# Three centuries of random lines: from Buffon's needle to scale-invariant networks

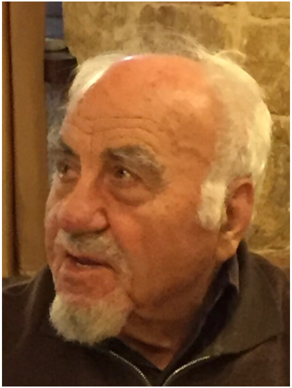
## Theo Cacoullos Memorial Lecture 2024

Wilfrid S Kendall

University of Warwick

16 May 2024

# Introduction



Theophilos Cacoullos 1932-2020

Cacoullos:
"An excellent teacher who never used notes but did use spontaneous humor"

(IMS obituary November 16, 2021)

I count it a real honour to be invited to present the second Theo Cacoullos Memorial lecture.These remarks on his presentational style make it clear that Theo Cacoullos Memorial lecturers have much to live up to! I will do my very best.

Handout available on the web: either use the `QR-code`

or visit https://wilfridskendall.github.io/talks/Kozani-Cacoullos-2024/

# Recent Origins: Particle Tracks in Apatite Crystals

An apatite crystal

- Nearly half a century ago, I assisted in a study (Laslett *et al.*, 1982) of fission track lengths in apatite crystals (tracks from fission of $^{238}$U nuclei; crystal is sliced then etched to reveal intercepting tracks);
- *In principle* the length histogram indicates thermal hence geological history; but direct observation is massively biased by
  - (a) length selection bias,
  - (b) truncation (a kind of *Wicksell Problem*),
  - (c) projection effects (if non-horizontal);
- How to model a pattern of segments? (*Poisson segment process*: a kind of *Boolean model*);
- How to get good length histogram? (select *horizontal* enclosed tracks hitting tracks that intersect slice).

# Historical Origins: pre-Revolutionary France

Georges-Louis Leclerc,
Comte de Buffon, 1707 – 1788

Buffon rose from undistinguished origins in the France of the *Ancien Regime*, to become a major and celebrated philosopher and scientist. He embodied a remarkable mixture of empiricism and rationalism.

Buffon: “Writing well consists of thinking, feeling and expressing well, of clarity of mind, soul and taste … The style is the man himself.”

# Throwing baguettes over one's shoulder (I)

Baguette as used by the Count.

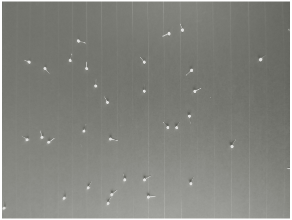


Table used by UCL experiment

- Comte de Buffon was very interested in probability (read him in Hey *et al.*, 2010);
- Probability from geometry not gambling;
- There ensued a whole sub-culture of experimental determinations of $\pi$ (see Lazzarini, 1901; Stigler, 1991),

- Most recently at the inaugural UCL Bloomsbury Probability Colloquium March 2024;
- Strips at 3cm, pins at 1.5cm, with 24 intersections out of 45 throws.
- This estimates $\pi \approx 45/21 = 2.14\ldots$: the true value of $\pi$ lies just inside the 95% confidence interval!

# Throwing baguettes over one's shoulder (II)

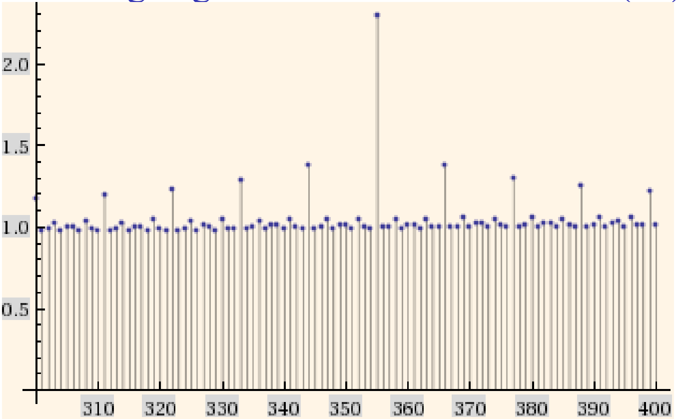
Suppose you have to design such an experiment.

