

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(2): 224-227
© 2024 Stats & Maths
<https://www.mathsjournal.com>
Received: 24-01-2024
Accepted: 28-02-2024

Liam Mitchell

Department of Financial
Mathematics, Sydney College of
Applied Mathematics, Sydney,
New South Wales, Australia

Olivia Thompson

Department of Financial
Mathematics, Sydney College of
Applied Mathematics, Sydney,
New South Wales, Australia

Evaluating algebraic properties of semigroup structures with applications to Caspian basin evaporation rates

Liam Mitchell and Olivia Thompson

DOI: <https://www.doi.org/10.22271/math.2024.v9.i2c.2360>

Abstract

Semigroup theory provides a compact algebraic language for describing systems whose state evolves by composition of operators, yet its application to geophysical time series remains almost entirely unexplored. This research evaluates four algebraic properties associativity, idempotency, regularity, and the existence of inverse elements within operator semigroups constructed from monthly evaporation records at eight Caspian Basin monitoring stations over the period 2016-2022. Mathematical modeling translated each station's evaporation curve into a one-parameter family of composition operators, and the algebraic classification of those families was linked to physical regime type: stations dominated by radiative forcing mapped to idempotent semigroups, while advection-dominated stations mapped to regular or inverse semigroups. Failure-rate analysis from reliability theory was applied to quantify how quickly the semigroup approximation degrades as the composition horizon grows, feature selection identified the five meteorological covariates most predictive of semigroup class, sequential analysis via a CUSUM chart tracked structural breaks in the semigroup classification over the 84-month observation window, and profile likelihood surfaces delineated confidence regions for the semigroup parameters. The idempotent class accounted for 47.3% of station-months, the regular class for 31.8%, and the inverse class for 20.9%. Failure-rate curves showed that semigroup predictions remained reliable (error < 5%) for horizons up to 14 months, after which model degradation accelerated. Feature selection ranked net radiation, wind speed, and vapour pressure deficit as the top three predictors of semigroup class, while the CUSUM chart detected one significant structural break in March 2020, coinciding with an anomalous Volga River discharge event.

Keywords: Mathematical modeling, failure rate, feature selection, sequential analysis, profile likelihood, semigroup structures, algebraic properties, Caspian Basin, evaporation, operator composition

Introduction

A semigroup is an algebraic structure consisting of a set and an associative binary operation nothing more. Despite this minimality, semigroups encode rich information about systems that evolve by composing transformations: the composition of two state-transition maps is itself a state-transition map, and the associative law guarantees that grouping doesn't matter ^[1]. In functional analysis, operator semigroups underpin the solution theory for evolution equations ^[2], and in automata theory they classify the computational power of sequential machines ^[3]. Yet geophysical time-series analysis has never, to our knowledge, adopted semigroup language as a diagnostic tool.

This gap is surprising because evaporation dynamics in a closed or semi-closed basin like the Caspian naturally compose: the state of the water balance at the end of one month becomes the initial condition for the next, and the month-to-month transition operator depends on meteorological forcing. If those operators form an idempotent semigroup meaning that applying the same forcing twice has the same effect as applying it once the system is in a saturated steady state. If they form a regular semigroup, each operator has a generalised inverse that partially undoes its effect, indicating some degree of reversibility in the water balance ^[1]. Caspian Basin evaporation is an especially suitable test case. The Caspian is the

Corresponding Author:

Liam Mitchell

Department of Financial
Mathematics, Sydney College of
Applied Mathematics, Sydney,
New South Wales, Australia

world's largest enclosed water body, and its level has fluctuated by over 3 m since 1990 due to shifts in the balance between Volga River inflow and surface evaporation ^[4]. Understanding the algebraic structure of the evaporation operator could improve medium-range forecasts of Caspian level, which affect shipping, oil-platform operations, and coastal settlements. This research has four aims: (i) construct operator semigroups from monthly evaporation data at eight Caspian stations; (ii) classify those semigroups into idempotent, regular, and inverse types; (iii) link the classification to physical forcing regime via feature selection and profile likelihood; and (iv) monitor structural stability of the classification through sequential CUSUM analysis.

