MACHINE LEARNING

When a Temporary Bypass Becomes the Bottleneck: Machine Learning and Discrete-Event Simulation for Scheduling Water Main Rehabilitation

Prince Keshi

Independent Researcher, Salford, United Kingdom

Published November 2026

Abstract

Rehabilitating small water mains by cement mortar lining or cured-in-place pipe usually keeps customers supplied through a temporary bypass. Bypass kits are scarce and must be cleaned and moved between streets, so releasing a kit at one site becomes a predecessor of work at another. A quantile regression forest was trained on a synthetic 10-year history of 480 sites calibrated to published utility documents. On a chronological hold-out of 144 sites it reduced mean pinball loss by 9% against historical distributions by work type, and its 90th percentile covered 87.5% of outcomes, against 21.5% for planned durations. In a discrete-event simulation of 1,000 ten-site programmes, better forecasts raised the probability that every kit was released on time from 0.293 to 0.329, and resequencing that protected every waiting site raised it to 0.344; protecting only the next site lowered it to 0.225. A lower booking confidence added 7 to 11 percentage points, and one extra kit added 15 to 17. Customer-hours of interruption varied by under 2% across policies. Results are scenario findings, not utility estimates.

Keywords: water main rehabilitation; temporary bypass; multi-project scheduling; quantile regression forest; discrete-event simulation; shared resources; backfilling

1 Practical applications

Rehabilitation programmes often run several streets at once, and each street needs a temporary bypass for about four to six weeks. If one street overruns, its bypass cannot move on time, and the next street may lose an approved work window such as a school holiday or a traffic permit. The study suggests four steps. First, forecast when each kit will be ready at its next site, including cleaning and transfer, as a range rather than a single date, and update the range as each stage finishes. Planned durations were too optimistic: almost four test sites in five overran them. Second, lend an idle kit to another ready street only when every street already waiting for a fixed window still has a kit it can count on. Checking only the next street in the queue made programmes later, not earlier. Third, decide openly how sure the planner must be before notifying customers; a lower bar finishes more programmes on time but withdraws more letters. Fourth, where deadlines matter, weigh the cost of one more kit against the cost of lateness. In these scenarios an extra kit did more than any forecast or sequencing rule.

2 Introduction

Trenchless rehabilitation of distribution mains has changed what limits progress on site. When a main is cleaned and lined with cement mortar or a cured-in-place liner, the excavation is small, but the customers on that street must be kept in water while the host pipe is out of service. The usual

answer is a temporary bypass: small-diameter pipe laid along the curb, tapped into each service and fed from a live hydrant or main. Municipal notices describe the pattern in plain terms. In Somerville, Massachusetts, residents on Oxford, Hershey, Gibbens and School Streets were told that a temporary above-ground bypass would carry their water while the contractor worked on the mains (City of Somerville 2023). East Providence, Rhode Island, tells customers that it aims to finish cleaning and lining a street in about a month and that a bypass is installed only where customers would otherwise lose service for more than one working day (City of East Providence 2026). Unitywater in Queensland requires a bypass wherever an interruption would exceed 8 h on mains up to DN200, and requires it to be designed to stay in place for no more than 10 working days (Unitywater 2026).

Programme schedules for this work are often built street by street, with crews, plant and permits as the visible constraints. The bypass is treated as part of each street. Yet the kit is a pooled, reusable asset. It must be removed, cleaned, disinfected and carried to the next site, and specifications such as Kalamazoo's require that used bypass pipe come from stock kept for potable work (City of Kalamazoo 2026a). When several streets share a small pool, the release of a kit at one street becomes a hidden predecessor of the start at another. A late test result at Site A can make Site B miss its window, even when the crew for Site B is ready and waiting.

This paper takes that coupling as its subject. The central question is whether planners who share a limited pool of bypass kits can use forecasts of when each kit will actually be free to reduce customer disruption and protect programme deadlines. The sharper operating question is this: when a site risks missing its next approved window, under what conditions should a planner move a shared kit to another site that is ready? The study answers both with a quantile regression forest (Meinshausen 2006) that forecasts kit occupancy using only information known at the time of the decision, and a discrete-event simulation that replays whole programmes under shared disruptions.

Three contributions follow. The first is a milestone structure for each site with four separate events: rehabilitation finished, permanent service restored, kit released, and kit ready at the next site. It makes the hand-over between sites explicit and keeps the customer outcome apart from the schedule outcome. The second is a link to backfilling from computer job scheduling (Mu'alem and Feitelson 2001; Tsafrir et al. 2007), which offers a tested vocabulary for deciding when a later, ready job may jump an earlier one that is waiting. The third is a controlled comparison, with common random numbers, of forecast quality and resequencing policy, together with a perfect-foresight bound, a booking confidence sweep and an extra-kit test. All results come from a synthetic case calibrated to public documents, and they are presented as scenario findings.

3 Background

3.1 Bypass practice in published utility documents

Public sources describe what a bypass is, how long it stays and what must happen before it moves. They do not say how kits are shared between streets. Table 1 lists the sources used and what each one supports. Three points matter for the model. Durations are of the order of weeks: for its Inman Street main replacement, Cambridge, Massachusetts, planned 1 to 2 weeks to install a bypass and about 4 weeks in service (City of Cambridge 2026), and a supplier states that customers are typically on temporary supply for 4 to 6 weeks (Core Aqua n.d.). Notice periods are fixed: the Madison, Wisconsin, specification calls for 72 h and 48 h notices and one working day before customers are moved onto the bypass (City of Madison n.d.-b), and Unitywater requires a planned interruption

notice 28 days ahead (Unitywater 2026). Hand-over has its own tasks: bypass lines are disinfected and tested before connection and removed once the main is accepted (City of Kalamazoo 2026a; Unitywater 2026).