- Unit-length needle, unit-ruled floor. Choose $n$ the number of trials.
- The number of successes $X$ is Binomial($n$, $2/\pi$).
- Measure success by $1/\sqrt{|2/\pi - X/n|}$.
- Use your skill and judgement (and your balance between accuracy and patience!) to decide on total number of trials $n \in [300, 400]$

In the next slide, we graph the mean measure of success against $n$ to see how much the choice of $n$ actually makes any difference!

Warwick Statistics

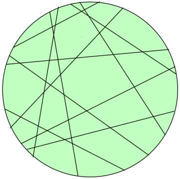
# Throwing baguettes over one's shoulder (III)




Log of mean of $1/\sqrt{|2/\pi - X/n|}$ \ for $X$ with distribution Binomial$(n, 2/\pi)$.

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{5 + \cfrac{1}{1 + \cfrac{1}{292 + \cfrac{1}{1 + \cfrac{1}{\cdots}}}}}} \quad \approx \quad 3 + \cfrac{1}{7 + \cfrac{1}{15 + 1/1}} = \frac{355}{113} = 3.14159292035398 25\,.$$

# Poisson line processes

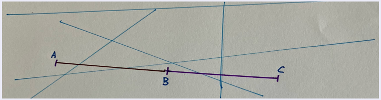
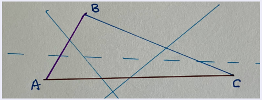
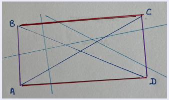


How to build random patterns of lines?

1. There is a theory of Random Closed Sets. More constructively,
2. Boolean model: union of sets located are Poisson points.
3. Random lines $\ell$ are not localized, so this must be adapted.
4. Simple solution: parametrize by signed perpendicular distance $r$, angle $\theta$:
   - (i) represent each line $\ell$ by a point $(r, \theta)$ in representation space;
   - (ii) view Poisson line process as Poisson point process in representation space. Use invariant line measure $\frac{1}{2}\,\mathrm{d}r\,\mathrm{d}\theta$.
5. The representation space is a cylinder with twist.
6. Calculations typically reduce to computation of probabilities that there are no lines of particular forms.

## Examples of calculations with lines

The probability of there being no lines in (the representation space subset) $E$ is $\exp(-\mu(E))$, where $\mathrm{d}\mu = \frac{1}{2}\,\mathrm{d}\,r\ \mathrm{d}\,\theta$ is the invariant measure. Compute:

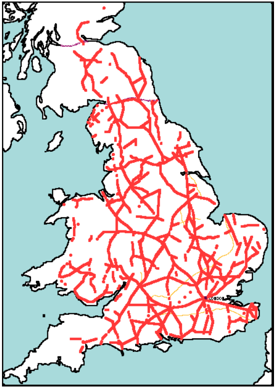
- $\mathbb{P}$ [no lines hit a unit segment]:
  - 
    
  - By (a) translation invariance, (b) choice of normalizing constant $\frac{1}{2}$, the probability can be computed to be $\exp(-1)$.
- $\mathbb{P}$ [all lines hitting a triangle $ABC$ actually hit side $AC$]:
  - 
    
  - By a counting argument(!) this is $\exp\left(-\frac{1}{2}(|AB| + |BC| - |AC|)\right)$.
- $\mathbb{P}$ [no lines pass between $AD$ and $BC$ of rectangle $ABCD$]:
  - 
    
  - Counting argument again(!): $\exp\left(-\frac{1}{2}(|AC| + |BD| - |AD| - |BC|)\right)$.

## Other relevant notions

- An alternate parametrization can be very convenient for calculations: $p$ is distance of intercept along a reference line, while $\phi$ is angle as before.
  - invariant measure now $\frac{1}{2}\sin\phi \; \mathrm{d}p \, \mathrm{d}\phi$;
- Notion of "conditioning on a line at the origin":
  - Slivnyak's theorem says that the remainder of the line pattern behaves just like the original Poisson line process! This is the theory of *Palm probability*, extremely useful in showing how to reason with general constructions;
- The use of marks for the points / lines:
  - *eg* speed-marked line process.

Warwick Statistics

# Application to network efficiency



Choose: spend money on expensive roads, or be unable to shoot trouble fast.

Can the Poisson line process help the Roman Emperor?

# Network Efficiency: mathematical model

Long-range transport can be almost maximally efficient (compared to a complete planar graph), simply by laying down straight roads according to a sparse Poisson line process.

