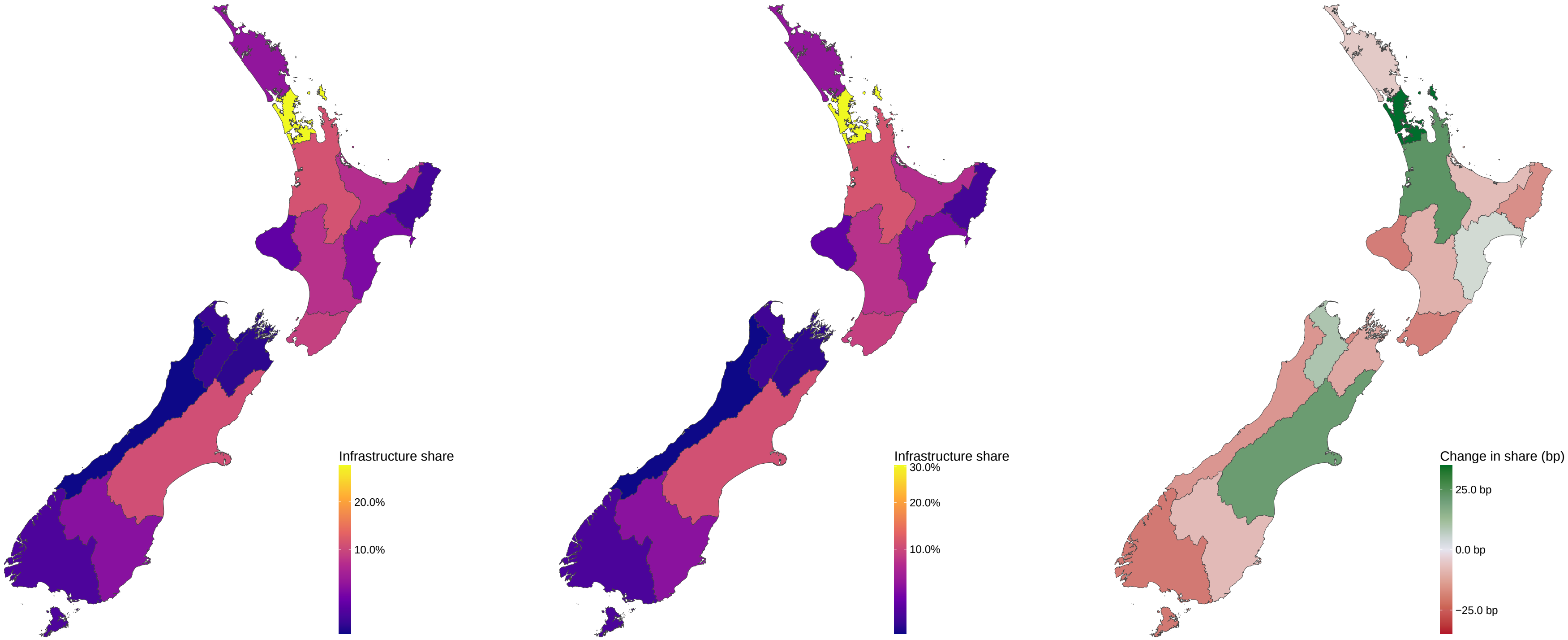
B.4 Electricity distribution



(a) 2023 share *(b) 2048 share* *(c) Change, 2023–2048*

Figure 17. Maps showing the share of modelled electricity distribution infrastructure expenditure for 2023, 2048, and the change in shares between 2023 and 2048.

