

The Significance of Three Methods of Grouping Biblical Hebrew Text Portions

A. Dean Forbes

<http://orcid.org/0000-0002-4633-9048>

University of the Free State, South Africa

adforbes@post.harvard.edu

Abstract

The paper first indicates the implications of the mixed results obtained by using three disparate analytical methods to infer relationships among biblical text portions based upon their spelling practices. Next, a sketch is provided of *matres lectionis* (“mothers of reading”) in Biblical Hebrew and of the Andersen-Forbes classification system. Vowel features are specified, and examples presented. The notion of transmissional textual change is introduced. The criticality of comparing the results provided by different analytical methods is emphasised. Next, three complementary analytical methods are introduced in turn, and their results are appraised. *Clustering* is a heuristic data exploration method, its prime result being that the spelling of the Torah sets it well apart from the other portions of the Hebrew Bible. Clustering, however, produces many other provocative portion groupings inviting investigation. While *multidimensional scaling* also gathers the Torah portions, it also yields its own tantalising juxtapositions. *Seriation* orders the portions along a timeline. It results in an expected horseshoe-shaped band of portions, albeit rather “puffy.” Also, some of its text-portion orders are suspicious. While many results produced by the three methods are encouraging, many are perplexing. Envisioned future application of evolving methods to our BH text-portion data may well enhance the trustworthiness of our inferences.

Keywords: *mater lectionis*; spelling; text-portion grouping; text-portion ordering



Journal for Semitics

<https://unisapressjournals.co.za/index.php/JSEM>

Volume 30 | Number 2 | 2021 | #9320 | 13 pages

<https://doi.org/10.25159/2663-6573/9320>

ISSN 2663-6573 (Online), ISSN 1013-8471 (Print)

© The Author(s) 2022



Published by the Southern African Society for Near Eastern Studies and Unisa Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International License (<https://creativecommons.org/licenses/by-sa/4.0/>)

Implications of Previous Work on BH Spelling

As will be documented below, three disparate analytical approaches yield intriguing results concerning ancient Hebrew orthography. However, some results are very puzzling. The unevenness of results strongly suggests that additional basic research must be carried out. Even when possibly confounding variables are handled and noise-tolerant methods are used, the recovered ordering will at best only reflect the copyist-supported text life. Forbes (2019) observed:

Transmission variants are generated during the “copyist-supported life” of a text, their generation rate likely varying with copyist, time, place, and genre. As a text moves from free to careful copying, innovation may stop, leading to *clamped* innovation. Unlike the classic wax-and-wane innovations in archaeology, clamping should make the BH innovation curves [S-curves, rising to and maintaining a maximum value]. [However, t]oo extensive changes can bleach a text of all reliable signs of earlier language performance. It is impossible, *ab initio*, to know how massive bleaching has been.

So, it may turn out that changes inflicted during transmission have bleached out the remnants of spelling information on which the methods feed. Current indications are that this is not the case, but we will not be sure until the full-fledged research has been completed.

A Brief Review of the Basics¹

Nomenclature and Emergence

According to Cross and Freedman (1952, 2), from about the ninth century BCE, a trio of *matres lectionis* “mothers of reading” (ו, י, and ה)² were inserted in Hebrew texts to assist reading. Consider for example Qoh 1:6, where we find סוֹבֵב סָבֵב, “round and round” (RSV). The first word has the *mater lectionis* ו while the second lacks it. The first is said to be written *plene*, while the second is written *defectivi*. The emergence of *matres lectionis* was characterised by Orlinsky (quoted in Cross and Freedman 1952, 60, footnote 3) thus: “From all the information that we have about North West Semitic orthography, the *scriptio plena* was at first utterly unknown; writing was consonantal. Gradually vowel letters began to come into use, at first in final position only, later also medially.”

Vowel Feature Specification

After evaluating the evidence advanced by Bange (1971) and Zevit (1980), Andersen relied primarily upon Cross and Freedman (1952) in devising his taxonomy of vowel

1 For an introduction to *matres lectionis* in Biblical Hebrew, see Andersen and Forbes (2013). On diachrony, see Forbes and Andersen (2012).