- Network connecting $N$ cities in rectangle side $\sqrt{N}$.
  - (A) Measure efficiency by minimizing connecting stuff? (Steiner tree)
  - (B) Measure efficiency by average excess of connection distance over Euclidean? (Complete planar graph)
- Aldous & WSK (2008): start with Steiner tree:
  - Add sparse set of random lines from a Poisson line process;
  - Add sparse rectilinear grid connecting lines and tree;
  - Add some box structures to avoid hotspots.
  - Resulting network (large $N$) is economical with connection stuff (not much more than Steiner tree), but average excess over complete planar graph is only logarithmic in $N$.
- Debunks a “natural” statistic for network efficiency (more on this in Aldous & Shun, 2010).

Warwick Statistics

# Network Efficiency: some details

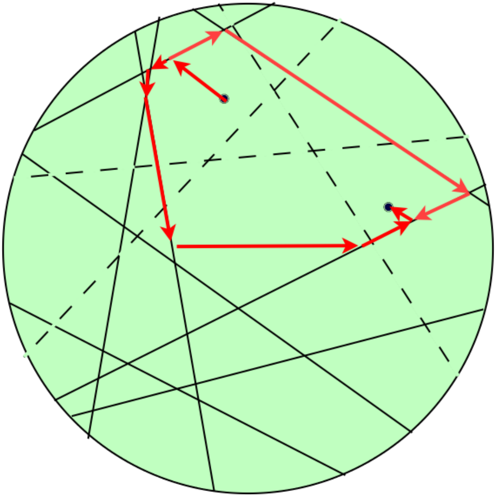
- Logarithmic upper bound for "mean connection distance minus Euclidean distance" using Poisson line process.
    - (Steinhaus estimator for distance, study intersections of original Poisson line process and independent copy, focus on a certain Poisson polygon related to "near-geodesics": lazy approximations to geodesics that avoid "over-shooting" the target);
- The logarithmic upper bound is of the correct order for a network based on a Poisson line process. This justifies use of *near geodesics* ("approximate geodesics").
    - (Stereological estimation using 60-year-old refinements of Mills ratio inequality);
- Controlling fluctuations by bounding variance of excess connection distance.
    - (represent perimeter of a Poisson polygon using theory of Lévy processes and self-similar processes).

# Traffic in networks

# The Poissonian City

## Design an idealized city!

Complete the 2 possible routes using a final off-road section.

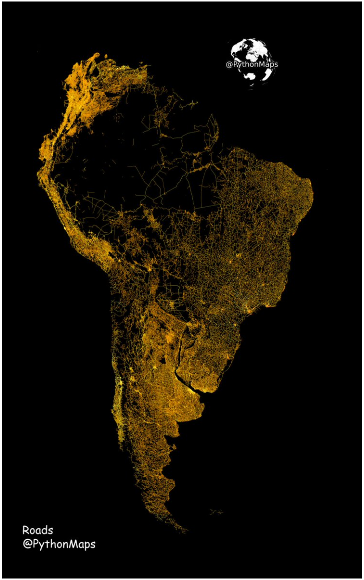


# The Poissonian City (results)

Recall: traffic generated uniformly between all pairs in a disk of radius $n$, routes formed by near-geodesics from standard Poisson line process, divided equally between two possible routes.

- As disk radius $n$ tends to $\infty$, traffic through the centre behaves as $2n^3$;
- Re-scaling, there is a limiting distribution for traffic through the centre;
- Limiting behaviour of traffic can be expressed as volume of a certain 4-dimensional Poisson polytope generated by an improper Poisson line process.
- Gameros Leal (2017) worked this out for a modified model and compared it to empirical results concerning British rail traffic before and after the 1960's Beeching railway cuts.

# Stochastic geometry of online maps



# What do we require of a model for an online map?

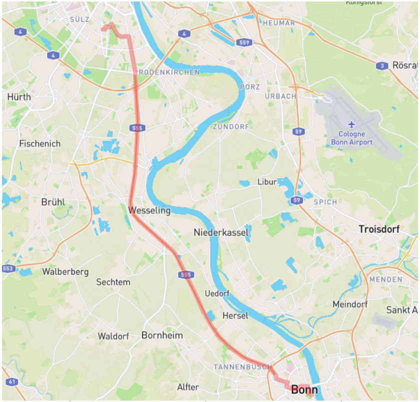
Axioms (following Aldous & Ganesan, 2013; Aldous, 2014).