Madison offers the clearest public chronology of sequential phases under one contractor. In 2025 Fer-Pal lined mains in two phases. Phase 1 (Churchill Drive, Todd Drive, Greenway View) was listed for June to September and Phase 2 (Pelham Road and nearby streets) for September to November. On September 8 the city reported Phase 2 bypass connections due that week while Phase 1 customers were being returned to the permanent main, and on September 18 the Phase 1 temporary system was being removed while Phase 2 lining was under way (City of Madison 2025). By October 13 Greenway View still needed a new ductile iron main. Phases overlapped, and the end of one fed the start of the next. The pages do not say whether the same pipe moved between phases, and this paper does not assume that it did. The 2026 east-side project shows the stage rhythm: contract award on May 20, work from June 22, bypass installed and under test by July 16, lining complete on August 19, services reconnected in about a week, and completion expected in mid-September (City of Madison 2026). Providence Water's Contract 1-25 goes further and asks for the bypass to be phased rather than built as one continuous system over about 8,500 m (28,000 ft) of main (Providence Water 2025), which is an explicit, if indirect, sign that bypass stock is a constraint.

Table 1. Documented evidence base and its use in the model

Source	What it documents	Use in this study
City of Somerville (2023)	Temporary above-ground bypass for four streets during main works; contractor named	Practice exists in New England programmes
City of East Providence (2026)	Cleaning and lining steps; target of about one month per street; bypass only if outage over one working day; 4 to 8 h shutdowns otherwise	Stage list; stage ranges; bypass trigger
City of Madison (2025, 2026)	Two overlapping CIPP phases in 2025; dated stage milestones in 2026	Stage order and durations; worked example of phase hand-over
City of Madison (n.d.-a, n.d.-b)	Bypass plan lengths of about 550 m (1,800 ft); 72 h, 48 h and one working day notices	Kit size classes; booking notice
Unitywater (2026)	Bypass above 8 h outage; mains to DN200; designed for at most 10 working days; disinfection before connection; prompt removal; 28-day notice	Hand-over tasks; notice; size class limits
City of Kalamazoo (2026a, 2026b)	Used bypass pipe must come from potable-only stock; pressure and flow limits; 2,390 m (7,840 ft) of CIPP	Cleaning step before reuse; compatibility classes
City of Cambridge (2026)	Bypass installation 1 to 2 weeks; in place about 4 weeks	Occupancy check
Core Aqua (n.d.)	Customers on temporary supply about 4 to 6 weeks	Occupancy check
Providence Water (2024, 2025)	1,280 m (4,200 ft) of 51 mm (2 in.) bypass on a 1,680 m (5,500 ft) contract; bypass to be phased on an 8,500 m (28,000 ft) contract	Evidence that bypass stock is staged and limited
City of Melrose (2025)	Bypass installed July 21, hook-ups July 28 to 30	Set-up stage duration
Ofwat (2025)	Supply interruption measure counts interruptions of 3 h or more	Customer-hours subset of 3 h or more

3.2 Scheduling shared resources across projects

The decision studied here is a resource-constrained multi-project scheduling problem. Gómez Sánchez et al. (2023) survey the field and note that most work assumes resources move between projects instantly. Krüger and Scholl (2009) are an exception: their formulation adds sequence-dependent transfer times when a resource leaves one project for another, which is the role that cleaning and haulage play for a bypass kit. Van Eynde and Vanhoucke (2020) provide benchmark sets and show that the best priority rules depend on the objective. These studies are mostly deterministic. Herroelen and Leus (2005) set out the main ways to handle uncertainty, namely reactive repair, stochastic policies and robust baselines, and the approach here sits in the stochastic and reactive group. A baseline programme is fixed in advance, and a rule decides at each booking point whether to keep or change the order.

Backfilling, from parallel computing, is a close analogue. In EASY backfilling a smaller job may run ahead of the queue if it will not delay the job at the head (Mu'alem and Feitelson 2001). Conservative backfilling gives every queued job a reservation. Tsafrir et al. (2007) showed that runtime predictions made by the system, rather than by users, improve such schedulers. A water programme differs in two ways. Starts are tied to approved windows, so a short delay can become a long one, and moves carry a cleaning and transfer cost. The study tests both the EASY and conservative forms.

3.3 Forecasting durations as distributions

A booking decision needs the probability that a kit will be ready by a given day, not a single best guess. Quantile regression forests estimate the full conditional distribution by keeping every observation in each leaf (Meinshausen 2006). They handle mixed inputs and need little tuning. Forecast quality is judged with the pinball loss, a proper scoring rule for quantiles, and by coverage on held-out data (Gneiting and Raftery 2007). Park et al. (2024) embedded learned operation-time predictions inside a discrete-event simulation of a container terminal, which is the pattern followed here. No published study was found that forecasts bypass occupancy or links it to sequencing.

4 Methods

4.1 Site milestones and kit occupancy

Each site passes through six stages: bypass set-up and hook-up, pits and valves, the rehabilitation core (cleaning and lining, or liner installation and cure), reinstatement of services, testing and disinfection, and reconnection with bypass removal. Four milestones are recorded (Figure 1). M1 is the end of rehabilitation work. M2 is the return of customers to the permanent main, set at 60% of the way through the last stage. M3 is the release of the kit from the site. M4 is the moment the kit is ready at the next site after cleaning, disinfection and a one-day transfer. Kit occupancy is the time from the start of set-up to M3, and the cleaning and transfer time from M3 to M4 is modelled separately. Customer outcomes depend on M2, and programme outcomes depend on M3 and M4.