2 A fourth *mater lectionis* (ח) was introduced much later (Cross and Freedman 1952, 2).

types (Andersen and Forbes 1986, 162–204). For the *ketiv* text of the Leningrad Codex (L), Andersen assigned a vowel type and stress level to each of the 108,943 candidate vowels identified. An algorithm then assigned the spelling choice (*plene* or *defectivi*) and text portion to each candidate vowel, the latter being drawn from a catalogue of 76 possibilities.³

Vowel Types (65 in all)

To convey some idea of the nature of the vowel types, Table 1 characterises two of the 26 vowel types involving /ō/ vowels.⁴

Table 1: Two sample vowel types for words containing /ō/⁵

		<i>defectivi</i>	<i>plene</i>
Type 33: <i>suffixed nota accusativi</i>	ׁōt-	1,031	357
Type 47: <i>plural suffix of hollow root noun</i>	-ō(t)	719	1,464

Vowel Stress (3 levels in all)

For an example of stress effects, consider Psalm 74:4b, where stress levels and spelling choice appear to interact:

↓ plene
 ↓ defective
 [suffix]
 ↓ defective
 ↓ plene
 שָׁמוּ אוֹתוֹת יָם אֲתוֹת:

Text Portion (76 in all)

Spelling choice also can be affected by text portion. For example, while *David* is always *defectivi* in Samuel (דָּוִד, x576), it is always *plene* in Chronicles (דָּוִד, x261). Or consider the association of spelling choice and text portion for Type 33. The percentage of opportunities spelled *plene* for Type 33 vowels in the Torah, Former Prophets, and Latter Prophets are 8%, 36%, and 51%, respectively (Andersen and Forbes 1986, 189).

- 3 Statistical sample size constraints limit the number of text portions to 76 or fewer. Where possible, natural boundaries have been adopted. For example, the Psalms are divided into six portions: Book I, Book II, Book III, Book IV, Book Va (Ps 107–119) and Book Vb (Ps 120–150). For specifics, see Table 2.
- 4 We excluded from the analysis 26 lexemes because their spelling was invariant (Andersen and Forbes 1986, 153–54). For example, אֲרָץ “oracle/oracle of” (x376) was omitted because it is always written *defectivi*.
- 5 For a discussion of Type 33 and Type 47, see Andersen and Forbes (1986, 189–91 and 197).

Transmissional Changes

While vowel type, vowel stress, and/or text portion can affect vowel choice, so can transmissional factors, such as copying errors or scribal fiat. For an instance of transmissional change, consider Ezek 41:16. *BHS* (p. 975, l. 13) has הַחֲלוֹנוֹת while L (p. 610, col. 1, l. 17)—as well as *BHK* (p. 882, l. 4) and *BHL* (p. 791, l. 17)—has הַחֲלוֹנוֹת. *BHS* disagrees with the L facsimile, a *plene* spelling having been mistakenly copied *defectivi*.

A Simple Analogy: Portion Apartment Buildings

For each text portion, envision a three-floor building having 65 apartments running along each side of the hallway on each floor. The situation is rendered in Figure 1. Imagine that each candidate vowel has its location in the apartment complex determined by its text portion (building), stress level (floor), position along the hallway (vowel type), and side of the hallway (spelling choice).

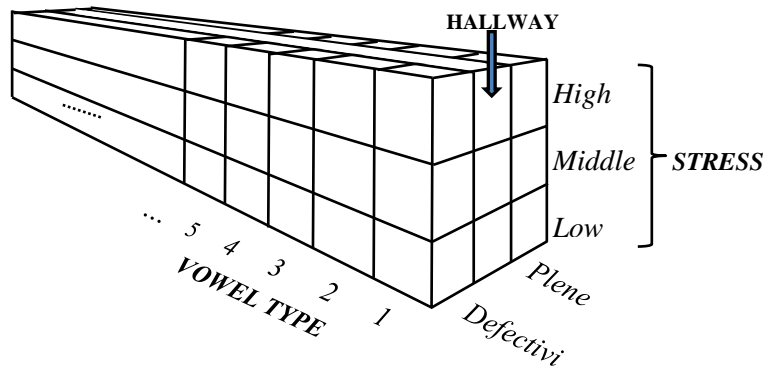


Figure 1: Vowel analogy—Apartment building structure for each portion

Given the fully populated apartment complex consisting of 76 identically built apartment houses (each house representing a text portion), our problem is as follows: Group the apartment houses based upon their occupancy similarities. In so doing, take into account that some occupants may have wandered to the wrong side of their hallway (due to copying noise/change).

