

DOI: 10.5281/zenodo.23435442

GREEDY APPROXIMATION AND EXACT OPTIMISATION OF HISTORICAL LIGHTHOUSE COVERAGE IN ENGLAND, WALES AND FRANCE

Luis Eduardo Muñoz Guerrero

Universidad Tecnológica de Pereira, Pereira, Colombia; Email: lemunozg@utp.edu.co

Received: 08/12/2024
Accepted: 25/12/2024

Corresponding Author: Luis Eduardo Muñoz Guerrero
(lemunozg@utp.edu.co)

ABSTRACT

To measure how close the lighthouse networks of England and Wales (1693–1911) and France (1775–1929) came to optimal siting, and whether the centrally planned French programme of 1825 sited lights better than incremental English provision. Siting was formulated as a maximum coverage location problem over a 20 km sea band, solved exactly by integer programming and approximately by a lazy greedy algorithm with a $(1-1/e)$ guarantee and certified bounds, on two public datasets (603 and 1,139 light records) in a reproducible pipeline with perturbation replicates and 15 alternative specifications. With a common range of 18.5 km, networks too small to cover the band reached 0.50–0.77 of the optimum. English siting in 1693 was no better than random, and French waves before 1825 were near random (normalised efficiency 0.01 and 0.05); programme waves reached 0.53–0.55 with observed ranges, against 0.38 in England and Wales, a small, specification-dependent margin under a common range. At every benchmark, greedy solutions stayed within 1.1% of exact optima; common-range optima took at most 0.249 s. Conclusions: Under the French programme, construction moved from near random to closing about half the gap to optimal siting with observed ranges; exact benchmarks can audit infrastructure heritage.

KEYWORDS: Facility Location Problem, Submodular Set Function, Computational Geometry, Visibility Range, Maritime Cultural Heritage, Fresnel Lens, Trinity House, Commission Des Phares.

1. INTRODUCTION

The objective of this work is to measure, by exact optimisation and by greedy approximation with a provable guarantee, how much of their coastal sea the historical lighthouse networks of England and Wales (1693–1911) and France (1775–1929) covered compared with the best siting attainable with the same number of lights. In computational terms the question is a maximum coverage location problem: given the lights that a coastal state had built by a given date, how much sea could their best placement have covered, and how much did the actual network cover? The answer rests on exact optima computed by integer programming, a greedy algorithm with certified bounds, and a pipeline in which every reported number is produced by code.

Lighthouses are among the best documented maritime monuments, and their architecture, fuels and optics have long been studied as heritage (Hague and Christie, 1975; Naish, 1985). They are also landmarks of the history of science: the dioptric lens of Augustin Fresnel, first installed at Cordouan in 1823, changed the range and cost of coastal lights (Levitt, 2013). The two national systems examined here were organised in opposite ways. In France, the Commission des phares adopted in 1825 a programme for lighting the whole coast as a system, with lights of different orders (Commission des phares, 1825; Levitt, 2013). In England and Wales, Trinity House, private patentees and local harbour bodies all provided lights, and the remaining private lights passed to Trinity House only after an act of 1836 (Taylor, 2001; Bogart et al., 2022).

The English case is a classic of economics, where the lighthouse has served since Coase (1974) as the test of whether public goods can be provided privately (van Zandt, 1993; Bertrand, 2006; Candela and Geloso, 2018). A quantitative study by Bogart et

al. (2022) used the English and Welsh light data to compare the costs and fees of competing organisations around 1832 and the light dues of the reform period 1822–1836. That literature, however, evaluates organisations, not the spatial design of the networks they produced, and, to the author’s knowledge, no study has measured where the lights stood against where they could have stood.

The tools for such a measurement exist. The maximal covering location problem of Church and ReVelle (1974) is one of the central models of location science (ReVelle and Eiselt, 2005; Murray, 2016; Church and Murray, 2018), and its objective is a monotone submodular set function, for which greedy selection has a guaranteed approximation ratio (Nemhauser et al., 1978; Krause and Golovin, 2014).

Computational and spatial methods are also established in archaeology and heritage studies (Conolly and Lake, 2006; Brughmans, 2013). What has been missing is the combination: a formal statement of historical siting as an optimisation problem, exact and certified solutions, and a reproducible comparison between systems.

This paper provides that combination and makes four contributions. First, it formulates historical lighthouse siting as a maximum coverage location problem (MCLP) over a national sea band, together with metrics that separate siting from technology. Second, it analyses the complexity of the problem and derives a certified interval for the optimum that does not require an exact solver. Third, it measures the static and incremental siting efficiency of both systems at every benchmark in two public datasets and tests whether the French programme of 1825 sited lights better than incremental English provision. Fourth, it documents a tested pipeline (Figure 1) whose results can be regenerated with one command.

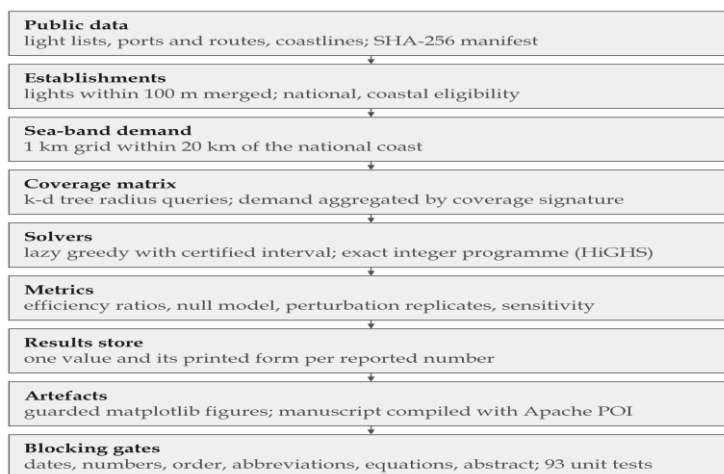


Figure 1: Architecture Of the Reproducible Pipeline. Each Stage Writes a Versioned Artefact; The Manuscript

Reads Numbers Only from the Results Store, And the Build Fails If Any Gate Fails.

2. DATA

The English and Welsh data are the light aids to navigation (LAN) dataset described by Dunn and Alvarez-Palau (2020) and deposited at the UK Data Service (Buxton-Dunn and Alvarez-Palau, 2020). It records 603 lights with coordinates, their presence at six benchmarks (medieval, 1693, 1753, 1831, 1851 and 1911) and, from 1831, their night-time visibility range.

The French data are the French light aids to navigation (F-LAN) dataset described by Litvine and Dunn (2021) and deposited at the UK Data Service

(Litvine and Dunn, 2021), with 1,139 records, presence at eight benchmarks between 1775 and 1929 (the data paper describes nine) and nominal ranges at six of them. The shipping routes described by Alvarez-Palau and Dunn (2019) and deposited at the UK Data Service (Alvarez-Palau et al., 2019) include flags that mark route segments visible at night in two benchmark years; they are used here only for validation. Table 1 lists the source behind each benchmark and whether it records ranges; the French ranges of 1775 and 1800 were estimated by the dataset authors rather than recorded.

Table 1: Historical Source Behind Each Benchmark of the Two Datasets and the Kind of Range Information It Provides.

Year	Source	Ranges
England and Wales		
1693	Great Britain's Coasting-Pilot, charts (Collins, 1693)	none
1753	Great Britain's Coasting-Pilot, charts (Collins, 1753)	none
1831	The Light-Houses of the British Islands (Hydrographic Office of Great Britain, 1832)	recorded
1851	The Light-Houses of the British Islands (Admiralty Hydrographic Department, 1851)	recorded
1911	The Admiralty List of Lights (Admiralty Hydrographic Department, 1911)	recorded
France		
1775	compiled from maps and memoirs (Fichou et al., 1999; Thévenard, 1799) and opening dates listed in 1825 and 1837	assumed
1800	compiled as for 1775	assumed
1825	Rapport contenant l'exposition du système adopté par la Commission des phares (Commission des phares, 1825)	none
1837	Description sommaire des phares et fanaux (Direction générale des Ponts et Chaussées et des Mines, 1837)	recorded
1846	Description sommaire des phares et fanaux (Direction générale des Ponts et Chaussées et des Mines, 1846)	recorded
1863	Mémoire sur l'éclairage et le balisage des côtes de France (Reynaud, 1864)	recorded
1872	État de l'éclairage et du balisage des côtes de France (Ministère des Travaux publics, 1872)	recorded
1882	État de l'éclairage et du balisage des côtes de France (Ministère des Travaux publics, 1882)	absent from the file
1929	État de l'éclairage des côtes de France (Ministère des Travaux publics, 1929)	derived only

Sources as documented by Dunn and Alvarez-Palau (2020) and Litvine and Dunn (2021).

Each English benchmark describes the lights of the year before publication; the LAN data paper dates the last two lists 1852 and 1912 in its text and 1851 and 1911 in its references, which are cited here. Recorded: nominal ranges given by the source; assumed: ranges estimated by the dataset authors from fuel, optics and height; derived only: visibility distances without nominal ranges.

Four properties of the published French file were checked in the data themselves. It contains no attributes for the benchmark of 1882 that the data paper describes and no ranges for 1825; for 1929 it gives no nominal ranges, only two columns of derived visibility distances that differ by a factor of 1,000; 11 records of 1929 have no coordinates and were excluded; and its counts of active lights equal

those of the data paper in two benchmarks and fall short by one light in seven. The analysis uses the published file as it stands; Table S6 gives the counts.

Lights less than 100 m apart were merged into one establishment, since the model places at most one light per site. An establishment belongs to a national system if the nearest national coast is closer than any foreign coast (with a tolerance of 2 km) and if it stands at sea or no more than 5 km inland. The rule removes the lights of the Channel Islands and the Isle of Man from the English and Welsh set, and lights far up rivers such as the Thames, the Avon, the Seine and the Loire, which served inland navigation (Table S5). Table 2 summarises the resulting establishments. Coastlines and land come from Natural Earth at scale 1:10m, version 5.1.2 (Natural Earth, 2022); the Global

Self-consistent Hierarchical High-resolution 1996) are used for a sensitivity analysis.
Geography (GSHHG) shorelines (Wessel and Smith,

Table 2: Light Records and Eligible Establishments by Benchmark, With the Number of Establishments That Have a Recorded Range and Their Median Nominal Range.