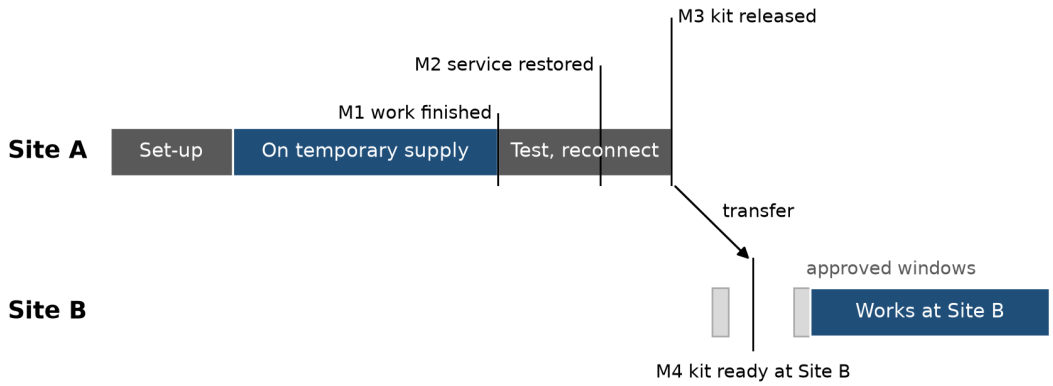

Figure 1. The hand-over between two sites that share a kit. Site B can start only at an approved window after M4 at Site A

4.2 Synthetic history anchored on documents

No public data set records bypass kit movements, so a generator was built and calibrated against the documents in Table 1. Sites are cement mortar lining (60%) or cured-in-place pipe (40%), and small-diameter (70%) or large, with main length drawn from a lognormal distribution with a median of 300 m (clipped to 120 to 750 m). Each site has 0.08 to 0.16 service connections per metre (at least six), a Poisson number of side streets (mean 1.5), one of three contractors with duration multipliers of 0.90, 1.00 and 1.15, a 20% chance of needing a booster pump on the bypass and a 15% chance of priority status. Stage durations are triangular, with ranges taken from the East Providence sequence and the Madison chronologies, scaled by length and connection count, and multiplied by 0.75 to allow for the overlap of steps that the East Providence page describes. Hidden factors, unknown at planning time, add variation: a lognormal ground-condition factor on the pits stage, a factor of 0.9 to 1.5 on reconnection at cured-in-place sites for delays in gaining access to properties, and failed bacteriological tests (10% for lining, 15% for liners) followed by a retest of 2 to 5 days. A lognormal programme factor (standard deviation 0.10 on the log scale) is shared by all sites in a programme. Two types of shared disruption act on whole programmes. Crew shocks halt the core stage for one work type at about 0.8 events per 13 weeks, lasting 3 to 10 days. Weather halts the pits and reinstatement stages at about 1.5 events per 13 weeks, lasting 1 to 4 days. Planned durations are the expected stage times with no hidden factors, contractor or programme effects, or shocks, which is how a contractor's bar chart is usually built.

The resulting history covers 2016 to 2025, with four programmes of 12 sites each year, or 480 site records, together with a snapshot at the end of each stage. The median occupancy is 25 working days for lining (10th to 90th percentile, 20 to 33) and 31 for liners (24 to 42). These match the documented 4 to 6 weeks on temporary supply. In the programme simulation each connection spent about 20 working days on temporary supply, which is close to the Cambridge figure of about 4 weeks. Table 2 lists the main settings.

Table 2. Main settings of the programme simulation

Setting	Value	Basis
Sites per programme	7 small, 3 large	Scenario
Kits	5 small, 2 large	Scenario; Providence phasing

Setting	Value	Basis
Restricted-window sites	2 small, 1 large	Scenario (school, traffic or event windows)
Windows	Weekly; restricted sites: weekly within two 11-day blocks starting 30 days apart	Scenario
Booking notice L	5 working days	Madison notices; Unitywater
Kit cleaning and transfer	1 to 3 days plus 1 day	Kalamazoo; Unitywater
Site not ready at window	5% per attempt	Scenario
Deadline D	65 working days	Scenario (13 weeks)
Planned finish	Day 49 to 52 (13 to 16 days of float)	Selection rule
Programmes and replications	1,000, common random numbers	Design

4.3 Forecasting models

Two forecasts are made. The planning forecast predicts total occupancy before a site starts from 12 features known at that time: work type, size class, length, number of connections, number of side streets, contractor (three indicator variables), booster pump need, priority status, start month and planned occupancy. The delivery forecast predicts remaining occupancy at the end of each stage, adding the number of stages completed, elapsed days, slip against plan and planned remaining time. Both use a quantile regression forest with 400 trees, a minimum leaf size of five and half the features tried at each split, implemented with the quantile-forest package for Python and scikit-learn 1.9.1. Two benchmarks are used. The planned benchmark is the contractor's planned duration as a single value. The historical benchmark is the empirical distribution of occupancy by work type and, in delivery, of remaining occupancy by work type and stages completed, conditioned on the time already spent in the current stage. A perfect-foresight oracle is used only as an upper bound in the simulation.

To avoid leakage, models were trained on programmes from 2016 to 2022 (336 sites) and tested on 2023 to 2025 (144 sites), so no programme contributes to both sets and every test site lies in the future of the training data. A random split of the same size was run as a check. Accuracy is reported as the mean pinball loss over the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles, the mean absolute error of the median, and coverage, meaning the share of outcomes at or below each forecast quantile. For use in the simulation, final models were refitted on all 10 years. The probability that a kit is ready for a window is the share of forecast quantiles of its release day, each increased by the matching quantile of the cleaning and transfer time, that fall at or before the window.

4.4 Discrete-event simulation

The simulation advances one working day at a time. A site can start only when four prerequisites hold on the same day: a kit of its class is physically ready, a crew is free (crew limits enter through the readiness check and the crew shocks), the day is an approved window, and the site is ready (a 5% chance of failure on any attempt covers permits, access and traffic management). Five working days before each window, the planner decides whether to book the site and notify customers. Booking requires the forecast probability that a committed kit will be ready by the window to reach the booking confidence q, which is 0.8 unless stated. For a site's last remaining window the threshold is

relaxed to 0.5; a site that still cannot be booked, or that runs out of windows, is deferred and its programme counts as late. If the kit is not ready when the day arrives, mobilisation is cancelled and the notified connections are counted as notified then cancelled. Shared disruptions are drawn in advance for each programme, so the same weather and crew events strike every policy, and the stage durations of each site are drawn in advance as well. This use of common random numbers means that differences between policies are paired and not masked by luck.