Initial Results via Three Methods

To explore the BH spelling data properly, the analyst must be aware of, and responsive to, a range of statistical hazards.⁶ To understand the results produced, however, requires little knowledge of statistics. The reader might naturally wonder why the outcomes of three methods of analysis are important. The reason is simple: Mathematical methods may be *brittle*. That is, the answers that they yield may depend critically upon the detailed underlying assumptions. That being the case, it makes sense when analysing a body of data to apply several versions of a given method and/or several different methods and then to contrast the results produced by these approaches to assess their consistency.

Andersen and Forbes analysed the BH spelling data using four contrasting methods:

- *Contingency table analysis*: Analyse tables of counts. This is the main method used in Andersen and Forbes (1986). It is difficult and extensive. Thus, it is not treated here.
- *Clustering*: Create family tree (Andersen and Forbes 1986, 294–308; Forbes 1992a, 29–34; Forbes 1992b 95–103).
- *Multidimensional scaling*: Project high-dimension data onto a special low-dimension space (Forbes 1992b, 104–10).
- *Seriation*: Convert computed distances between text portions into a plot having text portions ordered temporally (Forbes 1992c, 125–34).

The minutiae of these methods are well beyond the scope of this essay.⁷ Rest assured that the methods have been carried out with great care and thorough scrutiny.⁸ For now, let us turn simply to critiquing the results yielded by the final three methods listed above.

The Results of Clustering

Figure 2 clusters BH portions based on vowel type, stress, and spelling choice. To produce the tree, algorithmic choices were necessary (Forbes 1992b, 94–96).⁹ So, “[t]he main value of this figure is [a practical aid to problem solving]. It identifies affinities among the portions which require interpretation and leads to further investigation by other techniques” (Forbes 1992b, 96).

6 The threatening statistical hazards are considered at length in Andersen and Forbes (1986).

7 Those interested in understanding our methods in detail are encouraged to study our two books on BH spelling: Andersen and Forbes (1986) and Freedman, Forbes, and Andersen (1992).

8 My work using statistical methods has benefitted enormously by advice received from Karen Kafadar, Commonwealth Professor at the University of Virginia and 2019 President of the American Statistical Society.