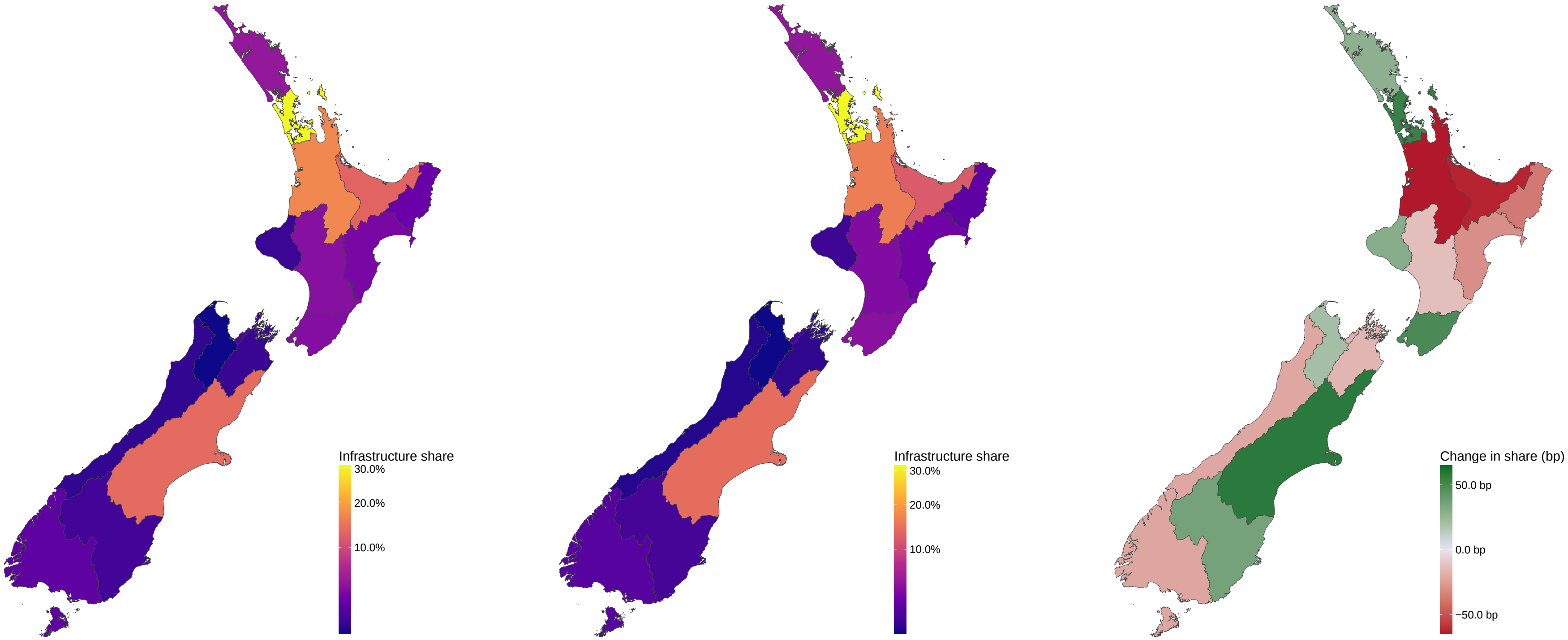
B.5 Education



(a) 2023 share *(b) 2048 share* *(c) Change, 2023–2048*

Figure 18. Maps showing the share of modelled education infrastructure expenditure for 2023, 2048, and the change in shares between 2023 and 2048.

B.6 Health



(a) 2023 share *(b) 2048 share* *(c) Change, 2023–2048*

Figure 19. Maps showing the share of modelled health infrastructure expenditure for 2023, 2048, and the change in shares between 2023 and 2048.

References

[1] New Zealand Infrastructure Commission / Te Waihanga. The infrastructure needs analysis forecast: Results and modelling technical report. Technical report, New Zealand Infrastructure Commission / Te Waihanga, Wellington, New Zealand, 2025. URL https://media.umbraco.io/te-waihanga-30-year-strategy/j54irnae/ina-model-technical-report-for-draft-plan.pdf. Modelling technical report supporting the draft National Infrastructure Plan.

[2] New Zealand Infrastructure Commission / Te Waihanga. Paying it forward: Understanding our long-term infrastructure needs. Research insights series, New Zealand Infrastructure Commission / Te Waihanga, Wellington, New Zealand, 2024. URL https://media.umbraco.io/te-waihanga-30-year-strategy/43ikcme0/paying-it-forward-understanding-our-long-term-infrastructure-needs.pdf. Research report.

[3] K. Chakraborty, B. Biswas, and WC. Lewis. Economies of scale in public education: an econometric analysis. *Contemporary Economic Policy*, 18(2):238–247, 2000. doi: https://doi.org/10.1111/j.1465-7287.2000.tb00021.x.