Each programme starts from a baseline plan. Restricted sites are placed first in their earliest feasible block and the others fill the earliest kit lane and window, using planned durations plus a 2-day allowance for cleaning and transfer. Only programmes whose plan finishes by day 52 are kept, which gave planned finishes between days 49 and 52 and a planned float of 13 to 16 working days, or about 25%. Four booking rules are compared:

- **Fixed sequence.** Sites are booked strictly in plan order. A site that cannot be booked holds up those behind it in its class.
- **Conservative resequencing.** A later site that is ready for the coming window may take a free kit if every earlier site still waiting keeps a reservation. For each such site, either a different free kit is forecast ready by its next window with probability at least τ, or the moved kit is forecast to return, cleaned and transferred, by that window with probability at least τ. Each kit can cover only one reservation. τ is 0.8 unless stated.
- **Next-site resequencing (EASY).** As above, but only the first waiting site is protected.
- **First ready site.** Any ready site may take any free kit in plan order, with no protection.

The 2 × 2 core of the design crosses historical against forest forecasts with the fixed sequence against conservative resequencing. Planned and oracle forecasts give the lower and upper reference points. The sensitivity runs cover the EASY and first-ready rules, τ of 0.5, 0.8 and 0.95 under conservative resequencing and 0.5 to 0.95 under the EASY rule, booking confidence q of 0.5, 0.65 and 0.9, a sixth small kit, and an ablation in which shocks strike sites independently rather than in common. In all, 49 policy arms were run on the same 1,000 programmes.

4.5 Outcomes

The two primary outcomes are the probability that the programme is on time, meaning every kit is released by day 65, and customer-hours of interruption. Customer-hours are counted once per connection and event, from three sources kept apart: the planned changeover when a connection moves onto and off the bypass (15 min to 1 h each way), unplanned bypass incidents such as hose damage (0.025 per site-day, affecting from three connections up to half of the street, for 1 to 8 h), and bursts on mains still awaiting rehabilitation (0.008 per km per working day, or about 2 per km per year, interrupting the whole street for 3 to 10 h). The burst rate is high for a network as a whole but plausible for mains chosen for rehabilitation because of their condition. Following the Ofwat (2025) measure, the share of unplanned customer-hours from events of 3 h or more is also reported. Secondary outcomes are the makespan, kit-days spent idle while a site of the same class is waiting, cancelled mobilisations, connections notified then cancelled, start delay at priority sites, kit travel distance and pump hours. Carbon is not estimated because no public data link pump hours to fuel use on these projects. Differences between policies are reported as paired means over the 1,000 programmes with 95% confidence intervals, and differences in on-time probability are given in percentage points.

5 Results

5.1 Forecast accuracy and calibration

Table 3 and Figure 2 summarize accuracy on the 144 held-out sites. Planned durations were biased: actual occupancy exceeded plan at 78.5% of sites, by 4.9 working days on average. The forest reduced mean pinball loss by 9% against the historical benchmark (1.236 against 1.360) and the error of the median from 4.49 to 4.25 days, with narrower 80% intervals (13.8 against 15.9 days). Its 90th percentile covered 87.5% of outcomes, slightly short of the nominal 0.9 but within sampling error for 144 sites (the 95% band is about 0.85 to 0.95). The historical benchmark covered 91.7% because its intervals are wider. Both models were weak in the tail. For the 28 test sites that took more than 32 days (the 75th percentile of the training data), the 90th percentile covered only 57%, so long overruns are understated by both methods. The random split gave a slightly worse pinball loss (1.307) than the chronological split, so there is no sign of optimism from leakage. Planned occupancy was the most important feature (0.29 of total impurity reduction), followed by work type, connections, the slowest contractor and length.

Table 3. Planning-stage forecast accuracy on held-out programmes, 2023 to 2025 (*n* = 144 sites)

Measure	Planned	Historical	QRF	QRF, random split
Mean pinball loss (days)	2.88	1.36	1.24	1.31
Error of median (days)	5.76	4.49	4.25	4.55
Coverage of P90	0.215	0.917	0.875	0.847
Coverage of 80% interval	0.056	0.875	0.840	0.813
Width of 80% interval (days)	0	15.9	13.8	13.1
P90 coverage, sites over 32 days (n = 28)	n/a	0.57	0.57	n/a

Note: QRF = quantile regression forest; P90 = 90th percentile forecast.

During delivery, the forest beat the historical benchmark at every stage. After the pits stage, for example, the mean pinball loss on remaining occupancy was 0.87 days against 1.00, and after the core stage 0.60 against 0.71 [Figure 2(b)]. The gain comes from the observed slip, which the forest uses and the historical benchmark ignores.