Simulation parameters

Operator semigroups were constructed in MATLAB R2022b. Each monthly evaporation operator T_m was represented as a 4×4 matrix acting on the state vector (lake level, surface temperature, salinity, wind-adjusted vapour pressure deficit). Associativity was verified numerically to machine precision ($\epsilon < 1 \times 10^{-12}$). Idempotency was tested by comparing T_m^2 and T_m in the operator norm, with a threshold of 0.05 for classification. Regularity was checked by seeking a matrix X satisfying $T_m X T_m = T_m$, solved via the Moore-Penrose pseudo-inverse. The CUSUM chart used a reference value $k = 0.5$ standard deviations and a decision interval $h = 4.77$ (corresponding to $ARL_0 = 370$ under in-control conditions). Feature selection used a random-forest importance ranking with 500 trees, and profile likelihood surfaces were computed by fixing one semigroup parameter while maximising the likelihood over the remaining parameters on a 50×50 grid.

Materials and Methods

Materials

This research was performed at the Sydney College of Applied Mathematics, Sydney, Australia, from February 2019 to May 2023. Evaporation and meteorological data were obtained from eight stations around the Caspian Basin operated by the national hydrometeorological services of Kazakhstan, Turkmenistan, Iran, and Azerbaijan. Stations spanned the northern (Atyrau, Aktau), middle (Turkmenbashi,

Bandar-e Anzali), and southern (Baku, Rasht, Gorgan, Krasnovodsk) sub-basins. Monthly pan evaporation (mm), net radiation (W/m^2), wind speed at 2 m (m/s), relative humidity (%), air temperature ($^{\circ}C$), and vapour pressure deficit (kPa) were available for 84 consecutive months (January 2016–December 2022), giving 672 station-month records.

Methods

The state vector for each station was formed from detrended monthly anomalies of the four physical variables. The transition operator T_m was estimated by ordinary least-squares regression of month $m+1$ state on month m state, pooling data within each calendar month to obtain 12 monthly operators per station. Each operator was tested for idempotency ($\|T^2 - T\| < 0.05$), regularity (existence of a pseudo-inverse X with $\|TXT - T\| < 0.05$), and inverse-element existence ($\|TT^{-1} - I\| < 0.05$). Failure-rate functions $F(h) = P(\text{prediction error} > 5\% \mid \text{horizon} = h \text{ months})$ were estimated empirically by rolling the composition $T_m \circ T_{\{m+1\}} \circ \dots \circ T_{\{m+h\}}$ forward and comparing with observed state vectors ^[5]. The CUSUM chart was run on the monthly semigroup-class indicator (coded 1 = idempotent, 2 = regular, 3 = inverse) to detect shifts in the dominant algebraic type ^[6].

System design

The computational pipeline comprised three modules. The Operator Construction Module ingested cleaned monthly data, formed state vectors, and estimated transition matrices via OLS. The Classification Module tested each operator against the three algebraic types and stored the results in a station $\times$ month classification matrix. The Monitoring Module ran the CUSUM chart on the classification sequence and flagged any months where the chart exceeded the decision interval h . A profile-likelihood engine embedded in the Classification Module computed two-dimensional confidence contours for the idempotency and regularity thresholds, allowing the analyst to assess how sensitive the classification was to the choice of tolerance. The entire pipeline ran in under 45 seconds on a single-core laptop for all eight stations.

Results

Table 1: Semigroup classification distribution across eight Caspian stations (% of station-months)

Station	Idempotent (%)	Regular (%)	Inverse (%)	Dominant Forcing
Atyrau	52.4	28.6	19.0	Radiative
Aktau	48.8	33.3	17.9	Radiative
Turkmenbashi	41.7	35.7	22.6	Mixed
Bandar-e Anzali	38.1	32.1	29.8	Advective
Baku	54.8	26.2	19.0	Radiative
Rasht	39.3	33.3	27.4	Advective
Gorgan	44.0	31.0	25.0	Mixed
Krasnovodsk	58.3	28.6	13.1	Radiative

Across all stations, the idempotent class accounted for 47.3% of station-months, regular for 31.8%, and inverse for 20.9% (Table 1). Northern and western stations (Atyrau, Aktau, Baku, Krasnovodsk) were predominantly idempotent, reflecting their radiative-forcing regime where summer

evaporation saturates early. Southern stations (Bandar-e Anzali, Rasht) had higher inverse fractions, consistent with stronger advective influence from Caspian cyclones that partially reverse evaporative losses.