Recorded ranges					
Year	Records	Establ.	Eligible	Number	Median (km)
England and Wales					
1693	26	26	25	–	–
1753	35	35	33	–	–
1831	94	94	86	91	22.2
1851	179	177	163	150	18.5
1911	551	534	481	304	20.4
France					
1775	20	17	17	17	9.0
1800	28	27	27	27	9.0
1825	63	60	60	–	–
1837	103	98	98	98	16.7
1846	140	135	135	135	18.5
1863	285	274	268	274	18.5
1872	328	314	307	314	18.5
1929	799	746	697	–	–

Records: active rows of the published file; Establ.: establishments with coordinates after merging lights less than 100 m apart; Eligible: establishments of the national system (see text); Recorded ranges: establishments with a recorded range and their median range. A dash means that no range is recorded for that benchmark.

Table 3 lists the notation. The demand is the national sea band. With Λ_h the points of a square grid of spacing h km, L the land, Γ the national coast, Γ' the foreign coasts and d the Euclidean distance, the demand points are the grid points at sea that lie within D km of the national coast and whose nearest coast is national,

3. METHODS

$$I = \{i \in \Lambda_h \setminus L: d(i, \Gamma) \leq \min\{D, d(i, \Gamma')\}\} \quad (1)$$

each weighted by the area w_i of its cell. Measuring coverage as sea area rather than coastline length avoids the dependence of length on resolution (Mandelbrot, 1967) and follows work on spatial

representations of coverage demand (Murray et al., 2008; Tong and Murray, 2009). Figure 2 maps the two bands, the candidate sites and the establishments excluded from each national system.

Table 3: Notation Used in the Formulation and the Metrics.

Symbol	Meaning
I, w_i	demand points of the sea band (Equation 1) and their weights (cell areas, km ²)
J, m	candidate sites (historical establishments) and their number
d, D, h	Euclidean distance, band width and grid spacing (km)
$r, N_i(r)$	light range (km) and set of sites within r of point i
x_j, y_i	binary site variables and coverage variables of the integer programme
p, S	number of lights and a set of sites
$f(S)$	area covered by S (km ² ; Equation 2)
$\text{OPT}(p), \text{psat}$	maximum of f over sets of p sites; smallest p whose optimum covers the coverable band $f(j)$
A_t, p_t	historical network at benchmark t and its size
R_t, π_t	random network of p_t sites and share of random networks that A_t outperforms
p_t, v_t, λ_t	static, normalised and lights-needed efficiencies
K, N, OPTK	kept and new sites of a wave and best coverage with $ N $ new sites
η, η_n	wave efficiency and its normalised form
$C, r \square c, p_c$	range classes of the observed-range lens, their ranges and their numbers of lights
$\Delta_j(S)$	marginal gain of site j given S (Equation 22)
zLP, γ	optimal value of the linear relaxation of the integer programme; certified gap

A light at site j with range r covers the demand points within distance r of it. For a set S of sites, the

covered area is

$$f(S) = \sum_{i \in I} w_i 1[S \cap N_i(r) \neq \emptyset] \quad (2)$$

where the neighbourhood of point i is

$$N_i(r) = \{j \in J: d(i, j) \leq r\} \quad (3)$$

and $N(j) = \{i \in I: j \in N_i(r)\}$ is the set of points that site j covers. The MCLP selects p sites to maximise f .

Following Church and ReVelle (1974), with binary site variables x_j and coverage variables y_i ,

$$\max_{x, y} \sum_{i \in I} w_i y_i \quad (4)$$

$$y_i \leq \sum_{j \in N_i(r)} x_j \quad \forall i \in I \quad (5)$$

$$\sum_{j \in J} x_j = p, \quad x_j \in \{0, 1\}, \quad 0 \leq y_i \leq 1 \quad (6)$$

4–6 gives $\text{OPT}(p)$. For a benchmark t with network A_t of $p_t = |A_t|$ establishments, static siting efficiency

$$\rho_t = \frac{f(A_t)}{\text{OPT}(p_t)} \quad (7)$$

Because ρ_t rises mechanically as the band saturates, it is complemented by an efficiency

$$v_t = \frac{f(A_t) - E[f(R_t)]}{\text{OPT}(p_t) - E[f(R_t)]} \quad (8)$$

so that v_t is zero for random and one for optimal siting. With $m = |J|$, a demand point escapes a random network only if none of the $|N_i(r)|$ sites that

cover it is drawn, so the expectation has the closed form

$$E[f(R_t)] = \sum_{i \in I} w_i \left(1 - \frac{\binom{m - |N_i(r)|}{p_t}}{\binom{m}{p_t}} \right) \quad (9)$$

The analysis estimates it from $M = 2,000$ random networks $R_t^{(\ell)}$, which also give the share of random

$$\pi_t = \frac{1}{M} \sum_{\ell=1}^M 1[f(R_t^{(\ell)}) < f(A_t)] \quad (10)$$

Static efficiency is also described by the share of lights an optimal planner would have needed to

reach the same coverage,

$$\lambda_t = \frac{1}{p_t} \min\{p: \text{OPT}(p) \geq f(A_t)\} \quad (11)$$

The coverable band $f(J)$ is the area within range of at least one candidate site, and the smallest network

whose optimum covers it,

$$p_{\text{sat}} = \min\{p: \text{OPT}(p) = f(J)\} \quad (12)$$

separates unsaturated networks ($p_t < p_{\text{sat}}$) from saturated ones. Because $f(A_t) \leq f(J)$, λ_t never exceeds p_{sat}/p_t , so in a saturated network a small λ_t , like a p_t near one, is partly mechanical.

evaluated conditionally on the lights already in place. If K is the set of establishments kept from t to $t+1$ and N the set of new ones, the best coverage attainable with the same number of new lights is

Construction between two benchmarks is

$$\text{OPT}_K = \max_{S \subseteq J \setminus K, |S|=|N|} f(K \cup S) \quad (13)$$

the optimum of Equations 4–6 with $p = |K| + |N|$ is

$$\eta = \frac{f(K \cup N) - f(K)}{\text{OPT}_K - f(K)} \quad (14)$$

and its normalised version, with R a set of $|N|$ sites drawn uniformly from $J \setminus K$, is

$$\eta_n = \frac{f(K \cup N) - E[f(K \cup R)]}{\text{OPT}_K - E[f(K \cup R)]} \quad (15)$$

which, like v_t , is zero for random and one for optimal additions. Its expectation is $f(K)$ plus the sum of Equation 9 over the points that K leaves uncovered, with $J \setminus K$ in place of J and $|N|$ in place of p_t ; the analysis estimates it from 500 random additions per wave, and replacing the Monte Carlo means by the closed forms would change no common-range v_t or η_n by more than 0.013.

Identity between benchmarks is resolved by

$$\psi^* = \arg \min_{\psi \in \Psi_\varepsilon} \sum_{(j,j') \in \psi} d(j, j') \quad (16)$$

matching the establishments of consecutive benchmarks, because the sources do not always give an unchanged light the same coordinates: between 1872 and 1929, 105 of the 258 French establishments kept moved by more than 100 m (median 0.42 km; Table S7 gives every wave).

Among the matchings Ψ_ε that pair as many establishments as possible within $\varepsilon = 1$ km, the one of least total displacement is chosen (Kuhn, 1955),

and K holds the establishments of $t+1$ that it pairs, N the others.

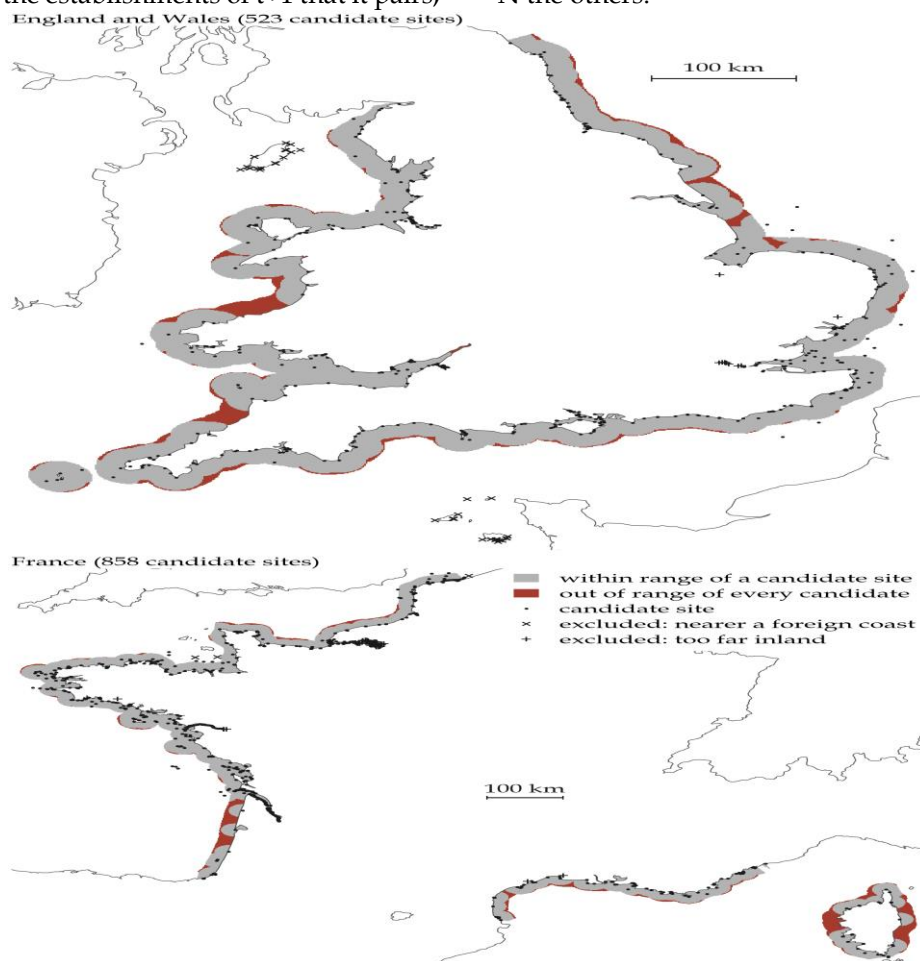


Figure 2: Sea Bands of England and Wales (First Panel) And France (Second Panel), With the Parts Within and Out of the Common Range of the Candidate Sites, The Candidate Sites and the Excluded Establishments.