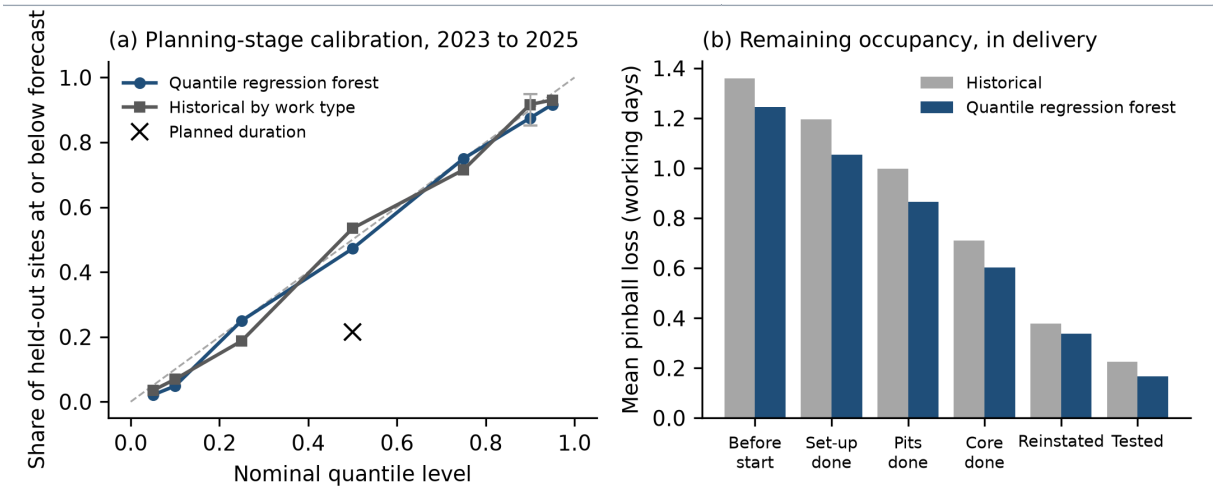

Figure 2. (a) Share of held-out sites at or below each forecast quantile; the error bar shows the 95% sampling band at the 90th percentile; and (b) mean pinball loss of remaining occupancy at the end of each stage

5.2 Programme outcomes

Under the fixed sequence with historical forecasts, 29.3% of programmes released every kit by day 65, although all of them had been planned to finish by day 52. The median makespan was 71 working days (10th to 90th percentile, 60 to 90). The gap between plan and outturn reflects the planning bias above, compounded through kit hand-overs. Table 4 and Figure 3 give the results for each arm.

Table 4. Mean outcomes per programme (1,000 programmes, common random numbers)

Forecast and rule	P(on time)	Makespan (days)	Un-planned CH	CH of 3 h or more	Idle kit-days while a site waits	Notified then can-celled	Priority delay (days)
Planned, fixed	0.346	72.3	423	392	78.6	49.8	3.09
Historical, fixed	0.293	73.8	425	394	84.0	0.0	3.52
QRF, fixed	0.329	73.0	423	392	81.6	0.1	3.33
Oracle, fixed	0.415	70.9	422	391	75.9	0.0	2.84
Historical, con-servative	0.311	73.1	424	393	79.8	0.0	3.18
QRF, conservat-ive	0.344	72.4	423	392	77.8	0.2	2.99
Oracle, conser-vative	0.433	70.2	421	390	72.8	0.0	2.50
QRF, next-site (EASY)	0.225	76.4	425	394	82.7	0.3	2.81
QRF, first ready site	0.030	82.4	438	408	96.3	0.2	3.43
QRF, fixed, q = 0.5	0.402	71.1	422	391	76.1	12.6	2.85
QRF, conservat-ive, q = 0.5	0.409	70.8	422	390	73.9	13.7	2.62

Forecast and rule	P(on time)	Makespan (days)	Un-planned CH	CH of 3 h or more	Idle kit-days while a site waits	Notified then can-celled	Priority delay (days)
Historical, fixed, 6 small kits	0.459	69.2	417	386	95.7	0.0	1.71
QRF, conservat-ive, 6 small kits	0.495	68.4	410	379	84.7	0.1	0.70
QRF, fixed, q = 0.5, 6 small kits	0.550	67.1	415	385	91.7	8.9	1.27

Note: CH = customer-hours of interruption; QRF = quantile regression forest. Booking confidence q = 0.8 and reservation threshold τ = 0.8 unless stated. Notified then cancelled counts connections per programme.

Forecast quality. With the order fixed, forest forecasts raised the on-time probability by 3.6 percentage points over historical forecasts (95% confidence interval 2.4 to 4.8) and shortened the makespan by 0.76 days (0.52 to 0.99). The mechanism is simple. A narrower, better-centred forecast reaches the booking confidence sooner, so fewer windows are passed up while a kit that would in fact be ready is treated as doubtful. Idle kit-days while a site waited fell by 2.4. The oracle shows how much room remains: perfect foresight added 12.2 points over historical forecasts, so the forest closed about 30% of the gap.

Resequencing. Conservative resequencing added 1.5 points with forest forecasts (0.7 to 2.3) and 1.8 points with historical forecasts (1.0 to 2.6). The two effects were additive: the interaction was −0.3 points (−0.9 to 0.3). Together they took the on-time probability from 0.293 to 0.344, a gain of 5.1 points (3.7 to 6.5), and cut 1.4 days from the makespan. Resequencing improved the makespan in 10.9% of programmes and worsened it in 0.5%. It mattered most where kits sat idle under the fixed order: in the third of programmes with the most idle kit-days it saved 1.65 days, against about 0.1 days elsewhere. The EASY variant, which protects only the next site, lost 10.4 points against the fixed sequence even with forest forecasts (8.2 to 12.6), and the first-ready rule left only 3% of programmes on time. Tracing single programmes showed the cause. A flexible site took one of two free kits. The next site in line passed the test because the second kit was still free, but the site after it, which had a short restricted block, then found no kit and was pushed 30 days to its next block. With perfect foresight the EASY rule still lost 12.1 points against the fixed order, so the failure lies in the rule, not in the forecasts.