[4] Cameron Speir and Kurt Stephenson. Does sprawl cost us all?: Isolating the effects of housing patterns on public water and sewer costs. *Journal of the American Planning Association*, 68(1):56–70, 2002. doi: 10.1080/01944360208977191. URL https://doi.org/10.1080/01944360208977191.

[5] Céline Nauges and Caroline van den Berg. Economies of density, scale and scope in the water supply and sewerage sector: a study of four developing and transition economies. *Journal of Regulatory Economics*, 34(2):144–163, 2008. ISSN 1573-0468. doi: 10.1007/s11149-008-9063-2. URL https://doi.org/10.1007/s11149-008-9063-2.

[6] Ángel M. Prieto, José L. Zofío, and Inmaculada Álvarez. Cost economies, urban patterns and population density: The case of public infrastructure for basic utilities*. *Papers in Regional Science*, 94(4):795–817, 2015. ISSN 1056-8190. doi: https://doi.org/10.1111/pirs.12096. URL https://www.sciencedirect.com/science/article/pii/S1056819023017815.

[7] Antony Andrews. On the productivity of new zealand state and state-integrated schools: Is bigger better? *Public Administration Quarterly*, 44(4):612–635, 2020. doi: 10.37808/paq.44.4.5. URL https://doi.org/10.37808/paq.44.4.5.

[8] StatsNZ, Tatauranga Aotearoa. Statistical area 1 2018 clipped (generalised), 2020. URL https://datafinder.stats.govt.nz/layer/92211-statistical-area-1-2018-clipped-generalised/. Retrieved from StatsNZ Geographic Data Service.

[9] StatsNZ, Tatauranga Aotearoa. Statistical area 2 2018 clipped (generalised), 2021. URL https://datafinder.stats.govt.nz/layer/92213-statistical-area-2-2018-clipped-generalised/. Retrieved from StatsNZ Geographic Data Service.

[10] StatsNZ, Tatauranga Aotearoa. Territorial authority 2018 clipped (generalised), 2021. URL https://datafinder.stats.govt.nz/layer/92215-territorial-authority-2018-clipped-generalised/. Retrieved from StatsNZ Geographic Data Service.

[11] StatsNZ, Tatauranga Aotearoa. Regional council 2018 clipped (generalised), 2021. URL https://datafinder.stats.govt.nz/layer/92205-regional-council-2018-clipped-generalised/. Retrieved from StatsNZ Geographic Data Service.

[12] StatsNZ, Tatauranga Aotearoa. New zealand grid 250 metre, 2023, 2024. URL https://datafinder.stats.govt.nz/layer/115044-new-zealand-grid-250-metre-2023/. Retrieved from StatsNZ Geographic Data Service.

[13] StatsNZ, Tatauranga Aotearoa. New zealand grid 500 metre, 2023, 2024. URL https://datafinder.stats.govt.nz/layer/115045-new-zealand-grid-500-metre-2023/. Retrieved from StatsNZ Geographic Data Service.

[14] StatsNZ, Tatauranga Aotearoa. New zealand grid 1 kilometre, 2023, 2024. URL https://datafinder.stats.govt.nz/layer/115046-new-zealand-grid-1-kilometre-2023/. Retrieved from StatsNZ Geographic Data Service.

[15] Toitū Te Whenua, Land Information New Zealand. Nz roads (addressing), 2025. URL https://data.linz.govt.nz/layer/53382-nz-roads-addressing/. Retrieved from LINZ Data Service.

[16] Infrageomatics/Open Infrastructure Maps. Openstreetmap infrastructure export (new zealand), 2025. URL https://www.infrageomatics.com/products/osm-export. Data export purchased from Infrageomatics and accessible via Open Infrastructure Maps at: https://openinframap.org/#5.72/-42.223/172.617.

[17] Wellington Water. Regional water pipes, 2025. URL https://services7.arcgis.com/2ECs938g489DMWjt/ArcGIS/rest/services/Regional_Water_Pipes/FeatureServer/0. Retrieved from Wellington Water ArcGIS API REST endpoints searchable at: https://data-wellingtonwater.opendata.arcgis.com/search.