9 Technical note: Used Mahalanobis rather than Euclidean distance and cohesive rather than isolated clustering.

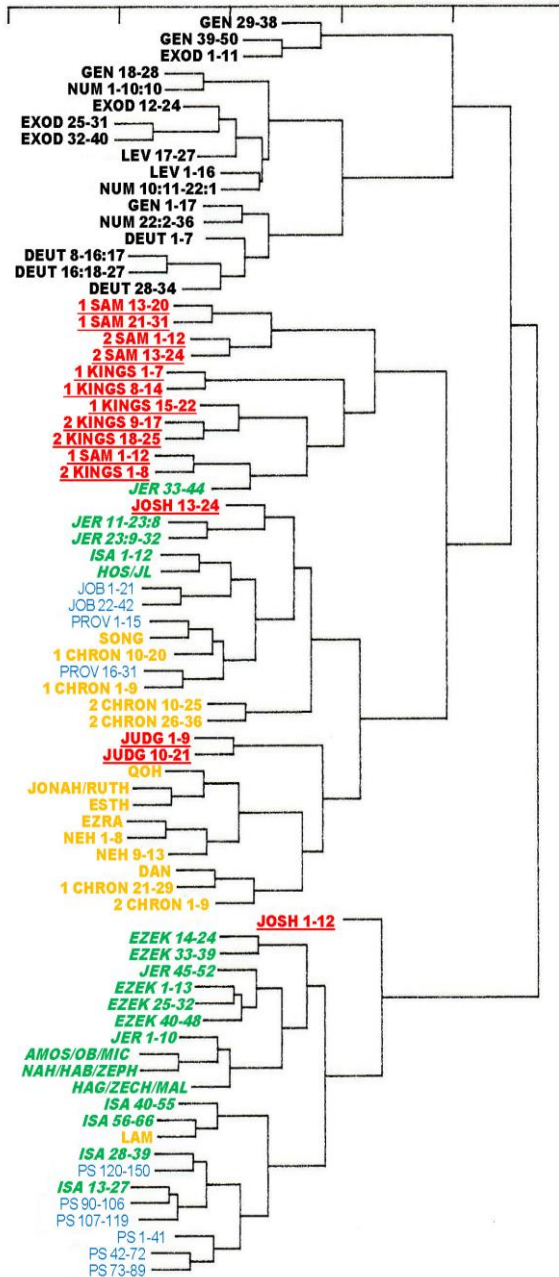


Figure 2: Spelling-based clustering of 76 BH text portions. Penn State University Press permission given. Colour and emphasis added: Torah: **Black**. Former prophets: **RED**. Latter Prophets: **GREEN**. Poetry: **BLUE**. Other Writings: **ORANGE**.

Affinities among the Portions

Note that 11 book portion pairs are sisters in the tree.¹⁰ For example, Exod 25–31 and Exod 32–40 form a pair. That pair then merges with Exod 12–24 to form an Exod trio.¹¹ Exod 1–11, however, pairs with Gen 39–50. The portions making up Deuteronomy form a single cluster, as do those making up Judges and also those making up Job. The six portions making up Psalms plus Isa 13–27 and Isa 28–39 also combine to form a cluster. The 17 portions of the Torah combine into a single grand cluster. The 11 portions making up Samuel-Kings plus Jer 33–44 also combine into a grand cluster.

Fractionations among the Portions

As regards curious group participation, consider Isaiah, Jeremiah, and Joshua. Isaiah 1–12 is far removed from the rest of Isaiah, merging first with Hosea-Joel and these portions then merging with Job. At the bottom of the tree diagram, Trito-Isaiah (Isa 56–66) merges first with Lamentations and the resulting pair merges with Deutero-Isaiah (Isa 40–55). Finally, as noted above, Isa 13–27 and Isa 28–39 appear in the Psalms cluster.

Jeremiah is scattered, appearing in the main Minor Prophets cluster (Jer 1–10), the Samuel-Kings cluster (Jer 33–44), and the Ezekiel cluster (Jer 45–52). The remaining portions of Jeremiah (the pair Jer 11–23:8 and Jer 23:9–32) combine with Josh 13–24. The other portion of Joshua (Josh 1–12) is orthographically the most atypical of the MT, joining the tree last by merging into a 21-portion cluster consisting of Isaiah, two portions of Jeremiah, Ezekiel, most of the Minor Prophets, and Psalms with Lamentations also included.

Falsifiability

Methods for checking the validity of clusters exist and have been used repeatedly in our work. Since the clustering method that we use is agglomerative, it produces clusters no matter what distance data are provided. For example, if we replace spelling choice by whether the given vowel appears in an even-numbered or odd-numbered verse and cluster the resulting data, as expected the resulting clusters are bizarre (Andersen and Forbes 1986, 307). For instance, the two closest portions are Exod 25–40 and Ezra-Nehemiah.

10 Reading Figure 1 from top down, the portion pairs are: Exod 25–31 and Exod 32–40, Deut 8–16:17 and Deut 16:18–27, 1 Sam 13–20 and 1 Sam 21–31, 2 Sam 1–12 and 2 Sam 13–24, 1 Kings 1–7 and 1 Kings 8–14, 2 Kings 9–17 and 2 Kings 18–25, Jer 11-23:8 and Jer 23:9–32, Job 1–21 and Job 22–42, 2 Chron 10–25 and 2 Chron 26–36, Judg 1–9 and Judg 10–21, and Ps 42–72 and Ps 73–89. (Number strings lacking semicolons refer to chapters.)

11 There are three such trios: Exod 25–31, Exod 32–40, and Exod 12–24; Deut 8–16:17, Deut 16:18–27, and Deut 28–34; Ps 1–41, Ps 42–72, and Ps 73–89.

The Results of Multidimensional Scaling

Detailed investigations of the observed behaviours may be found in Forbes (1992b, 98–103). Rather than going into these matters, we take up the next approach to analysis, classical multidimensional scaling (CMDS).