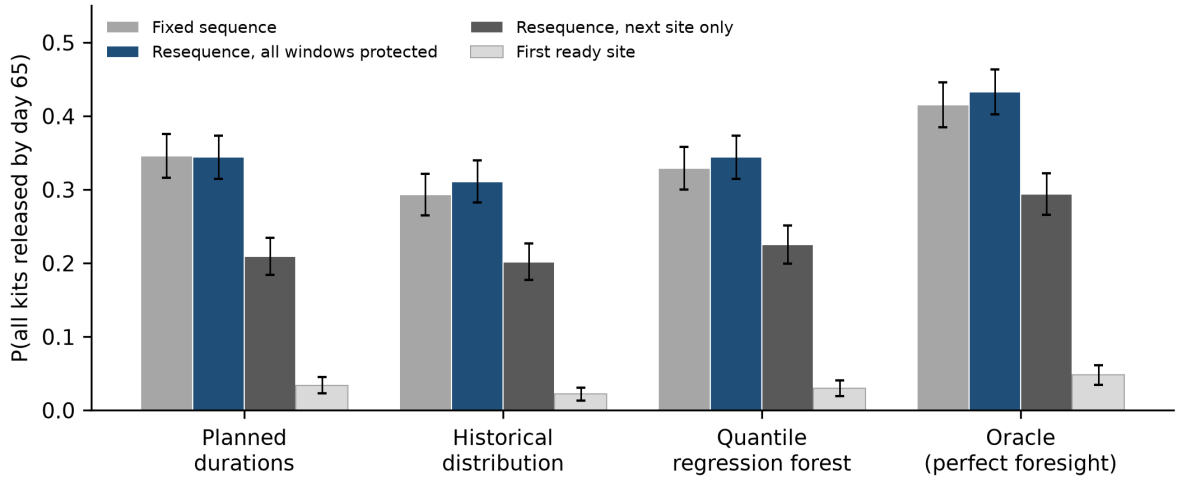

Figure 3. Probability that every kit is released by day 65, by forecast and booking rule. Bars show means over 1,000 programmes and whiskers show 95% confidence intervals

Booking confidence. Planned durations treat every kit as certain to be ready. With the fixed sequence they did better than historical forecasts on time (0.346), but at the cost of 1.6 cancelled mobilisations and about 50 connections notified then cancelled in each programme. Lowering the booking confidence for the probabilistic forecasts traced a clear trade-off [Figure 4(a)]. With forest forecasts, moving q from 0.8 to 0.5 added 7.3 points of on-time probability (5.7 to 8.9) and about 12.5 cancelled notices per programme; with historical forecasts it added 10.5 points (8.6 to 12.4). At q = 0.5 the forest beat planned durations by 5.6 points with 37 fewer cancelled notices. At that setting, however, the forest and the historical benchmark were almost equal (0.402 and 0.398). The advantage of the forest appeared only when the planner demanded high confidence, which is where its tighter intervals count. The value of τ hardly mattered. Across 0.5 to 0.95 the on-time probability under conservative resequencing stayed between 0.339 and 0.345, because the binding question was usually whether any kit could cover a reservation at all, not how sure the forecast was.

One extra kit. A sixth small kit raised the on-time probability by 16.6 points under the historical fixed policy (14.3 to 18.9) and by 14.8 points under the forest at q = 0.5 (12.6 to 17.0), reaching 0.550. That is more than any forecasting or sequencing change in the study [Figure 4(b)]. Under the historical fixed policy the extra kit also cut priority start delay from 3.5 to 1.7 days and kit travel by 1.7 km per programme. It raised idle kit-days by about 12, which is the cost of spare capacity. With six small kits, conservative resequencing with forest forecasts reached 0.495, 3.6 points (2.4 to 4.8) above the historical fixed policy with the same pool. The gain from resequencing alone shrank, however: with historical forecasts it fell to 0.8 points (0.3 to 1.4), because a larger pool leaves fewer sites waiting for a kit.

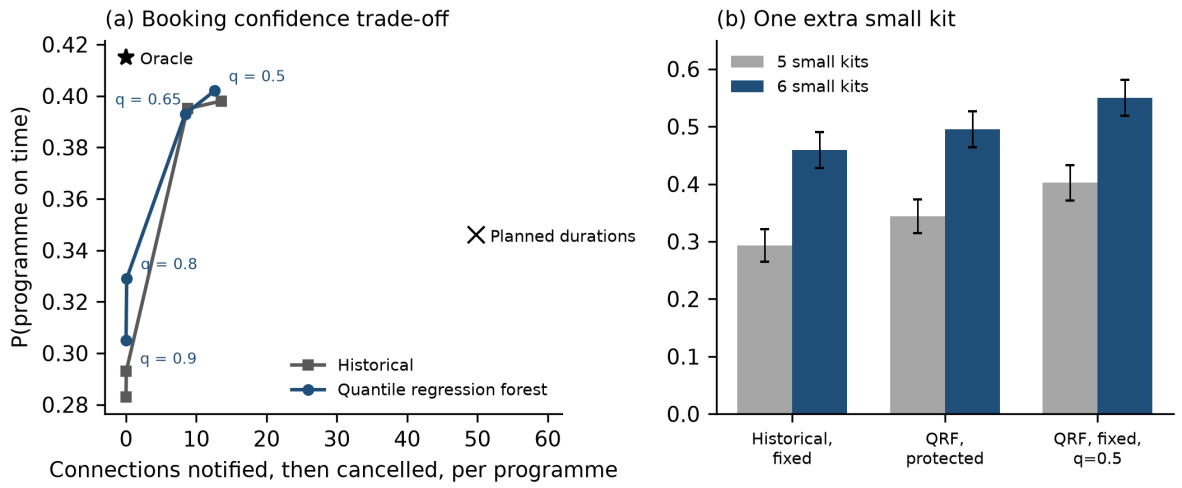

Figure 4. (a) On-time probability against connections notified then cancelled per programme as the booking confidence q changes, fixed sequence; and (b) effect of a sixth small kit; whiskers show 95% confidence intervals

Customer disruption. Planned changeover accounted for 419 of the roughly 842 customer-hours per programme and was the same under every policy, since every connection is moved on and off a bypass once. The unplanned part moved little. Against the historical fixed policy, forest forecasts with conservative resequencing cut unplanned customer-hours by 2.3 (0.3 to 4.4) out of 425, and the Ofwat-style subset of 3 h or more by 2.2 (0.2 to 4.2). Both cuts came from mains being rehabilitated sooner, which shortened their exposure to bursts. The sixth kit cut unplanned customer-hours by 8.3 (4.3 to 12.3), about 2%, and the first-ready rule added 13 (3 to 23). Connection-days on temporary supply did not differ by policy, because each site's bypass period is set by its own stages and not by when it starts. In this setting, sequencing protects deadlines much more than it protects customers.