[18] Wellington Water. Regional stormwater pipes, 2025. URL https://services7.arcgis.com/2ECs938g489DMWjt/ArcGIS/rest/services/Regional_Stormwater_Pipes/FeatureServer/0. Retrieved from Wellington Water ArcGIS API REST endpoints searchable at: https://data-wellingtonwater.opendata.arcgis.com/search.

[19] Wellington Water. Regional wastewater pipes, 2025. URL https://services7.arcgis.com/2ECs938g489DMWjt/ArcGIS/rest/services/Regional_Wastewater_Pipes/FeatureServer/0. Retrieved from Wellington Water ArcGIS API REST endpoints searchable at: https://data-wellingtonwater.opendata.arcgis.com/search.

[20] Kapiti Coast District Council. Water services, 2025. URL https://maps.kapiticoast.govt.nz/server/rest/services/Public/Services/MapServer/2. Retrieved from Kapiti Coast District Council ArcGIS API REST endpoints accessible at: https://maps.kapiticoast.govt.nz/LocalMaps/viewer/?webmap=627d29f22676457ca22bc92c19a095cc.

[21] Kapiti Coast District Council. Stormwater, 2025. URL https://maps.kapiticoast.govt.nz/server/rest/services/Public/Services/MapServer/20. Retrieved from Kapiti Coast District Council ArcGIS API REST endpoints accessible at: https://maps.kapiticoast.govt.nz/LocalMaps/viewer/?webmap=627d29f22676457ca22bc92c19a095cc.

[22] Kapiti Coast District Council. Wastewater, 2025. URL https://maps.kapiticoast.govt.nz/server/rest/services/Public/Services/MapServer/11. Retrieved from Kapiti Coast District Council ArcGIS API REST endpoints accessible at: https://maps.kapiticoast.govt.nz/LocalMaps/viewer/?webmap=627d29f22676457ca22bc92c19a095cc.

[23] South Wairarapa District Council. Water public masterton, 2025. URL https://gis.mstn.govt.nz/arcgis/rest/services/Services/WaterPublic/MapServer/5. Retrieved from South Wairarapa District Council ArcGIS API REST endpoints accessible at: https://gis.mstn.govt.nz/WairarapaViewer/?map=d6f5378092314eb4ac10c276906e554d.

[24] South Wairarapa District Council. Stormwater public masterton, 2025. URL https://gis.mstn.govt.nz/arcgis/rest/services/Services/StormwaterPublic/MapServer/6. Retrieved from South Wairarapa District Council ArcGIS API REST endpoints accessible at: https://gis.mstn.govt.nz/WairarapaViewer/?map=d6f5378092314eb4ac10c276906e554d.

[25] South Wairarapa District Council. Sewer public masterton, 2025. URL https://gis.mstn.govt.nz/arcgis/rest/services/Services/SewerPublic/MapServer/4. Retrieved from South Wairarapa District Council ArcGIS API REST endpoints accessible at: https://gis.mstn.govt.nz/WairarapaViewer/?map=d6f5378092314eb4ac10c276906e554d.

[26] South Wairarapa District Council. Water public carterton, 2025. URL https://gis.mstn.govt.nz/arcgis/rest/services/Services/WaterPublic/MapServer/12. Retrieved from South Wairarapa District Council ArcGIS API REST endpoints accessible at: https://gis.mstn.govt.nz/WairarapaViewer/?map=d6f5378092314eb4ac10c276906e554d.

[27] South Wairarapa District Council. Stormwater public carterton, 2025. URL https://gis.mstn.govt.nz/arcgis/rest/services/Services/StormwaterPublic/MapServer/13. Retrieved from South Wairarapa District Council ArcGIS API REST endpoints accessible at: https://gis.mstn.govt.nz/WairarapaViewer/?map=d6f5378092314eb4ac10c276906e554d.

[28] South Wairarapa District Council. Sewer public carterton, 2025. URL https://gis.mstn.govt.nz/arcgis/rest/services/Services/SewerPublic/MapServer/10. Retrieved from South Wairarapa District Council ArcGIS API REST endpoints accessible at: https://gis.mstn.govt.nz/WairarapaViewer/?map=d6f5378092314eb4ac10c276906e554d.