Two lenses are used. The common-range lens gives every light the same range r , so that ρ_t , ν_t , η and η_n measure siting alone.

The observed-range lens keeps the recorded range

$$\bar{r}_c = \text{med}\{r_j: q_{c-1} < r_j \leq q_c\} \quad (17)$$

where q_0 is zero and the last quantile is infinite. It re-replaces the same multiset of lights optimally: with

$$y_i \leq \sum_{c \in C} \sum_{j \in N_i(\bar{r}_c)} x_{jc} \quad \forall i \in I \quad (18)$$

$$\sum_{j \in I} x_{jc} = p_c \quad \forall c, \quad \sum_{c \in C} x_{jc} \leq 1 \quad \forall j \quad (19)$$

so that each site holds at most one light (Table S3). With observed ranges, ρ_{obs} and η_n follow Equations 7 and 15 with the class ranges; in a wave the kept lights stay fixed with their classes, and random additions keep the classes of the new lights.

The recorded ranges take few distinct values (Figure 3): the French ranges of 1775 and 1800, estimated by the dataset authors, take only two and three values, and in 1911 210 of the 481 English and Welsh establishments have no recorded range and receive the median.

Measurement error is propagated by $B = 200$ perturbation replicates. In each replicate every site is displaced by an independent Gaussian error and the range is rescaled by a common random factor,

$$\omega = \frac{1}{B} \sum_{b=1}^B \mathbf{1}[\bar{\eta}_n^{(b)}(G_1) > \bar{\eta}_n^{(b)}(G_2)] \quad (21)$$

Table 4 lists every analytic parameter, the reason for its value and the alternatives tested in the

r_j of each light, imputing the median recorded range where none is recorded, and groups the ranges into four classes $c \in C$ bounded by their quantiles q_c of order $c/|C|$, each represented by its median range,

a binary variable x_{jc} for a light of class c at site j and p_c lights in class c , it maximises Equation 4 subject to

$$\tilde{P}_j = P_j + \sigma \xi_j, \quad \tilde{r} = (1 + \zeta) r \quad (20)$$

where P_j is the position of site j , ξ_j a standard bivariate normal vector, $\sigma = 0.5$ km and ζ uniform between $-\alpha$ and α , with $\alpha = 10\%$; the demand grid is also shifted by a uniform offset within one cell.

The class ranges of the observed-range lens are rescaled by the same factor, and identity between benchmarks is taken from the unperturbed coordinates.

Intervals are the central 95% of the replicate values, and groups of waves G_1 and G_2 are compared through the share of replicates in which the mean normalised efficiency η_n of the first exceeds that of the second,

sensitivity analysis.

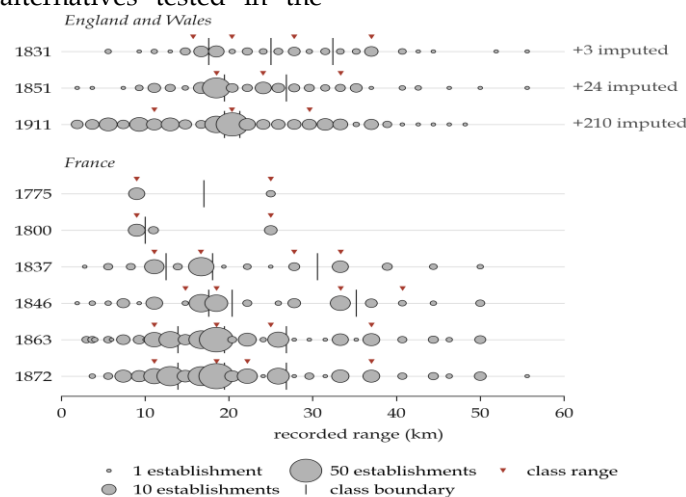


Figure 3: Recorded Ranges of the Observed-Range Lens at Each Benchmark. The Area of Each Circle Is Proportional to the Number of Establishments with That Range; Ticks Separate the Quantile Classes and Triangles Give the Range Assigned to Each Class. Establishments Without a Recorded Range, Which Receive the Median, Are Counted on the Right.

Table 4: Analysis Parameters, The Reason for Each Value and the Alternative Values Examined in the Sensitivity Analysis.

Parameter	Value	Reason	Also tested
Common range r	18.5 km (10 nautical miles)	median recorded range in England and Wales in 1851 and in France in 1846–1872 (Table 2)	9.3, 27.8 and 37.0 km
Band width D	20 km	about one common range from the coast	10 and 30 km
Grid spacing h	1 km	well below the median spacing of historical lights (6.2–10.7 km)	2 km
Merging distance	100 m	joins the lights of one establishment	50 and 250 m
Foreign-coast tolerance	2 km	keeps national headlands near a border	–
Inland limit	5 km	keeps estuary lights and drops river navigation lights	10 km; no limit
Candidate sites	historical sites	the planner chooses among sites used at some date	sites built by t ; coastal grid of 5 km
Demand	sea area	does not depend on the resolution of the coastline	coastline length
Coastline	Natural Earth	national boundaries consistent with the coastline	GSHHG
Matching tolerance	1 km	unchanged lights shift by a few hundred metres between lists (Table S7)	–
Range classes	4 quantile classes	keeps the observed-range ILP small	–
Random networks	2,000 (static), 500 (waves)	null distribution of v and η	–
Perturbation	200 replicates; $\sigma = 0.5$ km; range $\pm 10\%$; grid offset	uncertainty of historical positions and ranges	–
Intervals	central 95%	percentiles of the perturbation replicates	–
Random seeds	fixed	makes reruns reproducible	–

4. COMPLEXITY AND GUARANTEES

Proposition 1. The MCLP is NP-hard. Proof sketch. Given an instance of maximum coverage with sets $T_1, \dots, T_m$ over a ground set U and a budget p , create one demand point of unit weight per element of U and one candidate site j whose coverage set is T_j ; an optimal MCLP solution is an optimal maximum coverage solution. Maximum coverage generalises set cover, which is NP-complete (Karp, 1972; Garey

and Johnson, 1979); the MCLP is also NP-hard on general networks (Megiddo et al., 1983).

Moreover, no polynomial algorithm approximates maximum coverage within a factor better than $1 - 1/e$ unless $P = NP$ (Feige, 1998). The practical implication is that exact solutions can require exponential time in the worst case, so the size and structure of the instances at hand must be measured, not assumed.

For a set S and a site $j \notin S$, the marginal gain of j is

$$\Delta_j(S) = f(S \cup \{j\}) - f(S) \quad (22)$$

Lemma 1. The coverage function of Equation 2 is monotone and submodular: for $S \subseteq T \subseteq J$ and $j \in J \setminus T$,

$$\Delta_j(S) \geq \Delta_j(T) \geq 0 \quad (23)$$

Proof. The gain of adding j to S is the weight of the points covered by j and by no member of S ; enlarging S to T can only shrink that set, and weights are non-negative.

Theorem 1 (Nemhauser et al., 1978). The greedy algorithm, which repeatedly adds the site of largest marginal gain, returns a set S_g with

$$f(S_g) \geq \left(1 - \frac{1}{e}\right) \text{OPT}(p) \quad (24)$$

Proposition 2 (certified interval). For the greedy set S_g of size p , let top_p be the p sites with the largest marginal gains $\Delta_j(S_g)$, zLP the value of the linear

programming (LP) relaxation of Equations 4–6, $U_1 = f(S_g)/(1 - 1/e)$, $U_3 = \text{zLP}$ and

$$U_2 = f(S_g) + \sum_{j \in \text{top}_p} \Delta_j(S_g) \quad (25)$$

Then

$$f(S_g) \leq \text{OPT}(p) \leq \min\{U_1, U_2, U_3\} \quad (26)$$

Proof. The lower bound holds because S_g is feasible, and U_1 restates Theorem 1. For U_2 , by monotonicity and submodularity, $\text{OPT}(p) \leq f(S_g \cup O) \leq f(S_g) + \sum_{j \in O} \Delta_j(S_g)$ for an optimal set O of p sites, and the sum over O cannot exceed the sum

$$\gamma = \frac{\min\{U_1, U_2, U_3\} - f(S_g)}{f(S_g)} \geq \frac{\text{OPT}(p) - f(S_g)}{f(S_g)} \quad (27)$$

an upper bound on the shortfall of the greedy set, relative to its value, that holds without knowing $\text{OPT}(p)$. For the observed-range lens, the site-class pairs form a partition matroid by class, and greedy selection attains at least half of the optimum (Fisher et al., 1978); placing two lights on one site never increases coverage, so the one-light-per-site rule does not lower that optimum.

5. ALGORITHMS AND IMPLEMENTATION

Algorithm 1 is the lazy greedy procedure of

$$m + p - 1 \leq n_{\text{lazy}} \leq n_{\text{plain}} = pm - \frac{p(p-1)}{2} \quad (28)$$

The saving of the lazy version therefore depends on the data, whereas the cost of the plain one is fixed by m and p .

of the p largest gains (Leskovec et al., 2007). U_3 is valid because the LP is a relaxation of the ILP.

The interval certifies the quality of a greedy solution without an exact solver; its width is reported as the certified gap

Minoux (1978) with the certificate of Proposition 2. Its invariant is that every key in the priority queue is an upper bound of the current marginal gain of its site; by Lemma 1 gains can only fall as the solution grows, so a site whose freshly computed gain still tops the queue is the true greedy choice. The plain greedy algorithm recomputes every remaining gain at each of the p steps, whereas the lazy one computes each gain once at the start, at least one fresh gain at every later step and never the same gain twice in one step, so the numbers of gains they compute satisfy

Algorithm 1. Lazy greedy with certified interval (input: coverage sets $N(j)$, weights w , size p ; output: S_g , interval for $\text{OPT}(p)$)

Algorithm 1. Lazy greedy with certified interval (input: coverage sets $N(j)$, weights w , size p ; output: S_g , interval for $\text{OPT}(p)$)	
11	$S \leftarrow \emptyset$; $U \leftarrow$ all demand points (uncovered)
22	for each site j : push $(\Delta_j = \sum_{i \in N(j)} w_i, j, \text{stamp } 0)$ into a max-priority queue Q
33	while $ S < p$:
44	pop $(\delta, j, \text{stamp})$ with the largest key from Q
55	if stamp = $ S $: add j to S ; remove $N(j)$ from U (invariant: keys $\geq$ current gains)
66	else: recompute $\Delta_j = \sum_{i \in N(j) \cap U} w_i$ and push (Δ_j, j, S)
77	compute the p largest current gains and zLP
88	return $S_g = S$ and the interval of Equation 26

The exact optimum is computed with the HiGHS solver (Huangfu and Hall, 2018) through SciPy (Virtanen et al., 2020).