Shared shocks. When disruptions struck sites independently instead of in common, the baseline on-time probability was almost unchanged (0.302 against 0.293; difference 0.9 points, −2.2 to 4.0), and the EASY rule remained well below the fixed sequence (0.198). Only these two arms were rerun under independent shocks. Common shocks bunch delays in time, but with only about one crew shock per type per programme the effect on kit hand-over was small. Kit travel and pump hours differed by less than 1% between policies apart from the extra-kit runs, so no environmental difference can be claimed.

6 Discussion

6.1 When to move a kit

The results give a direct answer to the operating question. When a site risks missing its next window, a planner should move a shared kit to another ready site only when three conditions hold together. First, the kit would otherwise stand idle until the waiting site's next window, so the move fills a gap rather than taking a kit that is about to be needed. Second, every site ahead in the plan that is still waiting, and above all every site with a restricted window, keeps a reservation. Either a different kit is forecast to be ready for its next window, or the moved kit is forecast to be back, cleaned and transferred, in time. The test must count each kit once. Third, those forecasts must be ranges that are updated as stages finish. A single planned date gives no basis for the test because planned dates were met at only about one site in five.

The size of the window gap counts for more than the confidence threshold. A restricted site that misses a block can lose a month, while a flexible site that misses a weekly slot loses a week. The EASY rule failed because it ignored this asymmetry. In computing, a job that misses its reservation is late by a little; in a water programme it can be late by a whole access block. The lesson for practice is to protect all windows, not just the next one, and to treat sites with short restricted blocks as the reservations that must not be broken.

6.2 Holding, booking and buying

The study separates three levers that are often merged in practice. Holding a kit for an earlier site, as the fixed sequence does, is a safe default and in these scenarios beat any resequencing that did not protect every window. Booking confidence is a policy choice about customers, not a technical setting. A utility that values a firm notice will set q high, and it is there that a better forecast pays. A utility willing to withdraw notices can gain more on time from a lower q than from the forecasting model, and in that case simple historical ranges serve almost as well. Buying or renting one more kit gave the largest gain in every comparison. Whether that is worth it depends on the cost of a kit for a season against the cost of programme lateness, which this paper does not price. The forecasting model can still help with that choice, because it shows how often the existing pool is short.

There is a link to responsible use of resources. A bypass kit is a reusable temporary works asset, and the idle kit-days reported here are a measure of how well that asset is used. Better forecasts and conservative resequencing reduced idle kit-days while a site waited by about 6 per programme (from 84.0 to 77.8). An extra kit raised them. Planners therefore face a trade between asset use and deadline security that can be shown in the same units.

6.3 Relation to earlier work

The findings agree with the multi-project scheduling literature in that the best rule depends on the objective (Van Eynde and Vanhoucke 2020), and they support the call by Krüger and Scholl (2009) to model transfer between projects explicitly. They extend the backfilling results of Tsafrir et al. (2007) to a setting with hard windows, where the conservative form is needed and the EASY form does harm. They also show that a quantile forest inside a discrete-event model (Park et al. 2024) can give useful probabilities, while the value of better probabilities depends on how strict the booking rule is.

7 Limitations

The evidence base has a gap that should be stated plainly. Public documents establish bypass practice, notice periods, hand-over tasks and stage durations, but none of them records how bypass kits are shared between streets or how often a site waits for one. Madison shows sequential, overlapping phases under one contractor, and Providence asks for phased bypass construction, but neither confirms kit reuse across sites. The kit pool, window structure, shock rates, failure rates and customer-hour parameters are therefore scenario assumptions anchored on the documents, not estimates. The synthetic history was generated by the same kind of process that drives the simulation, which favours any model able to learn that process. The forest's advantage over historical ranges should be read as an upper estimate of what real records would give.

Other limits follow from the design. The held-out set has 144 sites, so calibration at the 90th percentile is known only to about ±5 percentage points, and both models were poorly calibrated for long overruns. Crews are not modelled as a separate pool, so crew conflicts appear only through the

site readiness failure and the crew shocks. The baseline plan is built by a simple constructive rule, and an optimized baseline could reduce the gains from resequencing. Hydraulics are not modelled: the bypass is assumed able to supply every street, though Kalamazoo's pressure and flow limits show that this needs checking. EPANET (US EPA n.d.) could be used to test bypass capacity for a real network. Finally, the cost of kits, lateness and cancelled notices is not priced, so the paper ranks policies but does not choose between them on cost.

A retrospective study would remove most of these limits. The minimum data are, for each site, the dates of bypass set-up, M1, M2 and M3, the identity of the kit, and the windows approved. Contractor daily reports and utility notice logs would usually hold this. The code released with the paper accepts such records in place of the synthetic history.

8 Conclusions

Treating the release of a bypass kit as a predecessor of work elsewhere changes how a rehabilitation programme should be run. In a calibrated scenario study of 1,000 ten-site programmes, planned durations were too optimistic for reliable booking. A quantile regression forest gave sharper, broadly calibrated ranges that raised the on-time probability by about 4 percentage points at a fixed order. Resequencing added 1.5 to 1.8 points, but only when every waiting window was protected; protecting only the next site cut on-time completion by about 10 points. Booking confidence and kit numbers were larger levers than either forecasting or sequencing, and customer-hours barely changed under any policy. The decision rule for moving a kit is therefore narrow and testable: move it only into a gap, only when every waiting window still has a kit it can count on, and only on the basis of an updated range of release dates. The next step is to test the rule on a retrospective programme with real kit records.

Data availability statement

All data, models, and code that support the findings of this study are provided with this submission as a supplementary archive and will be deposited in a public repository on acceptance. They comprise the generator, the synthetic history of 480 sites with stage snapshots, the trained models, the simulation code, the raw results of 1,000 programmes for 49 policy arms, and the scripts for the analysis and figures. The documents used for calibration are public and are listed in the references.