The online map software generates routes between pairs of nodes, hence networks between any given set of nodes.

Given: (random) network $\mathcal{N}(\underline{x}_1, \ldots, \underline{x}_n)$ connecting nodes. Require:

1. Scale-invariance: for each Euclidean similarity $\mathfrak{S}$,
$\mathcal{L}\left(N(\mathfrak{S}\underline{x}_1, \ldots, \mathfrak{S}\underline{x}_n)\right) = \mathcal{L}\left(\mathfrak{S}N(\underline{x}_1, \ldots, \underline{x}_n)\right)$.
2. If $D_1$ is length of fastest route between two points at unit distance apart then mean length is finite: $\mathbb{E}\left[D_1\right] < \infty$.
3. Weak SIRSN property: network connecting points of (independent) unit intensity Poisson point process has finite average length density;
**or**
(Strong) SIRSN property: finite length intensity of "long-range" routes for points of *dense* Poisson point process.

# A visual introduction to axioms for online maps

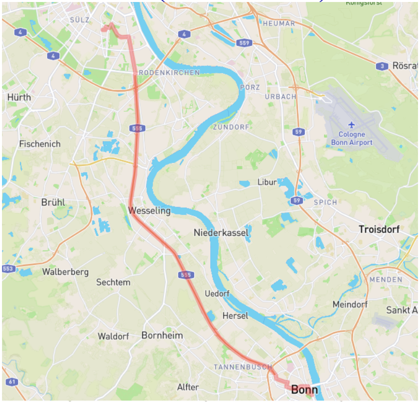


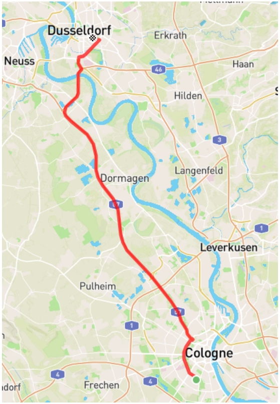

Given locations $\underline{x}_1, \ldots, \underline{x}_n$, the SIRSN should generate a *network*

$$\mathcal{N}(\underline{x}_1, \ldots, \underline{x}_n) \quad = \quad \bigcup \left\{ \mathcal{R}(\underline{x}_i, \underline{x}_j) \; ; \; 1 \leq i < j \leq n \right\}$$

of chosen *routes* $\mathcal{R}(\underline{x}_i, \underline{x}_j)$ connecting $\underline{x}_i$ to $\underline{x}_j$.

# Axiom 1: (Statistical) scale-invariance of networks



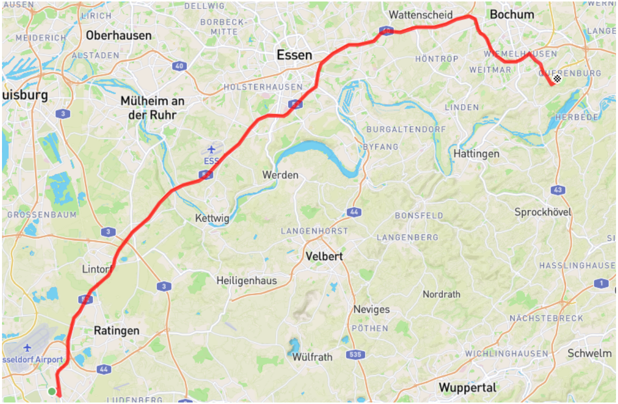




Axiom 1 requires that SIRSN networks must be *scale invariant* under similarities $\mathfrak{S}$ on $\mathbb{R}^m$:

$$\mathcal{L}\left(\mathfrak{S}\mathcal{N}(\underline{x}_1, \ldots, \underline{x}_n)\right) \quad = \quad \mathcal{L}\left(\mathcal{N}(\mathfrak{S}\underline{x}_1, \ldots, \mathfrak{S}\underline{x}_n)\right) .$$

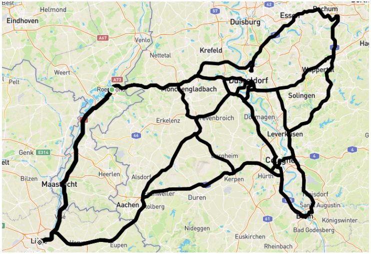
(Model route $\mathcal{R}(\underline{x}_i, \underline{x}_j)$ hence network $\mathcal{N}(\underline{x}_1, \ldots, \underline{x}_n)$ as random.)