Coverage sets are obtained with radius queries on k -d trees (Bentley, 1975). Before optimisation,

$$f(S) = \sum_{\theta \in \Theta} w_\theta \mathbf{1}[S \cap \theta \neq \emptyset] \quad (29)$$

for every set S , because points that share a coverage set are covered or uncovered together by every solution; unlike the lossy demand aggregation studied by Current and Schilling (1990), the reduction is therefore exact, and points that no site covers drop out. The reduction factor

demand points with identical coverage sets are merged into one weighted point. With Θ the distinct non-empty sets $N_i(r)$ and W_Θ the total weight of the points whose set is θ ,

$|I|/|\Theta|$ was 17.2 for England and Wales and 10.9 for France. Identity between benchmarks uses the assignment algorithm of Crouse (2016). Table 5 gives the size of every instance and the effort of its exact and greedy solutions.

Table 5: Size Of Each Instance and Computational Effort of the Exact Solution and Of the Plain and Lazy Greedy Algorithms Under the Common Range.

Evaluations Time (s)							
Year	p	ILP (s)	Plain	Lazy	Plain	Lazy	Speed-up
England and Wales: 54,905 demand points, 3,184 after aggregation; 523 candidate sites							
1693	25	0.080	12,775	588	0.056	0.003	22×
1753	33	0.072	16,731	651	0.075	0.002	30×
1831	86	0.060	41,323	1,282	0.178	0.006	29×
1851	163	0.033	72,046	1,826	0.306	0.009	36×
1911	481	0.029	136,123	2,158	0.552	0.013	42×
France: 61,044 demand points, 5,610 after aggregation; 858 candidate sites							
1775	17	0.138	14,450	947	0.063	0.005	13×
1800	27	0.166	22,815	1,035	0.102	0.006	18×
1825	60	0.236	49,710	1,558	0.221	0.008	29×
1837	98	0.249	79,331	2,098	0.349	0.011	30×
1846	135	0.207	106,785	2,669	0.466	0.014	34×
1863	268	0.069	194,166	3,111	0.810	0.016	51×
1872	307	0.062	216,435	3,150	0.883	0.018	49×
1929	697	0.056	355,470	3,540	1.44	0.022	66×

Evaluations: marginal gains computed by the greedy algorithm (Equation 28); the plain version recomputes every site at each step. Times are medians of repeated runs (three for the plain and five for the lazy greedy); speed-up: plain over lazy time.

The pipeline is organised as the stages of Figure 1. Raw files are downloaded by a script that records their size and SHA-256 hash; each analysis writes its results, with the printed form of every number, to a results store, and the manuscript is compiled from that store.

The build runs 93-unit tests, including

comparisons of the greedy, lazy greedy and ILP routines against exhaustive search on random instances, and blocking gates that check dates, the provenance of every number in the text, the order of figures and tables, abbreviations and equation numbering.

Random draws use fixed seeds; a second run of the data, static, wave and observed-range analyses reproduced all 756 of their values.

The software versions and hardware are listed in Table 6, together with the measured time and memory of each stage.

Table 6: Measured Wall-Clock Time and Peak Resident Memory of Each Pipeline Stage on the Author's Workstation.

Stage	Time (s)	Memory (MB)
Sea band and eligible sites, England and Wales	52.2	129
Sea band and eligible sites, France	44.2	184
Coverage sets and demand aggregation, England and Wales	0.066	123
Coverage sets and demand aggregation, France	0.093	137
Static analysis, all benchmarks	139	184
Construction waves	94.0	185
Observed-range analysis	206	343
Perturbation replicates (parallel)	5,582	10,093
Sensitivity analysis	2,378	222
Route validation	0.369	120
Greedy timing experiment	103	180

Hardware: AMD Ryzen 7 9800X3D 8-Core Processor (16 threads) with 15.6 GB of memory; the graphics processor was not used. Software: Python 3.12, NumPy 1.26.2 (Harris et al., 2020), SciPy 1.11.4 with HiGHS, matplotlib 3.8.2 (Hunter, 2007) and Apache POI 5.2.5. The parallel stage used 12 processes; the perturbation and sensitivity stages ran at the same time, so their times are upper bounds.

6. RESULTS

Table 7 reports the static results under the common-range lens (18.5 km, that is 10 nautical miles; a band of 20 km and cells of 1 km).

The coverable band measured 50,292 km² in England and Wales and 53,109 km² in France; another 4,613 and 7,935 km² of the bands lay out of range of every candidate site (Figure 2), so no siting of the historical establishments could cover them.

In every instance the lazy greedy value was

within 1.1% of the exact optimum and the certified interval of Proposition 2 contained the exact value; the LP bound equalled the integer optimum in 11 of the 13 instances and exceeded it by at most 0.003% in

the other two. In unsaturated networks ($p_t < p_{sat}$; Equation 12), ρ ranged from 0.50 to 0.77. The areas behind these ratios are in Table S1.

Table 7: Static Siting Efficiency by Benchmark: Establishments P , Efficiency P Under the Common Range with Its 95% Perturbation Interval, Normalised Efficiency N , Share Λ Of Lights an Optimal Planner Would Have Needed, Efficiency P_{obs} With the Observed Ranges, And Value of the Lazy Greedy Set Relative to the Exact Optimum.

Year	p	ρ	95% int.	v	λ	ρ_{obs}	Greedy
<i>England and Wales</i>							
1693	25	0.50	0.50–0.51	–0.02	0.48	–	1.000
1753	33	0.59	0.58–0.59	0.17	0.55	–	0.999
1831	86	0.64	0.63–0.66	0.22	0.45	0.78	0.993
1851	163	0.81	0.78–0.84	0.36	0.36	0.92	1.000
1911	481	1.00	0.99–1.00	0.74	0.25	0.99	1.000
<i>France</i>							
1775	17	0.68	0.67–0.70	0.38	0.71	0.65	0.998
1800	27	0.61	0.61–0.62	0.25	0.59	0.63	0.994
1825	60	0.55	0.54–0.55	0.12	0.45	–	0.992
1837	98	0.63	0.61–0.65	0.20	0.42	0.70	0.989
1846	135	0.77	0.73–0.80	0.41	0.42	0.90	0.998
1863	268	0.89	0.86–0.91	0.47	0.27	0.96	1.000
1872	307	0.91	0.89–0.93	0.50	0.25	0.99	1.000
1929	697	1.00	0.99–1.00	0.94	0.19	–	1.000

Common range of 18.5 km (10 nautical miles), band of 20 km and cells of 1 km. v uses 2,000 random networks and the intervals 200 perturbation replicates. p_{obs} : static efficiency with the observed ranges (a dash where ranges are not recorded). Greedy: coverage of the lazy greedy set divided by the exact optimum.

The certificate of Proposition 2 was tight in practice (Table 8): the LP bound U_3 was never larger than the other two and kept the certified gap γ at or below 1.12%, whereas U_2 and above all U_1 were

looser. Along the whole greedy curve, not only at the historical sizes, γ reached at most 1.38% (France, 75 lights), exceeded 0.1% only for networks of 6 to 147 lights and vanished once the greedy set covered the whole coverable band (Figure 4). At the 27 sizes where the exact optimum was also computed, γ differed from the actual gap by at most 0.01 percentage points: the LP bound was practically exact, so the certificate measured the true shortfall of the greedy set.

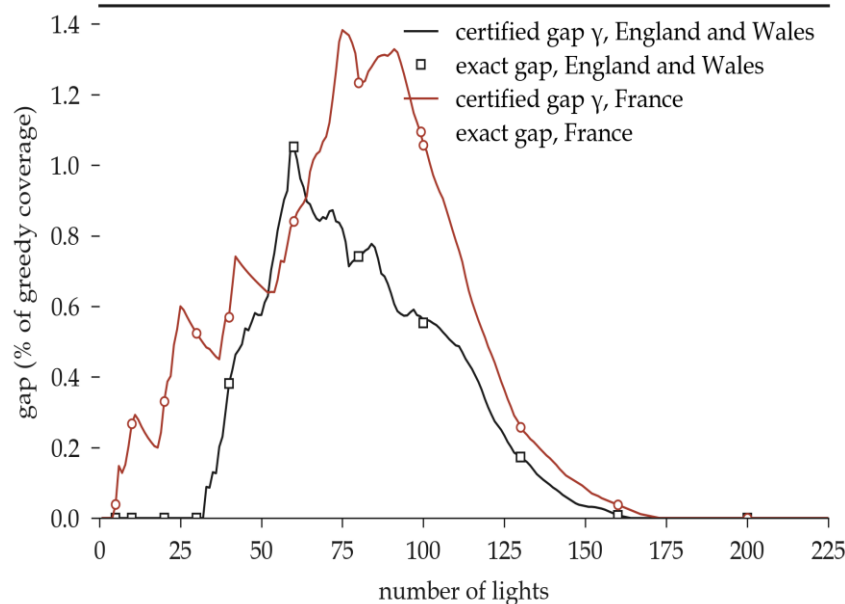


Figure 4: Certified Gap γ Of the Lazy Greedy Set at Every Network Size (Lines) And Actual Gap Where the Exact Optimum Was Computed (Markers), Under the Common Range; γ Is Zero for All Larger Networks.

Table 8: Certified Interval of Proposition 2 For Every Instance: Greedy Value, The Three Upper Bounds, Exact Optimum and the Largest Shortfall the Certificate Leaves Open.