The BH spelling information allows computation of the distances between all pairs of portions (Forbes 1992b, 94–95). Multidimensional scaling extracts the maximum information from the distances, allowing construction of a two-dimensional plot (a map) showing the relative positions of the portions. If that map is judged inadequate, one may easily add an optimised third dimension (making a topographical map or its equivalent). Table 2 allows one to use a portion number to learn the portion’s researcher-specified limits. For example, portion 11 runs from Num 1:1 through Num 10:10, encompassing the “sojourn at Sinai” [P].

Table 2: Portion numbers to portion limits

1	Ge 18–17	20	Jd 1–9	39	Je 11–23:8	58	Jb 1–21
2	Ge 18–28	21	Jd 10–21	40	Je 23:9–32	59	Jb 22–42
3	Ge 29–38	22	1S 1–12	41	Je 33–44	60	Pr 1–15
4	Ge 39–50	23	1S 13–20	42	Je 45–52	61	Pr 16–31
5	Ex 1–11	24	1S 21–31	43	Ek 1–13	62	Jn/Ru
6	Ex 12–24	25	2S 1–12	44	Ek 14–24	63	SS
7	Ex 25–31	26	2S 13–24	45	Ek 25–32	64	Qo
8	Ex 32–40	27	1K 1–7	46	Ek 33–39	65	La
9	Le 1–16	28	1K 8–14	47	Ek 40–48	66	Es
10	Le 17–27	29	1K 15–22	48	Hs/Jl	67	Da
11	Nu 1–10:10	30	2K 1–8	49	Am/Ob/Mi	68	Er
12	Nu 10:11–22:1	31	2K 9–17	50	Na/Hb/Zp	69	Ne 1–8
13	Nu 22:2–36	32	2K 18–25	51	Hg/Zc/Ml	70	Ne 9–13
14	De 1–7	33	Is 1–12	52	Ps 1–41	71	1C 1–9
15	De 8–16:17	34	Is 13–27	53	Ps 42–72	72	1C 10–20
16	De 16:18–27	35	Is 28–39	54	Ps 73–89	73	1C 21–29
17	De 28–34	36	Is 40–55	55	Ps 90–106	74	2C 1–9
18	Js 1–12	37	Is 56–66	56	Ps 107–119	75	2C 10–25
19	Js 13–24	38	Je 1–10	57	Ps 120–150	76	2C 26–36

Figure 3 shows the relative positions of the 76 portions in two dimensions (Forbes 1992b, 108–09). This plot manages to exploit only 32% of the information in the distances.