[29] WaterCare. Water pipe, 2025. URL https://data-watercare.opendata.arcgis.com/datasets/f7815fae85e84fa8ad5afc2158c5bdbe_4/explore. GeoJSON file retrieved from WaterCare open data portal searchable at: https://data-watercare.opendata.arcgis.com/.

[30] Auckland Council. Stormwater pipe, 2025. URL https://services1.arcgis.com/n4yPwebTjJCmXB6W/arcgis/rest/services/Stormwater_Pipe/FeatureServer/0. Retrieved from Auckland Council ArcGIS API REST endpoint searchable at: https://data-aucklandcouncil.opendata.arcgis.com/search?categories=

[31] Watercare. Wastewater pipe, 2025. URL https://data-watercare.opendata.arcgis.com/datasets/c6630b1357824f0a8fce0ebde6839f67_2/explore. GeoJSON file retrieved from WaterCare open data portal searchable at: https://data-watercare.opendata.arcgis.com/search?tags=Wastewater.

[32] CanterburyMaps. Canterbury water supply pipelines, 2025. URL https://services1.arcgis.com/RNxkQaMWQcgbiF98/ArcGIS/rest/services/Canterbury_Three_Waters_Data_2_view/FeatureServer/5. Retrieved from Canterbury Maps ArcGIS API REST endpoints accessible at: https://mapviewer.canterburymaps.govt.nz/.

[33] CanterburyMaps. Canterbury stormwater pipelines, 2025. URL https://services1.arcgis.com/RNxkQaMWQcgbiF98/ArcGIS/rest/services/Canterbury_Three_Waters_Data_2_view/FeatureServer/3. Retrieved from Canterbury Maps ArcGIS API REST endpoints accessible at: https://mapviewer.canterburymaps.govt.nz/.

[34] CanterburyMaps. Canterbury wastewater pipelines, 2025. URL https://services1.arcgis.com/RNxkQaMWQcgbiF98/ArcGIS/rest/services/Canterbury_Three_Waters_Data_2_view/FeatureServer/4. Retrieved from Canterbury Maps ArcGIS API REST endpoints accessible at: https://mapviewer.canterburymaps.govt.nz/.

[35] Southland District Council. Water supply, 2025. URL https://gis.southlanddc.govt.nz/server/rest/services/External_Utilities_Layers/MapServer/10. Retrieved from Southland District Council ArcGIS API REST endpoint available at: https://gis.southlanddc.govt.nz/localmapspublic/viewer/?webmap=63ff7e215bbe43b1b0d47aeeb893c530.

[36] Southland District Council. Stormwater main, 2025. URL https://gis.southlanddc.govt.nz/server/rest/services/External_Utilities_Layers/MapServer/49. Retrieved from Southland District Council ArcGIS API REST endpoint available at: https://gis.southlanddc.govt.nz/localmapspublic/viewer/?webmap=63ff7e215bbe43b1b0d47aeeb893c530.

[37] Southland District Council. Wastewater - main, 2025. URL https://gis.southlanddc.govt.nz/server/rest/services/External_Utilities_Layers/MapServer/31. Retrieved from Southland District Council ArcGIS API REST endpoint available at: https://gis.southlanddc.govt.nz/localmapspublic/viewer/?webmap=63ff7e215bbe43b1b0d47aeeb893c530.

[38] Top of the South Maps. Water pipes, 2025. URL https://www.topofthesouthmaps.co.nz/arcgis/rest/services/OpenData/OpenData1/MapServer/5. Retrieved from Top of the South Maps ArcGIS API REST endpoint available at: https://www.topofthesouthmaps.co.nz/app/.

[39] Top of the South Maps. Stormwater pipes, 2025. URL https://www.topofthesouthmaps.co.nz/arcgis/rest/services/OpenData/OpenData1/MapServer/7. Retrieved from Top of the South Maps ArcGIS API REST endpoint available at: https://www.topofthesouthmaps.co.nz/app/.