Year	f(Sg)	U1	U2	U3	OPT	γ
England and Wales						
1693	21,465	33,957	38,604	21,465	21,465	0.00%
1753	27,162	42,970	46,194	27,186	27,186	0.09%
1831	47,338	74,888	54,435	47,686	47,686	0.74%
1851	50,291	79,559	50,292	50,292	50,292	0.00%
1911	50,292	79,561	50,292	50,292	50,292	0.00%
France						
1775	15,708	24,850	28,557	15,740	15,740	0.20%
1800	23,277	36,824	41,513	23,410	23,410	0.57%
1825	42,347	66,992	61,126	42,703	42,703	0.84%
1837	51,112	80,858	57,440	51,686	51,684	1.12%
1846	52,921	83,720	53,675	53,030	53,030	0.21%
1863	53,109	84,017	53,109	53,109	53,109	0.00%
1872	53,109	84,017	53,109	53,109	53,109	0.00%
1929	53,109	84,017	53,109	53,109	53,109	0.00%

Areas in km² under the common range. U1 = $f(Sg)/(1 - 1/e)$ (Theorem 1); U2: $f(Sg)$ plus the p largest marginal gains; U3: value of the LP relaxation; γ : certified gap (Equation 27).

In England and Wales, ρ rose from 0.50 in 1693 to 0.81 in 1851. The early network was no better than random: its coverage exceeded that of only 44.6% of random networks of the same size, and v was -0.02 . Normalised efficiency then grew to 0.17 in 1753, 0.22 in 1831 and 0.36 in 1851. The French network of 1775 was better sited than random networks of its size (its coverage exceeded that of 100.0% of them; $v = 0.38$), but efficiency fell to 0.12 by 1825 and recovered to 0.20 in 1837, 0.41 in 1846 and 0.47 in 1863.

At comparable sizes the two systems were close: 86 English and Welsh establishments in 1831 reached $\rho = 0.64$, and 98 French ones in 1837 reached 0.63; with the observed ranges the same networks reached 0.78 and 0.70.

The French figure of 1837 still reflects the network inherited from before the programme, which supplied 59 of its 98 establishments; the wave measures below isolate what was built. The optimal frontier explains why ρ tends to one later in the period (Figure 5). Under the common range, ρ_{sat} was 156 in England and Wales and 160 in

France: that many optimally sited lights would have covered the whole coverable band.

The expected coverage of random networks (Equation 9) rises far more slowly: at ρ_{sat} lights it reached 69% of the coverable band in England and Wales and 66% in France, and the English and Welsh network of 1693 lay on the random curve ($v = -0.02$). Both networks exceeded those sizes after mid-century, so the English and Welsh network of 1911 and the French network of 1929 reached $\rho = 1.00$ and 1.00 with far more lights than coverage required ($\lambda = 0.25$ and 0.19, below the bounds of 0.32 and 0.23 that saturation sets); by their names alone, at least 114 of the 481 English and Welsh and 144 of the 697 French establishments were harbour, pier or jetty lights, whose purpose was local (Table S8).

Figure 6 compares the historical and the optimal networks of equal size in 1831 and 1837. In both countries the historical lights clustered, often at harbours and estuaries, while the optimal sites spread along open coasts, where the uncovered band lay: the median distance from each light to its nearest neighbour was 10.7 km in the historical and 23.9 km in the optimal network of England and Wales, and 6.2 and 22.5 km in France.

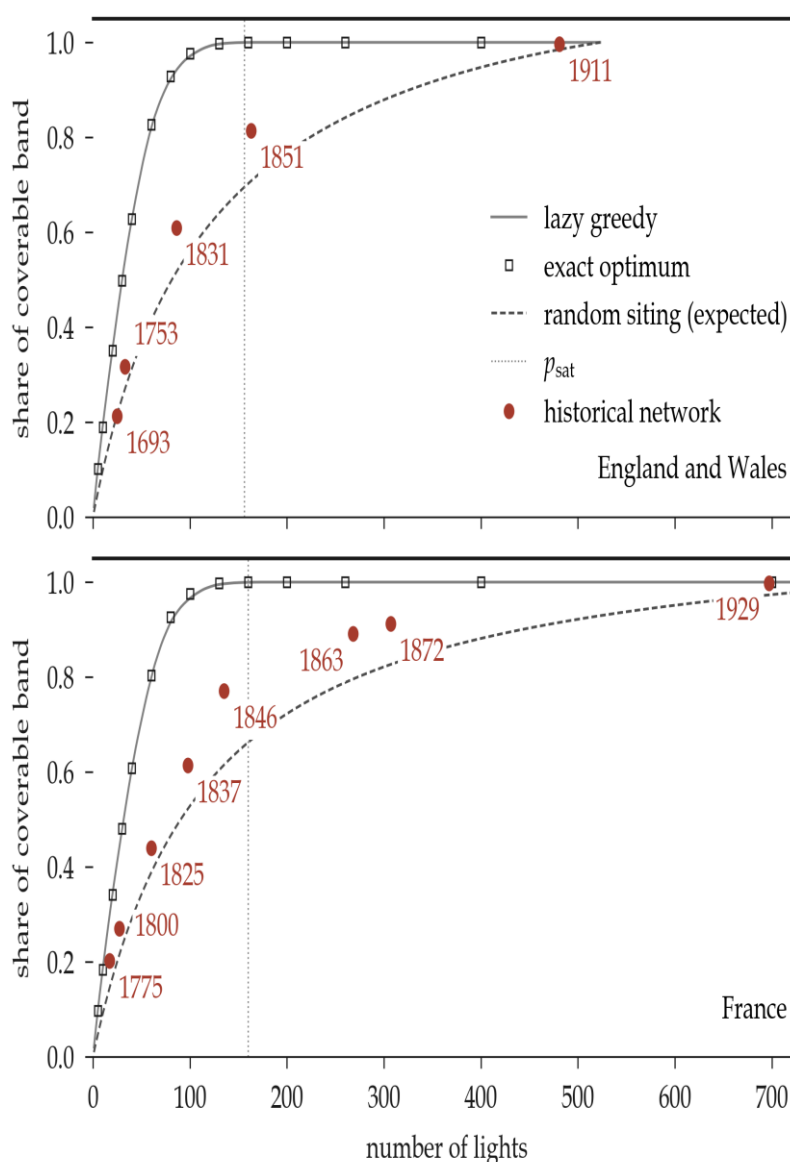


Figure 5: Optimal Coverage as a Function of the Number of Lights (Lazy Greedy Curve and Exact Optima), Expected Coverage of Random Networks of the Same Size and the Historical Networks at Each Benchmark, As Shares of the Coverable Band; The Dotted Line Marks p_{sat} .

Table 9 summarises the construction waves. Under the common range, the two French waves before 1825 were indistinguishable from random additions ($\eta n = 0.01$ and 0.05 ; better than 55.8% and 75.6% of random additions), whereas the English and Welsh waves up to 1851 reached 0.23, 0.13 and 0.16.

The French waves after 1825 reached 0.13, 0.32 and 0.17. With observed ranges the contrast widens: the waves that executed the programme reached 0.53 and 0.55, against 0.38 for the English and Welsh wave of 1831–1851 and 0.09 for the French wave of 1775–

1800, whose ranges were estimated by the dataset authors. The coverage gains behind these ratios are in Table S2. Figure 7 maps the two French waves on either side of the programme under the common range: before it, the new lights clustered on the Channel coast, whereas the optimal additions of the same number would have gone mainly to Brittany, Corsica and the Mediterranean coast; during it, the new lights went mainly to Brittany, and the area that only the optimal additions would have covered fell from 20,337 to 9,576 km².

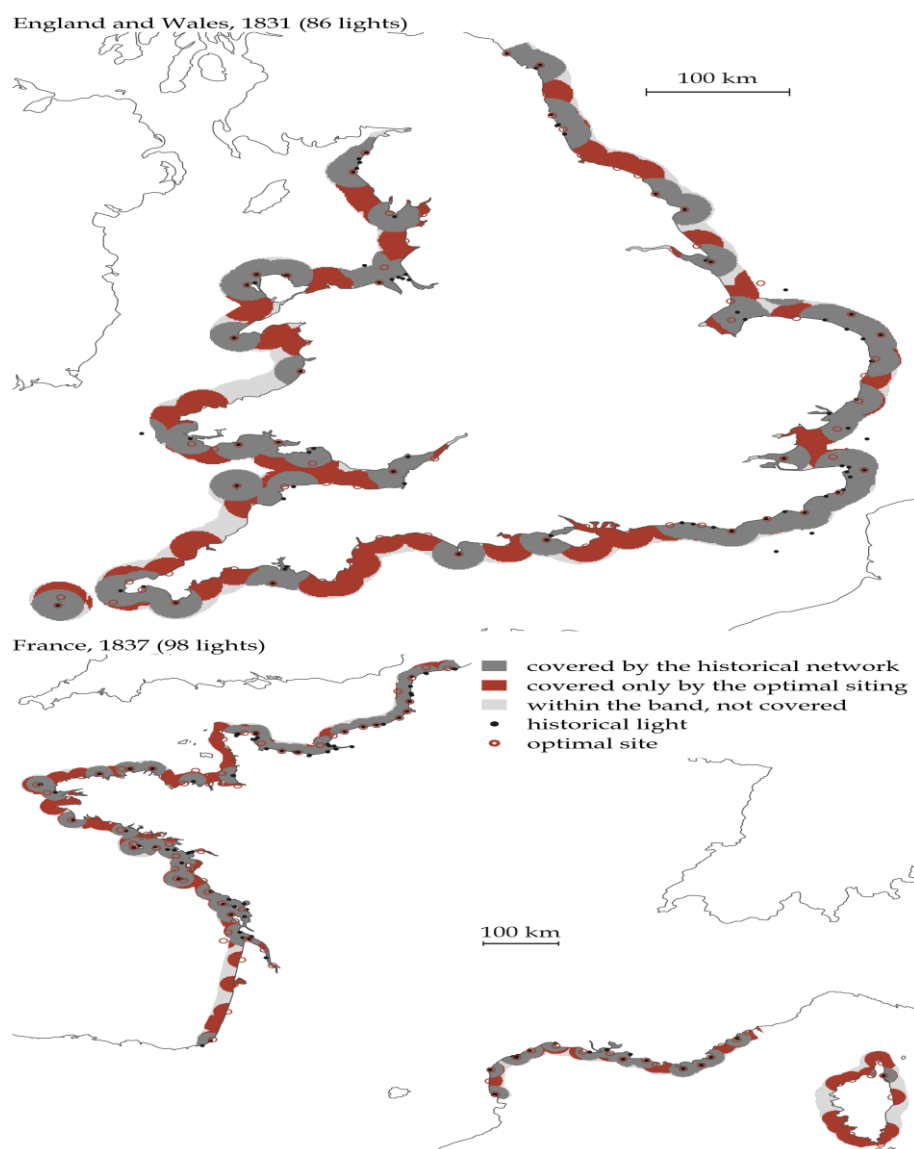


Figure 6: Sea Band Covered by the Historical Network and by the Optimal Siting with the Same Number of Lights, England and Wales in the First Panel and France in the Second, Under the Common Range.