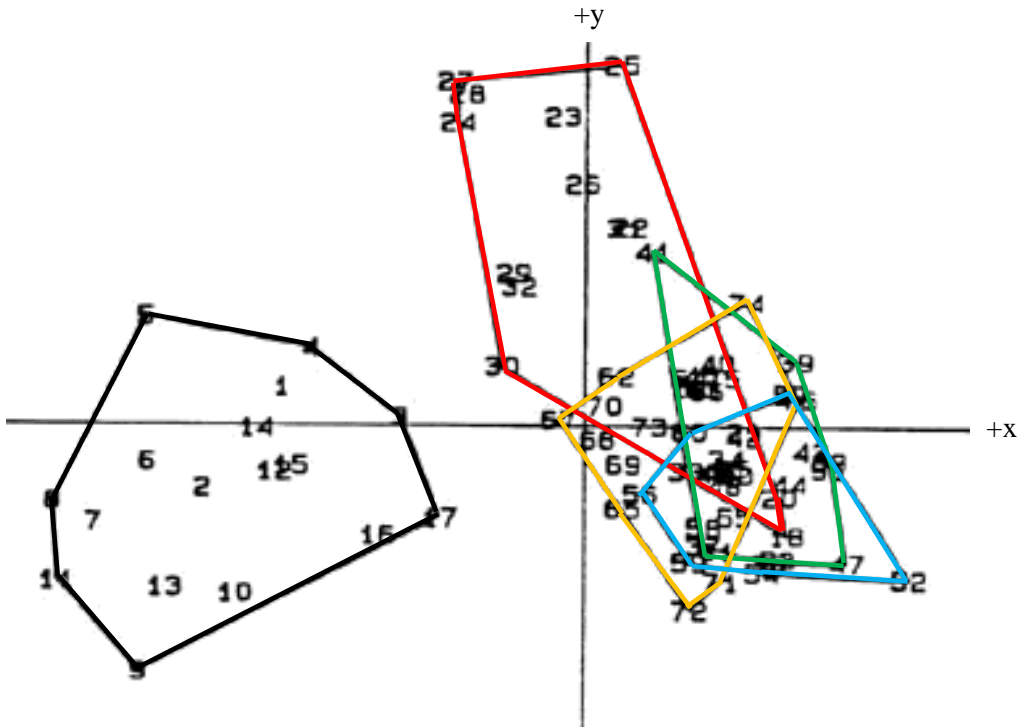


Figure 3: Multidimensional scaling x-y plot for 76 portions. Penn State University Press permission given. Colours added: Torah: **Black**. Former Prophets: **Red**. Latter Prophets: **Green**. Poetry: **Blue**. Other writings: **Orange**.

In Figure 3, the portions making up each section of the Hebrew Bible are bounded by (colour coded) irregular polygons. Thus, since portions 1–17 make up the Torah, these are all located within the (black) Torah polygon.¹²

Several behaviours are immediately evident:

1. As was the case with the family tree in Figure 2, the Torah is decisively separated from the rest of the MT.
2. Among the Former Prophets, the books of Samuel and Kings scatter around the positive y-axis, while Joshua and Judges are down among the other sections.
3. The Latter Prophets, Poetry, and Other Writings sections appear to be grossly intermixed in Figure 3. Do these portions constitute a spelling *mélange*, or do we

¹² For those viewing the figures without the colours: The Torah polygon boundary contains portion number 17 at its far right. The Former Prophets polygon boundary contains portion number 25 at its top right. The Latter Prophets polygon boundary contains portion number 47 (with the 4 hidden) at its bottom right. The Poetry polygon boundary contains portion number 52 (with the 5 hidden) at its bottom right. The Other Writings polygon boundary contains portion number 74 at its top. These portion numbers can also be located on the boundaries in Figure 4.

need to add z-axis information, perhaps to lessen the overlap? Figure 4 shows the x-z plot for our data.

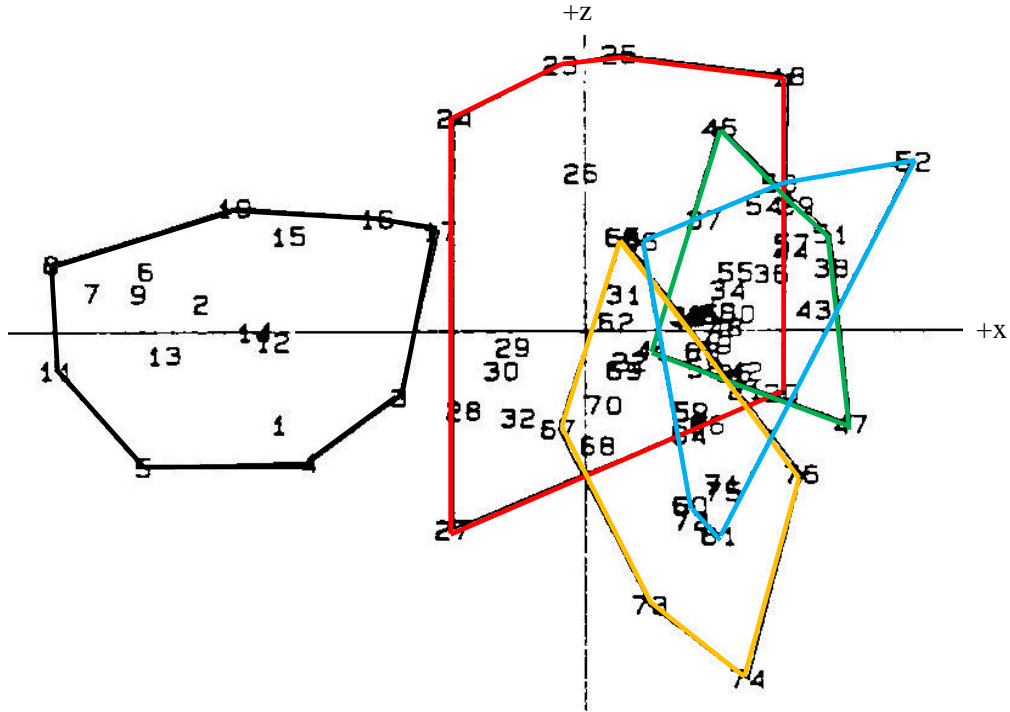


Figure 4: Multidimensional scaling x-z plot for 76 portions. Penn State University Press permission given. Colours added: Torah: **Black**. Former Prophets: **Red**. Latter Prophets: **Green**. Poetry: **Blue**. Other writings: **Orange**.