# Axiom 2: Finite mean length of routes




Axiom 2 insists that SIRSN routes must have finite mean length.

# Axiom 3: SIRSN property

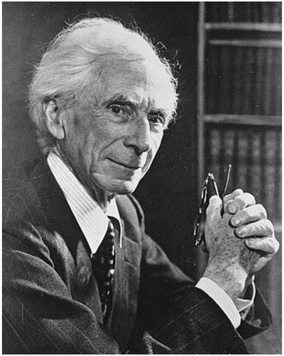



Axiom 3 requires that the (strong) SIRSN property holds (considerable re-use of routes).

(Weak) Source/terminus nodes spread evenly everywhere. “Route-re-use” means there is a relatively small mean-network-length-per-unit-area.

(Strong) Source/terminus nodes spread densely everywhere. “Route-re-use” means there is still only a relatively small mean-network-length-per-unit-area for the “long-distance” network.

# But do SIRSN exist at all?



Bertrand Russell (1872-1970):

"The axiomatic method has many advantages over honest work."
But it is useless if in fact there are no possible examples!

First example: Hierarchical SIRSN (Aldous, 2014),
a *randomized* dense dyadic speed-marked planar *lattice*.

# Do SIRSN exist at all? a more natural answer

Second example: Poisson line process (Plp) SIRSN: planar *and also* in higher dimensions too (Aldous, 2014; Kahn, 2016; WSK, 2017).

Construction uses speed-marked (improper) Plp $\Pi$.

$\Pi$ is governed by intensity measure $\varpi_{\mathrm{Plp}}$ on $(0,\infty)\times\mathbb{R}\times[0,\pi)$:

$$\mathrm{d}\varpi_{\mathrm{Plp}} \quad = \quad \tfrac{1}{2}(\gamma-1)v^{-\gamma}\,\mathrm{d}v\;\;\mathrm{d}r\,\mathrm{d}\theta \qquad (\text{require } \gamma > 2)\,.$$

NB: $\frac{1}{2}\,\mathrm{d}r\,\mathrm{d}\theta$: intensity measure for standard Poisson line process.

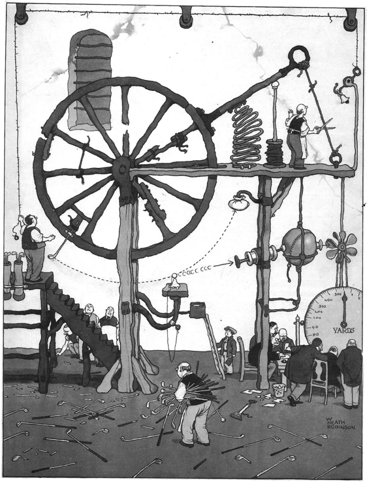
Use fastest $\Pi$-paths for routes. Require $\gamma > 2$ so

- $\Pi$-paths cannot reach $\infty$ in finite time;
- $\Pi$-paths can reach any prescribed point in finite time.

NB ($\gamma > 1$): Lines of speed $> 1$ form standard Poisson line process $\Pi^{(\geq 1)}$.

Proof of SIRSN axioms (especially third axiom) is non-trivial!

# Some further remarks




Testing Golf Drivers

34

- Lines are unsatisfactory models for roads: they are too long(!) and too straight.
- The proof that Plp-SIRSN satisfies the axioms has now been generalized so that it only depends on getting good bounds on the time taken to transit between the two “flat ends” of a cylinder.
- We’d like to move from lines to line segments (and then to curved fibres).
- Clearly we’d need to know when random patterns of line segments percolate (connect up enough to provide long-distance routes). This is a dramatically hard problem!
- However we can finesse by considering a ”thickness” property: if the length of a typical unit-speed line segment has finite first moment but infinite second moment then the two “flat ends” of any cylinder will be connected by (perhaps very slow) single fibres.
- Under a slight quantification of the divergence of the second moment, the SIRSN property can then be proved.