Table 9: Construction Waves: Establishments Kept and Added, And Normalised Wave Efficiency η_n (Zero for Random Additions, One for Optimal Ones) Under the Common Range and the Observed Ranges, With 95% Perturbation Intervals.

Common Observed						
Wave	Kept	New	η_n	95% int.	η_n	95% int.
England and Wales						
1693–1753	22	11	0.23	0.20–0.24	–	–
1753–1831	23	63	0.13	0.12–0.15	–	–
1831–1851	85	78	0.16	0.14–0.18	0.38	0.33–0.46
1851–1911	150	331	0.65	0.56–0.69	0.64	0.54–0.78
France						
1775–1800	15	12	0.01	0.00–0.04	0.09	0.07–0.15
1800–1825	24	36	0.05	0.04–0.07	–	–
1825–1837	59	39	0.13	0.11–0.14	–	–
1837–1846	98	37	0.32	0.30–0.34	0.53	0.47–0.58
1846–1863	135	133	0.17	0.13–0.24	0.55	0.50–0.56
1863–1872	264	43	0.08	0.06–0.11	0.66	0.44–0.83
1872–1929	258	439	0.88	0.70–0.86	–	–

Kept and new establishments are matched within 1 km. A dash means that ranges are not recorded at both ends of the wave. Intervals are the central 95% of the perturbed replicates and need not contain the unperturbed estimate when the potential gain of a wave is small, as in the saturated French wave of 1872–1929.

Across the perturbation replicates the contrasts

kept their sign (Table 10). With observed ranges, the programme waves exceeded the English and Welsh wave in $\omega = 100.0\%$ of replicates (Equation 21) and the French wave before the programme in 100.0% . Under the common range they exceeded the pre-programme French waves in 100.0% of replicates, but the English and Welsh waves only by a small margin (difference 0.04).

Table 10: Normalised Wave Efficiency of Groups of Waves and Differences Between Them Across the Perturbation Replicates, With the Observed Ranges and Under the Common Range.

Quantity (η_n)	Median	95% int.	Positive
Observed ranges			
Programme waves, 1837–1863 (mean)	0.53	0.49–0.57	100.0%
English and Welsh wave, 1831–1851	0.38	0.33–0.46	100.0%
French wave, 1775–1800	0.11	0.07–0.15	100.0%
Programme – English and Welsh	0.14	0.06–0.21	100.0%
Programme – French before	0.42	0.35–0.49	100.0%
Common range			
Programme waves, 1825–1863 (mean)	0.21	0.19–0.23	100.0%
French waves before 1825 (mean)	0.04	0.02–0.05	100.0%
English and Welsh waves, 1693–1851 (mean)	0.17	0.16–0.18	100.0%
Programme – French before	0.17	0.15–0.20	100.0%
Programme – English and Welsh	0.04	0.02–0.07	100.0%
English and Welsh – French before	0.13	0.12–0.15	100.0%

Normalised wave efficiency across 200 perturbation replicates; differences are computed within each replicate. Positive: share of replicates with a value above zero, which for a difference is ω of Equation 21.

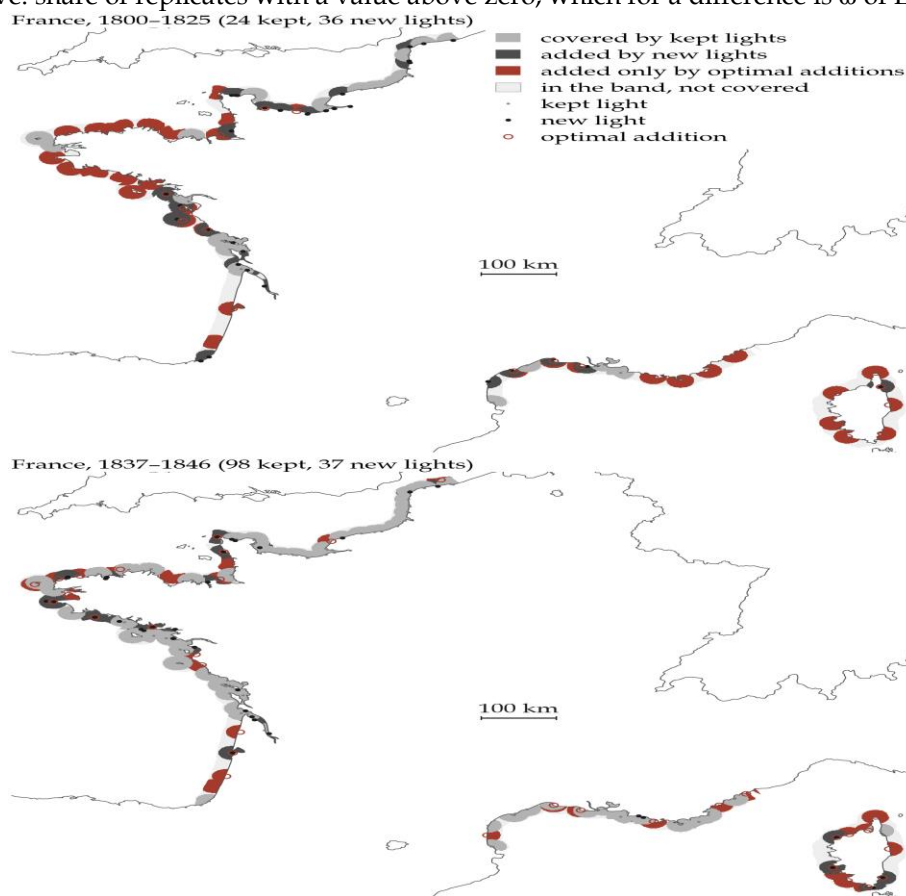


Figure 7: French Waves Before (1800–1825, First Panel) And During (1837–1846, Second Panel) The Programme, Under the Common Range: Coverage of the Kept Lights, Of the New Lights and of the Optimal Additions of the Same Number.

Two independent checks support the geometry. A route segment s counts as visible at night if it passes

$$v_s = 1[\exists j: d(s, j) \leq r_j] \quad (30)$$

where j runs over the lights of the benchmark year and $d(s, j)$ is the distance from the segment to the light. Computed in this way for the 2,652 route segments and the 94 lights of 1831, with their recorded ranges (the

within range of at least one light,

median range for the three without one), visibility agreed with the flags as of the dataset authors for 88.5% of segments.

Agreement beyond chance is measured by Cohen's κ

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \quad p_e = \frac{v_1 a_1 + v_0 a_0}{n^2} \quad (31)$$

where p_o is the share of the n segments on which v_s and a_s agree, v_1 and a_1 count the segments lit according to the model and to the flags, and v_0 and a_0 those dark; in 1831, $\kappa = 0.74$ (Table 11).

Most disagreements lay at the edge of the recorded ranges: the segments lit only by the model were reached at 87%–94% of the range of their best-placed light (interquartile range), against a median of 28% for the

segments lit by both, and agreement peaked when every recorded range was scaled by 0.85 (agreement 96.7%, $\kappa = 0.93$; Figure 8). For 1693, whose lights have no recorded range, agreement was highest when every light was given 5.6 km ($\kappa = 0.61$) and lowest with the common range of 18.5 km ($\kappa = 0.37$): the common range is a benchmark for siting, not a reconstruction of early visibility.

Table 11: Agreement Between the Night Visibility Computed for the Route Segments and the Flags of the Dataset Authors, With Recorded Ranges In 1831 And with Assumed Common Ranges In 1693.

Ranges	Both lit	Both dark	Model only	Flags only	Agree	κ
1831						
recorded	1,661	687	286	18	88.5%	0.74
1693						
3.7 km	122	2,340	35	155	92.8%	0.53
5.6 km	156	2,325	50	121	93.6%	0.61
9.3 km	196	2,240	135	81	91.9%	0.60
18.5 km	205	1,971	404	72	82.1%	0.37

Route segments (2,652) classified as visible at night by the model and by the flags of the dataset authors (both lit), by neither (both dark) or by only one of them, from the 94 lights of 1831 and the 26 of 1693. Lights of 1693 have no recorded range, so every light is given the stated range.

The recorded ranges were also physically plausible. The geographic range of a light, the sum of the horizon distances of a lantern at height H_j and an observer at eye height H_0 , is

$$r_{\text{geo}} = kH_j + H_0 \quad (32)$$

with $k = 3.92$ for kilometres and heights in metres

(1.17 for nautical miles and heights in feet; Bowditch, 2002). With $H_0 = 5$ m, recorded ranges correlated with geographic ranges (Spearman coefficients from 0.68 to 0.81) and 86.4%–94.2% of them exceeded the geographic range by no more than one nautical mile (Table S4).

The median ratio of recorded to geographic range rose with lantern height (Figure 9): it was 0.77 for the lowest quarter of lanterns and 0.92 for the tallest in England and Wales in 1831, and 0.60 and 0.80 in France in 1863, consistent with low lights being limited by their power rather than by the horizon.