Table 3: Portion positions with respect to x-y plane

Section	Torah	Former Prophets	Latter Prophets	Poetry	Other Writings
Above x-y	59%	47%	74%	60%	14%
Below x-y	41%	53%	26%	40%	86%

We see that the Latter Prophets and Other Writings mostly lie on opposite sides of the x-y plane. Together, Figures 3 and 4 exploit only 39% of the inter-portion distance information. That is, almost two-thirds of the spelling information is not used.

The Results of Seriation

Grouping text portions by successively gathering the most similar portions (clustering) or by optimally extracting positional information about their position in an abstract space (multidimensional scaling) yields interesting, if problematic, results. But ordering them in time would be a significant achievement. The method of seriation seeks to do exactly this. The behaviours and technical apparatus underlying seriation (ordering objects in time) are unfortunately daunting (Forbes 1992c, 125–34). Ideally, seriation of the spelling data would order the portions along a straight line. But it has long been known that positioning objects on a curved band¹³ is the likely outcome of seriation of noisy, real world data (Kendall 1971, 225–29). In practice, however, uncatered for variables and transmissional noise cause the band to be rather “puffy,” as the scraggly nature of many of the results obtained by Kendall (1971, 228 and 232–35) attest. For our spelling data, puffiness is decidedly found in the thirty-portion seriation shown in Figure 5.

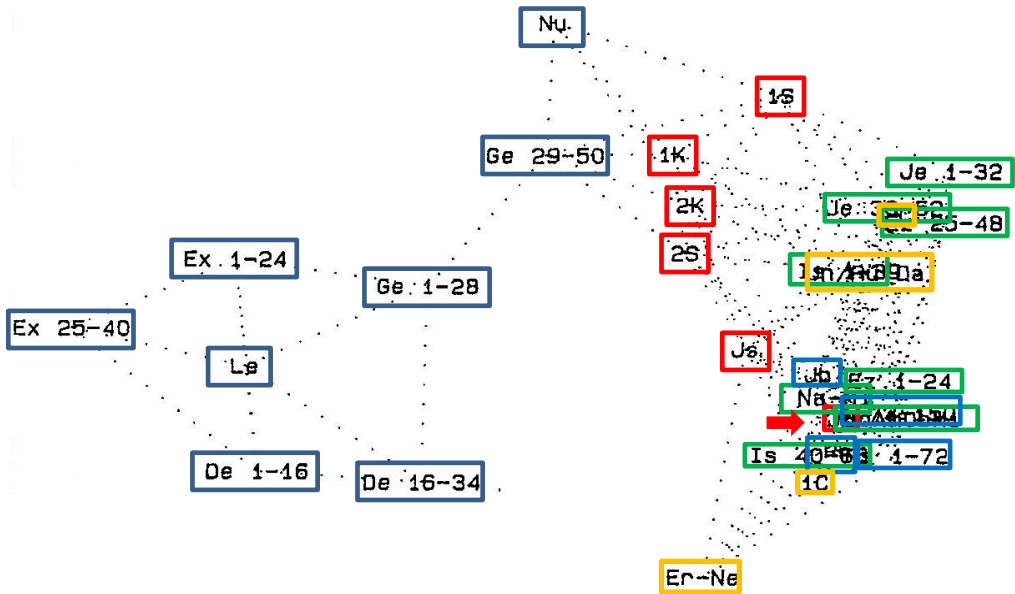


Figure 5: Spelling-based seriation of 30 BH text portions. Penn State University Press permission given. Colours added.

¹³ The curve is said to be a *horseshoe*, even though it can manifest as a closed circle-like curve (Kendall 1971, 225–27).

As with the two other sets of results, the seriation findings are mixed. While the ordering of some portions does not surprise, others are puzzling.