# Conclusion

Many questions remain!

1. "Thickness" is a simplifying assumption and still has reality problems:
   - if sticks have finite-mean-length but infinite second-length-moment, then *any* selected stick will have infinite-mean-length (length-biasing!);
   - Conjecture: it suffices to have super-critical percolation for the configuration of fibres of greater than unit speed;
2. Everything still works for lines in higher-dimensional space:
   - Dimension 3 is another case of interest (*eg* human connectome!);
   - Blanc has recently obtained results on an analogue in hyperbolic space.
3. One can now relax parts of the scaling symmetry assumption while maintaining the "SIRSN effect" (Axiom 3), *eg* under controlled spatial inhomogeneity. Thus we have a flexible framework for statistical and operational research modelling.
4. Introduce *traffic*, *eg* supposing traffic between any pair of points to be uniformly generated along the connecting route. One can now develop Gameros Leal (2017) comparisons.

Thank you for your attention! QUESTIONS?

# References I


Aldous, D.J. (2014) Scale-Invariant Random Spatial Networks. *Electronic Journal of Probability*, **19**, 1–41.

Aldous, D.J. & Ganesan, K. (2013) True scale-invariant random spatial networks. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 8782–8785.

Aldous, D.J. & Shun, J. (2010) Connected spatial networks over random points and a route-length statistic. *Statistical Science*, **25**, 1–28.

Aldous, D.J. & WSK (2008) Short-length routes in low-cost networks via Poisson line patterns. *Advances in Applied Probability*, **40**, 1–21.

Gameros Leal, R.M. (2017) Mean traffic behaviour in Poissonian cities (PhD Thesis).

Hey, J.D., Neugebauer, T.M., & Pasca, C.M. (2010) Georges-Louis Leclerc de Buffon's 'Essays on Moral Arithmetic'. *The selten school of behavioral economics* (Sadrieh, A., Ockenfels, A.(eds). ed). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 245–282.

Kahn, J. (2016) Improper Poisson line process as SIRSN in any dimension. *Annals of Probability*, **44**, 2694–2725.

Laslett, G.M., WSK, Gleadow, A.J.W., & Duddy, I.R. (1982) Bias in measurement of fission-track length distributions. *Nuclear Tracks and Radiation Measurements (1982)*, **6**, 79–85.

Lazzarini, M. (1901) Un'applicazione del calcolo della probabilita alla ricerca esperimentale di un valore approssimato di $\pi$. *Periodico di Matematica*, **4**, 49–68.

# References II


Schneider, R. & Weil, W. (2008) *Stochastic and Integral Geometry*, Probability and its applications. Berlin, Heidelberg: Springer Berlin Heidelberg.

Steinhaus, H. (1930) Zur Praxis der Rektifikation und zum Längenbegriff. *Sächsischen Akad. Wiss. Leipzig*, **82**, 120–130.

Stigler, S.M. (1991) Stochastic simulation in the nineteenth century. *Statistical Science*, **6**, 89–97.

WSK (2011) Geodesics and flows in a Poissonian city. *Annals of Applied Probability*, **21**, 801–842.

WSK (2014) Return to the Poissonian city. *Journal of Applied Probability*, **51**, 297–309.

WSK (2017) From random lines to metric spaces. *Annals of Probability*, **45**, 469–517.


Warwick Statistics

# Technical information

| Image | Original location | |
|---|---|---|
| Theo Cacoullos | IMS Bulletin obituary, November 2021 | |
| Apatite crystal | Apatite_taillée.jpg | By Didier Descouens - Own work, CC BY 3.0 |
| Comte de Buffon | Buffon_1707-1788.jpg | Free of copyright |
| Baguette | baguette-de-pain-.jpg | Free stock photo - Public Domain |
| Buffon and Number Theory | WSK calculation | |
| Buffon Table | By kind permission of Terry Soo (UCL) | |
| Poisson line process | WSK calculation | |
| S. American Roads | By kind permission of @PythonMaps | |
| Heath Robinson | Testing Golf Drivers | |

These notes were produced from `Kozani-Cacoullos-2024.qmd`:

| | |
|---|---|
| Author: | Wilfrid Kendall W.S.Kendall@warwick.ac.uk |
| Date: | Sat May 11 16:53:38 2024 +0100 |
| Summary: | First draft now completed. |