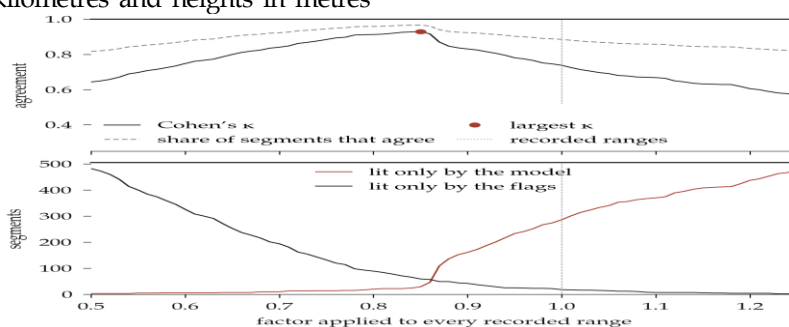


Figure 8: Agreement Between the Night Visibility Computed for the Route Segments Of 1831 and the Flags of the Dataset Authors.

the Dataset Authors When Every Recorded Range Is Multiplied by a Factor: Cohen's K And Share of Segments That Agree (First Panel) And Segments Lit Only by the Model or Only by the Flags (Second Panel).

Table 12 reports the main contrasts under alternative choices of range, band width, grid spacing, candidate set, merging distance, coastline source, demand type and inland limit. The programme waves exceeded the pre-programme French waves in 14 of the 15 specifications, but exceeded the English and Welsh waves in only 11: the

small common-range margin over England and Wales depends on analytic choices, whereas the gain over the pre-programme waves held in all specifications but one, the one that draws candidates from a coastal grid. Tables S9–S11 give every wave and benchmark under every specification.

Table 12: Sensitivity Of Static Efficiency P At Comparable Network Sizes and of the Mean Normalised Wave Efficiency H_n (Common-Range Lens) To Each Analytic Choice; The First Row Is the Main Specification.

Specification	ρ^a	ρ^b	η_n^c	η_n^d	η_n^e
Main specification	0.64	0.63	0.03	0.21	0.17
Range 9.3 km	0.61	0.58	0.06	0.16	0.22
Range 27.8 km	0.79	0.75	-0.04	0.32	0.20
Range 37.0 km	0.89	0.80	-0.12	0.38	0.23
Band width 10 km	0.63	0.64	-0.02	0.20	0.08
Band width 30 km	0.65	0.59	0.03	0.15	0.21
Grid spacing 2 km	0.64	0.63	0.03	0.21	0.17
Candidates built by t	0.96	0.98	-0.01	0.53	0.74
Coastal grid, 5 km	0.65	0.61	-0.10	-0.11	0.08
Merging 50 m	0.64	0.63	0.04	0.21	0.17
Merging 250 m	0.64	0.63	0.02	0.20	0.17
GSHHG coastline	0.64	0.62	0.00	0.18	0.18
Coastline demand	0.57	0.58	-0.21	0.08	-0.07
Inland limit 10 km	0.64	0.63	0.05	0.21	0.18
No inland limit	0.64	0.63	0.06	0.22	0.19

a England and Wales, 1831; b France, 1837; c mean over the French waves 1775–1825; d mean over the French waves 1825–1863; e mean over the English and Welsh waves 1693–1851.

Candidates built by t : only sites in use by the benchmark; Coastal grid: candidates every 5 km

along the coast instead of historical sites; Coastline demand: demand as coastline length instead of sea area. All values use the common-range lens; the first row repeats the main analysis and the other rows use 200 random additions per wave.

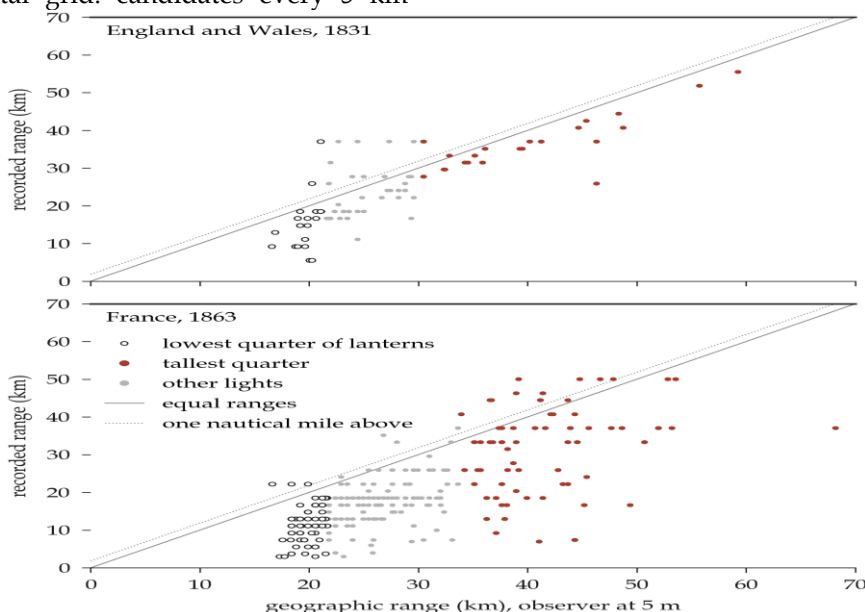


Figure 9: Recorded Range Against Geographic Range for an Observer At 5 M, In England and Wales In 1831 (First Panel) And in France in 1863 (Second Panel), With the Lowest and the Tallest Quarters of Lanterns Marked.

The exact optima were fast: every ILP of the main analysis was solved in at most 0.249 s and every observed-range ILP in at most 12.4 s. The lazy greedy evaluated between 15 and 100 times fewer marginal gains than the plain greedy, within the bounds of Equation 28, and ran 13 to 66 times faster (Table 5), with the largest gains in the largest instances.

7. DISCUSSION

The results give a quantitative content to a contrast that historians describe qualitatively. English lighthouse provision before the nineteenth century answered local petitions and patents (Coase, 1974; Taylor, 2001), and the network of 1693 covered no more sea than randomly placed lights would have. The French network was not better by default: the additions between 1775 and 1825 were as good as random, and French siting efficiency fell over that period.

What changed in the French system was the programme adopted in 1825: once each light keeps its recorded range, the waves built under it closed about half of the gap between random and optimal siting, a level the English and Welsh wave of the same period did not reach. The comparison is before and after, not a controlled contrast; the same decades brought the Fresnel lens and larger budgets, so the programme is the most direct but not the only explanation. The French advantage over England and Wales is clear with observed ranges but small and specification-dependent under a common range, which suggests that the assignment of lights of different power to different positions contributed to it, consistent with the programme's classification of lights into orders of different power (Commission des phares, 1825; Levitt, 2013).

These findings complement the organisational studies of the English system. Bogart et al. (2022) found that lights under more private control charged

higher fees and cost more to run, whereas lights with more local representation and funding served more local traffic and were maintained most cheaply; the present results show that the network those organisations produced improved in spatial efficiency over time but did not reach the observed-range wave efficiency of the centrally planned French system. The two measurements address different margins, cost per light and coverage per light, and a full welfare comparison would need both.

For heritage science the value of the method is that it turns a heritage landscape into an auditable design. The surviving towers of a coast are not a random sample of places: they are the remains of successive siting decisions, and an exact benchmark separates the lights that extended coverage from those whose value lay elsewhere, such as marking harbours or hazards. The same machinery could be applied to other networked heritage, such as signal stations, fortifications or milestones, wherever a service radius can be defined.

The computational results also matter. MCLP is NP-hard and inapproximable beyond $1-1/e$ in general, yet on these geometric instances the LP bound was within 0.003% of the integer optimum and every common-range optimum took at most 0.249 s. Worst-case hardness therefore did not bind for these two national systems, and exact benchmarks, rather than heuristics alone, are practical for historical questions of this kind.

8. LIMITATIONS

Table 13 lists the main threats to validity, their possible effect and the check that addresses each. Two of them bound the interpretation of the results: coverage measures the reach of lights, not the safety they gave, and the French comparison is before and after rather than controlled.

Table 13: Threats To Validity, Their Possible Effect on the Results and the Checks That Address Them.

Threat	Possible effect	Check
Coverage as a proxy: discs, no terrain, sectors or traffic	misstates navigational service; the same measure applies to both systems	same demand for every network; other band widths and demand as coastline length (Table 12)
Neighbouring systems ignored	border seas look uncovered	exclusion rule applied alike to both systems (Table S5)
Modern coastline	small errors in the area of the band	GSHHG coastline and other inland limits (Table 12)
Errors in historical coordinates	misplaced coverage	200 perturbation replicates (Table 10)
Ranges missing or assumed	observed-range lens limited to benchmarks with ranges	the common-range lens needs no ranges; imputation changed ρ_{obs} by at most 0.02 (Table S3)
Lights with local purposes	λ overstates redundancy	harbour, pier and jetty lights counted by name (Table S8)
Establishments merged within 100 m	lights of different character at one site are merged	merging within 50 and 250 m (Table 12)
Benchmarks are years of surviving sources	each wave spans several decades	sources listed in Table 1; results reported by wave
Before-and-after comparison of the programme	other changes of the period are confounded	stated as an association; contemporaneous English waves as comparison

Discrepancies in the French file	counts short by one light; benchmark of 1882 missing	file used as published; audit in Table S6
----------------------------------	--	---

9. CONCLUSIONS

Formulated as an MCLP, historical lighthouse siting can be benchmarked exactly and certified with greedy bounds. England and Wales moved from random siting in 1693 to a normalised efficiency of

0.36 in 1851; France moved from near-random additions before 1825 to construction waves that, with observed ranges, closed about half of the gap to optimal siting. The pipeline and its gates make these numbers reproducible and open the same audit to other networks of infrastructure heritage.

Declarations

Author Contributions: L.E.M.G.: conceptualisation, methodology, software, formal analysis, data curation, visualisation, writing (original draft, review and editing).

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Data Availability: All data are public: LAN (Buxton-Dunn and Alvarez-Palau, 2020), F-LAN (Litvine and Dunn, 2021), historic ports and shipping routes (Alvarez-Palau et al., 2019), Natural Earth (Natural Earth, 2022) and GSHHG (Wessel and Smith, 1996). The supplementary tables give the numbers behind every table of the article.