[40] Top of the South Maps. Wastewater pipes, 2025. URL https://www.topofthesouthmaps.co.nz/arcgis/rest/services/OpenData/OpenData1/MapServer/6. Retrieved from Top of the South Maps ArcGIS API REST endpoint available at: https://www.topofthesouthmaps.co.nz/app/.

[41] Waikato District Council. Water supply pipe, 2025. URL https://data.waikatodistrict.govt.nz/layer/95860-water-supply-pipe/. Geopackage file retrieved from Waikato District Council open data portal searchable at: https://data.waikatodistrict.govt.nz/data/.

[42] Waikato District Council. Storm water pipe, 2025. URL https://data.waikatodistrict.govt.nz/layer/95369-storm-water-pipe/. Geopackage file retrieved from Waikato District Council open data portal searchable at: https://data.waikatodistrict.govt.nz/data/.

[43] Waikato District Council. Waste water pipe, 2025. URL https://data.waikatodistrict.govt.nz/layer/95372-waste-water-pipe/. Geopackage file retrieved from Waikato District Council open data portal searchable at: https://data.waikatodistrict.govt.nz/data/.

[44] GNS Science. Vertical infrastructure capital value estimates at 250m grid and sa1 levels. Unpublished internal data release, 2025. Provided by GNS Science to the New Zealand Infrastructure Commission.

[45] StatsNZ, Tatauranga Aotearoa. Statistical area 1 dataset for 2018 census, total new zealand, 2020. URL https://www3.stats.govt.nz/2018census/SA1Dataset/Statistical%20Area%201%20dataset%20for%20Census%202018%20-%20total%20New%20Zealand%20-%20CSV_updated_16-7-20.zip. Retrieved from StatsNZ at: https://www.stats.govt.nz/information-releases/statistical-area-1-dataset-for-2018-census-updated-march-2020/.

[46] StatsNZ, Tatauranga Aotearoa. Statistical area 2 population projections, characteristics, 2018(base)-2048, 2021. URL https://api.data.stats.govt.nz/rest/data/STATSNZ,POPPR_SUB_002,1.0/all?dimensionAtObservation=AllDimensions&format=csvfilewithlabels. Retrieved from StatsNZ Aotearoa Data Explorer.

[47] StatsNZ, Tatauranga Aotearoa. New zealand estimated resident population grid 250 metre, 2025. URL https://datafinder.stats.govt.nz/layer/119709-new-zealand-estimated-resident-population-grid-250-metre/. Retrieved from StatsNZ Geographic Data Service.

[48] StatsNZ, Tatauranga Aotearoa. New zealand estimated resident population grid 500 metre, 2025. URL https://datafinder.stats.govt.nz/layer/120026-new-zealand-estimated-resident-population-grid-500-metre/. Retrieved from StatsNZ Geographic Data Service.

[49] StatsNZ, Tatauranga Aotearoa. New zealand estimated resident population grid 1 kilometre, 2025. URL https://datafinder.stats.govt.nz/layer/119989-new-zealand-estimated-resident-population-grid-1-kilometre/. Retrieved from StatsNZ Geographic Data Service.

[50] Toitū Te Whenua, Land Information New Zealand. Nz coastlines and islands polygons (topo 1:50k), 2025. URL https://data.linz.govt.nz/layer/51153-nz-coastlines-and-islands-polygons-topo-150k/. Retrieved from LINZ Data Service.

[51] Te Kaunihera o Tāmaki Makaurau, Auckland Council. Auckland council annual reports - financial statements, 2025. URL https://www.aucklandcouncil.govt.nz/en/plans-policies-bylaws-reports-projects/our-annual-reports.html.

[52] Wellington City Council. Wellinton city council annual reports - financial statements, 2025. URL https://wellington.govt.nz/your-council/plans-policies-and-bylaws/plans-and-reports/annual-report.

[53] Christchurch City Council. Christchurch city council annual reports - financial statements, 2025. URL https://ccc.govt.nz/the-council/how-the-council-works/reporting-and-monitoring/annualreport.

[54] OpenStreetMap contributors. Openstreetmap. https://www.openstreetmap.org, 2025. OpenStreetMap data, freely contributed and licensed under the Open Data Commons Open Database License (ODbL).