Specific Seriation Findings

Amplifying the clustering and scaling results, the entire Torah comes earliest.¹⁴ Next come the Former Prophets with Joshua somewhat domesticated while puzzling Judges (pointed at by the added arrow) ambles along much later. Among the Latter Prophets, Proto-Isaiah comes earlier than Deutero-Isaiah/Trito-Isaiah. The two portions of Jeremiah are close in time, but the first part of Ezekiel is positioned much later than the second part. The two portions of the Minor Prophets are close in time, but Nahum-Malachi come slightly earlier than the others. As for the Poetry, Job comes earliest, followed by Ps 73–150 and then by Ps 1–72 with Proverbs co-temporaneous. The behaviour of the Other Writings is also puzzling: 2 Chronicles is located between the second parts of Jeremiah and of Ezekiel, while 1 Chronicles is a bit later than Deutero-Isaiah/Trito-Isaiah. Our ill-advised combining of portions 62–67¹⁵ is positioned with Proto-Isaiah! And, of course, Ezra-Nehemiah conclude the temporal parade.

These quirky orderings remind us that “[s]eriation orders the portions in time based on the received spelling. Hence, because of the possibility of extensive spelling updating and/or copying errors, the relative times associated with portions may bear little relation to the time(s) of their creation” (Forbes 1992c, 134).

Concluding Remark

While the results produced by each of the methods reported share encouraging features, each also raises warnings that all is not well. The path ahead involves considerable work, especially concerning incorporation of an additional spelling feature. Regarding enhancing the noise-tolerance of seriation, help is on the way. Work in this latter area will be introduced in an upcoming paper.

References

- Andersen, F. I., and Forbes, A. D. 1986. *Spelling in the Hebrew Bible*. Rome: PBI.
- Andersen, F. I., and Forbes, A. D. 2013. “Matres Lectionis: Biblical Hebrew.” In *Encyclopedia of Hebrew Language and Linguistics*, edited by G. Khan, II, 607–11. Leiden: Brill.
- Bange, L. A. 1971. *A Study in the Use of Vowel-Letters in Alphabetic Consonantal Writing*. Munich: UMI-DRUCK.

14 Seriation does not supply the direction of time in its plots. Here, it is assumed that readers will accept that Ezra-Nehemiah marks the latest time in the plot.

15 This mélange consists of Jonah, Ruth, Song of Songs, Qohelet, Lamentations, Esther, and Daniel.

- Cross, F. M., and Freedman, D. N. 1952. *Early Hebrew Orthography: A Study of the Epigraphic Evidence*. New Haven: ASOR.
- Forbes, A. D. 1992a. "A Tutorial on Method: A Guide for the Statistically Perplexed." In *Studies in Hebrew and Aramaic Orthography*, edited by D. N. Freedman, A. D. Forbes, and F. I. Andersen, 17–35. Winona Lake, IN: Eisenbrauns.
- Forbes, A. D. 1992b. "Choice of Statistical Methods." In *Studies in Hebrew and Aramaic Orthography*, edited by D. N. Freedman, A. D. Forbes, and F. I. Andersen, 93–110. Winona Lake, IN: Eisenbrauns.
- Forbes, A. D. 1992c. "The Seriation of Portions." In *Studies in Hebrew and Aramaic Orthography*, edited by D. N. Freedman, A. D. Forbes, and F. I. Andersen, 125–34. Winona Lake, IN: Eisenbrauns.
- Forbes, A. D. 2019. "Ground-Truth, S-Curves, and Reduced Horseshoes." Paper presented at the annual meeting of the Society of Biblical Literature, San Diego, California, November 23–26.
- Forbes, A. D., and Andersen, F. I. 2012. "Dwelling on Spelling." In *Diachrony in Biblical Hebrew*, edited by C. Miller-Naudé and Z. Zevit, 127–45. Winona Lake, IN: Eisenbrauns. <https://doi.org/10.1515/9781575066837-009>
- Freedman, D. N., Forbes, A. D., and Andersen, F. I. 1992. *Studies in Hebrew and Aramaic Orthography*. Winona Lake, IN: Eisenbrauns.
- Kendall, D. G. 1971. "Seriation." In *Mathematics in the Archaeological and Historical Sciences*, edited by F. H. Hudson, D. G. Kendall, and P. Tăutu, 215–52. Edinburgh: Edinburgh University Press.
- Zevit, Z. 1980. *Matres Lectionis in Ancient Hebrew Epigraphs*. New Haven: ASOR.