REFERENCES

- Admiralty Hydrographic Department (1851) *The Light-Houses of the British Islands*. London: Clowes and Sons.
- Admiralty Hydrographic Department (1911) *The Admiralty List of Lights*. London: Wyman and Sons.
- Alvarez-Palau, E. J. and Dunn, O. (2019) Database of historic ports and coastal sailing routes in England and Wales. Data in Brief, Vol. 25, 104188. <https://doi.org/10.1016/j.dib.2019.104188>
- Alvarez-Palau, E. J., Dunn, O., Bogart, D., Satchell, A. E. M. and Shaw-Taylor, L. (2019) Historical ports and sailing shipping routes in England and Wales 1540-1914 [data collection]. UK Data Service. <https://doi.org/10.5255/ukda-sn-853711>
- Bentley, J. L. (1975) Multidimensional binary search trees used for associative searching. *Communications of the ACM*, Vol. 18, No. 9, 509–517. <https://doi.org/10.1145/361002.361007>
- Bertrand, E. (2006) The Coasean analysis of lighthouse financing: myths and realities. *Cambridge Journal of Economics*, Vol. 30, No. 3, 389–402. <https://doi.org/10.1093/cje/bei068>
- Bogart, D., Dunn, O. B., Alvarez-Palau, E. J. and Shaw-Taylor, L. (2022) Organizations and efficiency in public services: The case of English lighthouses revisited. *Economic Inquiry*, Vol. 60, No. 2, 975–994. <https://doi.org/10.1111/ecin.13060>
- Bowditch, N. (2002) *The American Practical Navigator: An Epitome of Navigation*, Bicentennial Edition (Pub. No. 9). Bethesda, MD: National Imagery and Mapping Agency.
- Brughmans, T. (2013) Thinking Through Networks: A Review of Formal Network Methods in Archaeology. *Journal of Archaeological Method and Theory*, Vol. 20, No. 4, 623–662. <https://doi.org/10.1007/s10816-012-9133-8>
- Buxton-Dunn, O. and Alvarez-Palau, E. (2020) Historical light aids to navigation 1514-1911 [data collection]. UK Data Service. <https://doi.org/10.5255/ukda-sn-854172>
- Candela, R. A. and Geloso, V. J. (2018) The lightship in economics. *Public Choice*, Vol. 176, No. 3–4, 479–506. <https://doi.org/10.1007/s11127-018-0573-x>
- Church, R. and ReVelle, C. (1974) The maximal covering location problem. *Papers of the Regional Science Association*, Vol. 32, No. 1, 101–118. <https://doi.org/10.1007/bf01942293>
- Church, R. L. and Murray, A. (2018) *Location Covering Models*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-99846-6>
- Coase, R. H. (1974) The Lighthouse in Economics. *The Journal of Law and Economics*, Vol. 17, No. 2, 357–

376. <https://doi.org/10.1086/466796>
- Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, 37–46. <https://doi.org/10.1177/001316446002000104>
- Collins, G. (1693) *Great Britain's Coasting-Pilot*. London: Richard Mount.
- Collins, G. (1753) *Great Britain's Coasting-Pilot*. London: W. and J. Mount, and T. Page.
- Commission des phares (1825) *Rapport contenant l'exposition du système adopté par la Commission des phares, pour éclairer les côtes de France*. Paris: Direction générale des Ponts et Chaussées et des Mines.
- Conolly, J. and Lake, M. (2006) *Geographical Information Systems in Archaeology*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511807459>
- Crouse, D. F. (2016) On implementing 2D rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 52, No. 4, 1679–1696. <https://doi.org/10.1109/taes.2016.140952>
- Current, J. R. and Schilling, D. A. (1990) Analysis of Errors Due to Demand Data Aggregation in the Set Covering and Maximal Covering Location Problems. *Geographical Analysis*, Vol. 22, No. 2, 116–126. <https://doi.org/10.1111/j.1538-4632.1990.tb00199.x>
- Direction générale des Ponts et Chaussées et des Mines (1837) *Description sommaire des phares et fanaux allumés sur les côtes de France au 1er août 1837*. Paris.
- Direction générale des Ponts et Chaussées et des Mines (1846) *Description sommaire des phares et fanaux allumés sur les côtes de France au 1er juillet 1846*. Paris.
- Dunn, O. B. and Alvarez-Palau, E. J. (2020) Dataset for historical lighthouses and light aids to navigation (LAN). England and Wales, 1514–1911. *Data in Brief*, Vol. 31, 105991. <https://doi.org/10.1016/j.dib.2020.105991>
- Feige, U. (1998) A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, Vol. 45, No. 4, 634–652. <https://doi.org/10.1145/285055.285059>
- Fichou, J.-C., Le Hénaff, N. and Mével, X. (1999) *Phares: histoire du balisage et de l'éclairage des côtes de France*. Douarnenez: Le Chasse-Marée-ArMen.
- Fisher, M. L., Nemhauser, G. L. and Wolsey, L. A. (1978) An analysis of approximations for maximizing submodular set functions—II. In *Mathematical Programming Studies*, 73–87. Springer Berlin Heidelberg. <https://doi.org/10.1007/bfb0121195>
- Garey, M. R. and Johnson, D. S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: W. H. Freeman.
- Hague, D. B. and Christie, R. (1975) *Lighthouses: Their Architecture, History and Archaeology*. Llandysul: Gomer Press.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. and Oliphant, T. E. (2020) Array programming with NumPy. *Nature*, Vol. 585, No. 7825, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Huangfu, Q. and Hall, J. A. J. (2018) Parallelizing the dual revised simplex method. *Mathematical Programming Computation*, Vol. 10, No. 1, 119–142. <https://doi.org/10.1007/s12532-017-0130-5>
- Hunter, J. D. (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, Vol. 9, No. 3, 90–95. <https://doi.org/10.1109/mcse.2007.55>
- Hydrographic Office of Great Britain (1832) *The Light-Houses of the British Islands*. London: William Clowes.
- Karp, R. M. (1972) Reducibility among Combinatorial Problems. In *Complexity of Computer Computations*, 85–103. Springer US. https://doi.org/10.1007/978-1-4684-2001-2_9
- Krause, A. and Golovin, D. (2014) Submodular Function Maximization. In *Tractability*, 71–104. Cambridge University Press. <https://doi.org/10.1017/cbo9781139177801.004>
- Kuhn, H. W. (1955) The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, Vol. 2, No. 1–2, 83–97. <https://doi.org/10.1002/nav.3800020109>
- Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. and Glance, N. (2007) Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 420–429. ACM.

- <https://doi.org/10.1145/1281192.1281239>
- Levitt, T. (2013) *A Short Bright Flash: Augustin Fresnel and the Birth of the Modern Lighthouse*. New York: W. W. Norton & Company.
- Litvine, A. D. and Dunn, O. (2021) Data set for French historical light aids to navigation (F-LAN) covering the period 1775–1929. Data in Brief, Vol. 36, 106980. <https://doi.org/10.1016/j.dib.2021.106980>
- Litvine, A. D. and Dunn, O. (2021) French Historical Light Aids to Navigation (F-LAN), 1775–1929 [data collection]. UK Data Service. <https://doi.org/10.5255/ukda-sn-854607>
- Mandelbrot, B. (1967) How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*, Vol. 156, No. 3775, 636–638. <https://doi.org/10.1126/science.156.3775.636>
- Megiddo, N., Zemel, E. and Hakimi, S. L. (1983) The Maximum Coverage Location Problem. *SIAM Journal on Algebraic Discrete Methods*, Vol. 4, No. 2, 253–261. <https://doi.org/10.1137/0604028>
- Ministère des Travaux publics (1872) *État de l'éclairage et du balisage des côtes de France au 1er janvier 1872*. Paris.
- Ministère des Travaux publics (1882) *État de l'éclairage et du balisage des côtes de France au 1er janvier 1882*. Paris.
- Ministère des Travaux publics (1929) *État de l'éclairage des côtes de France au 1er janvier 1930*. Paris.
- Minoux, M. (1978) Accelerated greedy algorithms for maximizing submodular set functions. In *Lecture Notes in Control and Information Sciences*, 234–243. Springer-Verlag. <https://doi.org/10.1007/bfb0006528>
- Murray, A. T., O'Kelly, M. E. and Church, R. L. (2008) Regional service coverage modeling. *Computers & Operations Research*, Vol. 35, No. 2, 339–355. <https://doi.org/10.1016/j.cor.2006.03.004>
- Murray, A. T. (2016) Maximal Coverage Location Problem. *International Regional Science Review*, Vol. 39, No. 1, 5–27. <https://doi.org/10.1177/0160017615600222>
- Naish, J. (1985) *Seamarks: Their History and Development*. London: Stanford Maritime.
- Natural Earth (2022) Natural Earth 1:10m physical and cultural vectors, version 5.1.2 [data collection]. Natural Earth.
- Nemhauser, G. L., Wolsey, L. A. and Fisher, M. L. (1978) An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, Vol. 14, No. 1, 265–294. <https://doi.org/10.1007/bf01588971>
- ReVelle, C. and Eiselt, H. (2005) Location analysis: A synthesis and survey. *European Journal of Operational Research*, Vol. 165, No. 1, 1–19. <https://doi.org/10.1016/j.ejor.2003.11.032>
- Reynaud, L. (1864) *Mémoire sur l'éclairage et le balisage des côtes de France*. Paris: Imprimerie impériale.
- Taylor, J. (2001) Private property, public interest, and the role of the state in nineteenth-century Britain: the case of the lighthouses. *The Historical Journal*, Vol. 44, No. 3, 749–771. <https://doi.org/10.1017/s0018246x01001832>
- Thévenard, A. (1799) *Mémoires relatifs à la marine*. Paris: Laurens jeune.
- Tong, D. and Murray, A. T. (2009) Maximising coverage of spatial demand for service. *Papers in Regional Science*, Vol. 88, No. 1, 85–97. <https://doi.org/10.1111/j.1435-5957.2008.00168.x>
- van Zandt, D. E. (1993) The Lessons of the Lighthouse: “Government” or “Private” Provision of Goods. *The Journal of Legal Studies*, Vol. 22, No. 1, 47–72. <https://doi.org/10.1086/468157>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima,

- T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O. and Vázquez-Baeza, Y. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, Vol. 17, No. 3, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wessel, P. and Smith, W. H. F. (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research: Solid Earth*, Vol. 101, No. B4, 8741–8743. <https://doi.org/10.1029/96jb00104>