Acknowledgments

This research received no specific funding.

References

[1] City of Cambridge (2026). Inman Street reconstruction project. *Department of Public Works*, Cambridge, MA. https://www.cambridgema.gov

[2] City of East Providence (2026). 2026 cleaning & lining work. *Water Utilities Division*, East Providence, RI. https://eastprovidenceri.gov/node/8255

[3] City of Kalamazoo (2026a). Bid 91244-052.0: Specifications and plans, 2026 water main pipe lining project. *Kalamazoo*, MI. https://www.kalamazoocity.org/files/assets/public/v/1/purchasing/bid-documents/2026/91244-052.0/91244-052.0-specifications_plans.pdf

[4] City of Kalamazoo (2026b). 2026 water main pipe lining project. *Kalamazoo*, MI. https://www.kalamazoocity.org/Community/Projects/2026/2026-Water-Main-Pipe-Lining-Project

[5] City of Madison (2025). Water pipeline rehabilitation project. *Madison Water Utility*, Madison, WI. https://www.cityofmadison.com/water/projects/water-pipeline-rehabilitation-project

[6] City of Madison (2026). Water pipeline rehabilitation project - East. *Madison Water Utility*, Madison, WI. https://www.cityofmadison.com/water/projects/water-pipeline-rehabilitation-project

[7] City of Madison (n.d.-a). *Contract 9633: Base bid plans*. Department of Public Works, Madison, WI. https://www.cityofmadison.com/water/projects/water-pipeline-rehabilitation-project

[8] City of Madison (n.d.-b). *Contract 9788: Signed contract and specifications*. Department of Public Works, Madison, WI. https://www.cityofmadison.com/water/projects/water-pipeline-rehabilitation-project

[9] City of Melrose (2025). Notice of work start and bypass hookup. *Melrose*, MA. https://www.cityofmelrose.org/DocumentCenter/View/2126/2025-07-17_Notice-of-Work-Start-and-Bypass-Hookup_Circuit

[10] City of Somerville (2023). Temporary water main bypass installation, March 23: Oxford, Hershey, Gibbens, School. *Somerville*, MA. https://voice.somervillema.gov/springhillsewer/news_feed/temporary-water-main-bypass-installation-march-23-oxford-hershey-gibbens-school

[11] Core Aqua (n.d.). Solutions. https://www.coreaqua.com/solutions

[12] Gneiting T, Raftery AE (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.

[13] Gómez Sánchez M, Lalla-Ruiz E, Fernández Gil A, Castro C, Voß S (2023). Resource-constrained multi-project scheduling problem: A survey. *European Journal of Operational Research*, 309(3), 958–976. https://doi.org/10.1016/j.ejor.2022.09.033

[14] Herroelen W, Leus R (2005). Project scheduling under uncertainty: Survey and research potentials. *European Journal of Operational Research*, 165(2), 289–306.

[15] Krüger D, Scholl A (2009). A heuristic solution framework for the resource constrained (multi-)project scheduling problem with sequence-dependent transfer times. *European Journal of Operational Research*, 197(2), 492–508. https://doi.org/10.1016/j.ejor.2008.07.036

[16] Meinshausen N (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7, 983–999.

[17] Mu'alem AW, Feitelson DG (2001). Utilization, predictability, workloads, and user runtime estimates in scheduling the IBM SP2 with backfilling. *IEEE Transactions on Parallel and Distributed Systems*, 12(6), 529–543. https://doi.org/10.1109/71.932708

[18] Ofwat (2025). Water supply interruptions: Performance commitment definition. *Ofwat*, Birmingham, UK. https://www.ofwat.gov.uk/wp-content/uploads/2025/03/Water-supply-interruptions-%E2%80%93-PC-definition.pdf

[19] Park K, Kim M, Bae H (2024). A predictive discrete event simulation for predicting operation times in container terminal. *IEEE Access*, 12, 58801–58822. https://doi.org/10.1109/ACCESS.2024.3389961

[20] Providence Water (2024). Replacement and rehabilitation of water mains and appurtenances within Edgewood, Cranston, Contract 3-24. *Providence*, RI. https://www.providenceri.gov/wp-content/uploads/2024/03/PVD24-25-Water-Supply-RFP-Replacement-Rehab-of-Water-Mains-Appurtenances-within-Edgewood-Cranston-Contract-3-24-1.5-Year-Contract-MT44043.pdf

[21] Providence Water (2025). Replacement and rehabilitation of water mains and appurtenances within Lower South Providence, North Section, Contract 2-25. *Providence*, RI. https://www.providenceri.gov/wp-content/uploads/2024/03/PVD24-25-Water-Supply-RFP-Replacement-Rehab-of-Water-Mains-Appurtenances-within-Edgewood-Cranston-Contract-3-24-1.5-Year-Contract-MT44043.pdf

[22] Tsafrir D, Etsion Y, Feitelson DG (2007). Backfilling using system-generated predictions rather than user runtime estimates. *IEEE Transactions on Parallel and Distributed Systems*, 18(6), 789–803. https://doi.org/10.1109/TPDS.2007.70606

[23] Unitywater (2026). Pr10179 specification for water supply temporary bypass, Rev. *15. Unitywater*, Maroochydore, QLD, Australia. https://www.unitywater.com/building-and-developing/-/media/unitywater/pdf-infrastructure-standards/pr10179---specification-for-water-supply-temporary-bypass.pdf

[24] US EPA (n.d.). EPANET: Application for modeling drinking water distribution systems. *US Environmental Protection Agency*, Washington, DC. https://www.epa.gov/water-research/epanet

[25] Van Eynde R, Vanhoucke M (2020). Resource-constrained multi-project scheduling: Benchmark datasets and decoupled scheduling. *Journal of Scheduling*, 23(3), 301–325. https://doi.org/10.1007/s10951-020-00651-w