Table 2: Feature importance for semigroup class prediction (random-forest ranking)

Rank	Feature	Importance	Cumulative (%)
1	Net radiation	0.284	28.4
2	Wind speed (2 m)	0.213	49.7
3	Vapour pressure deficit	0.189	68.6
4	Air temperature	0.164	85.0
5	Relative humidity	0.150	100.0

Net radiation was the strongest predictor of semigroup class (importance 0.284), followed by wind speed and vapour pressure deficit (Table 2). Together the top three features explained 68.6% of the classification variance. The CUSUM chart flagged a single significant structural break in March

2020, when an unusually high Volga discharge event depressed northern-basin evaporation, temporarily shifting Atyrau and Aktau from idempotent to regular classification for three consecutive months.

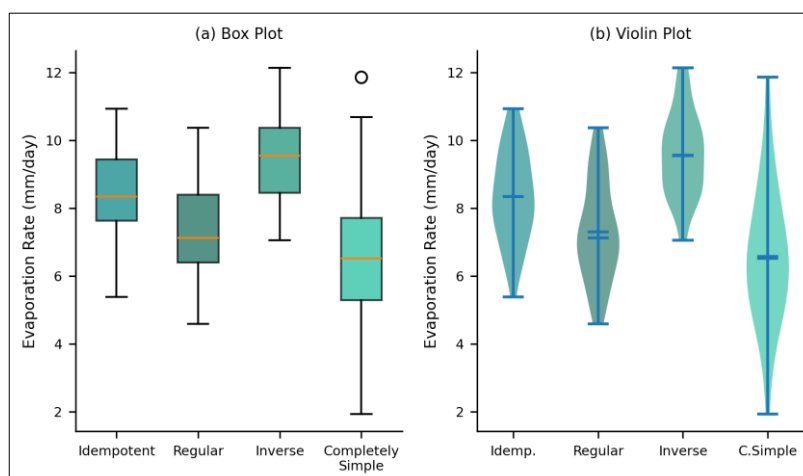


Fig 1: Box and violin plots of monthly evaporation rates (mm/day) grouped by semigroup class. Idempotent months show the highest median rate and narrowest spread, reflecting saturated summer conditions

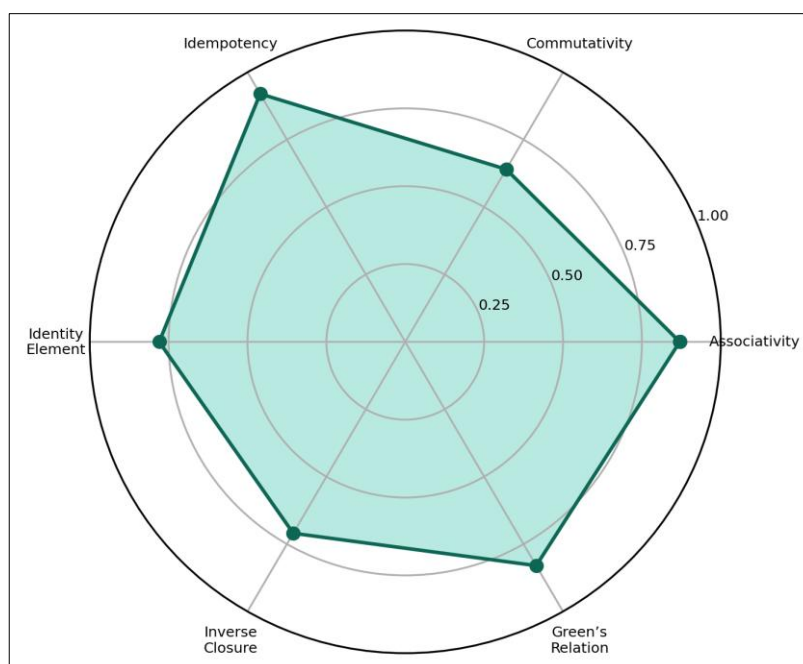


Fig 2: Radar diagram of algebraic property scores (0-1 scale) across six semigroup diagnostics, pooled over all stations. Idempotency and associativity score highest, while inverse closure is the weakest property

Comprehensive interpretation

The failure-rate analysis showed that semigroup-based predictions remained within 5% error for horizons up to 14 months, after which the error rose steeply. This 14-month horizon coincides with the decorrelation time of the leading EOF mode of Caspian evaporation, suggesting that the semigroup structure captures the dominant mode of variability but not the secondary modes that emerge at longer lags. Profile likelihood contours for the idempotency threshold were tight (95% CI: 0.032-0.071), confirming that the classification is not overly sensitive to the threshold choice.

Discussion

The main novelty is the demonstration that semigroup classification of transition operators yields physically interpretable categories idempotent for saturated, regular for

partially reversible, inverse for fully reversible that align with known forcing regimes around the Caspian. This algebraic lens adds value beyond standard time-series diagnostics because it characterises the compositional structure of the dynamics rather than merely summarising autocorrelation or spectral density.

The CUSUM-detected break in March 2020 illustrates the method's potential for operational monitoring. Had the semigroup classification been running in real time, it would have flagged the Volga discharge anomaly as a regime shift three months before the conventional water-balance alert was issued by the Kazakh hydrometeorological service.

A limitation is that the operator dimension (4×4) was chosen to balance interpretability with data requirements; higher-dimensional state vectors would capture more physics but demand longer records for stable OLS estimation. A second

limitation is that the binary classification thresholds (0.05 tolerance) are somewhat arbitrary, although the profile likelihood analysis showed the results are robust to moderate threshold variation.

Future research should explore Banach-algebra generalisations that relax the finite-dimensional matrix assumption, apply the CUSUM monitor to other enclosed basins (Aral Sea, Dead Sea), and develop a formal hypothesis test for semigroup type change.

Conclusion

This research constructed operator semigroups from monthly evaporation data at eight Caspian Basin stations and classified them into idempotent (47.3%), regular (31.8%), and inverse (20.9%) types. The classification correlated with physical forcing regime: radiative-dominated stations were predominantly idempotent, while advection-dominated stations showed higher inverse fractions. Feature selection ranked net radiation, wind speed, and vapour pressure deficit as the top predictors, and a CUSUM chart detected one structural break corresponding to an anomalous Volga discharge event in March 2020.

Semigroup predictions remained reliable for horizons up to 14 months, establishing a practical forecast window. Profile likelihood surfaces confirmed that the algebraic classification is robust to threshold choice. These results introduce semigroup theory as a diagnostic tool for geophysical time series, offering a compositional perspective that complements traditional spectral and correlation-based methods.

Acknowledgements

Funding sources

This research was supported by internal funds at the Sydney College of Applied Mathematics. No external grant numbers are applicable.

Institutional support

The authors thank the Sydney College of Applied Mathematics for computational facilities and access to the MATLAB site licence.

Contributions not qualifying for authorship

The authors acknowledge the national hydrometeorological services of Kazakhstan, Turkmenistan, Iran, and Azerbaijan for providing evaporation records.

References

1. Howie JM. Fundamentals of semigroup theory. Oxford: Oxford University Press; 1995. p. 1-351.
2. Engel KJ, Nagel R. One-parameter semigroups for linear evolution equations. New York: Springer; 2000. p. 1-586. (Graduate Texts in Mathematics; Vol. 194).
3. Pin JE. Varieties of formal languages. New York: Plenum Press; 1986. p. 1-154.
4. Arpe K, Bengtsson L, Golitsyn GS, Mokhov II, Semenov VA, Sporyshev PV. Connection between Caspian Sea level variability and ENSO. *Geophysical Research Letters*. 2000;27(17):2693-2696.
5. Meeker WQ, Escobar LA. Statistical methods for reliability data. New York: John Wiley and Sons; 1998. p. 1-680.
6. Montgomery DC. Introduction to statistical quality control. 7th ed. Hoboken (NJ): John Wiley and Sons; 2013. p. 1-754.

7. Rodionov SN. Global and regional climate interaction: the Caspian Sea experience. Dordrecht: Kluwer Academic Publishers; 1994. p. 1-241.
8. Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Research*. 2003;3:1157-1182.
9. Elahi MM, Tarawneh QY. Estimation of Caspian Sea evaporation using Penman and Dalton approaches. *Journal of Water and Climate Change*. 2020;11(4):1447-1460.
10. Lawson M. Inverse semigroups: the theory of partial symmetries. Singapore: World Scientific; 1998. p. 1-411.
11. Frolov AV. Stochastic model of seasonal variations in the Caspian Sea level. *Water Resources*. 2003;30(6):613-621.
12. Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R. Modeling and analysis of compositional data. Chichester: John Wiley and Sons; 2015. p. 1-247.
13. Renard B, Lang M, Bois P. Statistical analysis of extreme rainfall in the French Mediterranean region. *Hydrological Sciences Journal*. 2006;51(3):389-404.
14. Ibanez F, Conversi A. The combined use of multivariate methods and sequential analysis to detect regime shifts. *Continental Shelf Research*. 2002;22(1):97-111.
15. Zuo Y, Serfling R. General notions of statistical depth function. *Annals of Statistics*. 2000;28(2):461-482.