

方言測量學方法的穩定性檢驗

Testing the Stability of Dialectometric Methods

◎ 黃 河、宋華強

提 要：本文以語保工程項目全國範圍內新派和老派單字音調查材料為數據，以測試方言測量學能否抵抗社會變異的干擾，呈現底層的語言地理結構。在通過二項式分佈檢驗確定新派和老派單字音數據存在顯著差異的前提下，我們分別計算新派和老派數據集的語言距離，通過相關分析、降維分析，以及普通話、中古音、上古音等參照係，著重比較了新派、老派單字音地理分佈的差異。我們發現方言測量學的集成數據分析法具有良好的穩定性。

關鍵詞：漢語方言 方言地理學 方言測量學 地理語言學

Key words: Chinese dialects; dialect geography; dialectometry; geolinguistics

一、前言

地理語言學研究語言的地理分佈和空間擴散，早期稱之為方言地理學，主要是研究單個語言變項的地理分佈，然後根據地理分佈重建該語言變項的歷史，其背後依據的理念是“特定的地理過程塑造特定的歷史過程”。上世紀 70 年代之後，地理語言學迎來了計量革命，方言測量學（Dialectometry）應運而生。方言測量學是計量語言學和方言學的交叉學科，該方法以語言地圖集、語料庫為數據來

源，通過語言距離的測量、空間可視化等手段探索方言空間分佈。Séguy（1973）首次在羅曼語語言地圖集上嘗試使用方言測量學的方法探索方言之間的差異，在此後五十年的發展中，湧現出薩爾茲堡學派（Salzburg school）、格羅寧根學派（Groningen School）等流派。前者側重於範疇型數據的語言距離計算，並開發出一系列語言距離的空間可視化方法（Goebel 1982）。後者側重於基於字符串或詞頻的語言距離計算，並探索穩健的、適用於地理語言學的聚類和降維技術（Nerbonne 2008）。

方言測量學的基本步驟是計算各個方言點之間的語言距離，得到一個“地點x地點”的語言距離矩陣，可以把某個地點設置為參照係，呈現其他地點到這個參照係之間的語言距離，並可視化為參照係地圖（reference map）；然後，我們可以採用降維分析、聚類分析等無監督算法對語言距離矩陣做進一步的分析，以識別地理語言的底層結構。

方言測量學的本質是一種集成數據分析（aggregate data analysis）。集成數據分析是分類學中常用的一種計量方法，已經在語言類型學（Greenberg 1963）、方言測量學（dialectometry, Séguy 1971; Geobl 1982）、多維語域研究（multidimensional register studies, Biber 1988）中得到廣泛的應用。我們知道，不同的語言變項有着不同的地理分佈，應盡量避免給單個語言變項提供特設性（ad hoc）的解釋，陷入“只知樹木，不見森林”的困境。方言測量學是採用計量手段綜合考慮一個數據集中的所有語言變項，以探索整個數據集的全貌，“通過對不同的語言變項做集成分析以加強信號”（Nerbonne 2010）。

語言距離的測量是方言測量學的核心課題，目前較為流行的測量方法是萊氏距離，又叫萊文施坦距離（Levenshtein Distance）、編輯距離（Edit Distance），這種方法把國際音標轉寫的字音當做個字符串，然後計算字符串之間的距離，作為詞形之間的語言距離。黃河（2022）首次將方言測量學的方法應用到漢語方言上，並提出杰卡德加權距離（Weighted Jaccard Distance）的算法，詳述了萊氏距離和杰卡德加權距離各自的特點和局限性，並用方言測量學的方法討論了語言的結構化程度和各向漸變性的問題。我們知道，漢語有聲調，這是和大部分印歐語不同的地方。方言測量學的語言距離測

量方法最早基於英語、荷蘭語、德語、法語等主要的印歐語開發，這些語言因為缺乏聲調，因此至今方言測量學在聲調距離的測量上仍然薄弱。

本文旨在討論一個問題：就全國漢語方言的範圍，不同方言點的社會變異（如新派、老派字音的不同）多大程度上會幹擾方言測量學識別底層的方言地理結構？即方言測量學這種集成數據分析的方法在識別漢語方言地理面貌上的穩定性如何？

漢語方言有新派和老派字音的差別，方言學家知道新老派字音之間往往存在明顯差異，比如蘇州城關的老派發音人尚有分尖團者，新派則很難見到分尖團的人。在社會語言學上已經有諸多研究，年齡是社會方言變體的重要參量（Labov 2001）。另一方面，就同一個地點而言，新派和老派口音儘管存在差異，但是畢竟是同一個地點方言的社會變體，相較於不同地點的方言，理應更為接近。大多數情況下，社會變異比地理變異要小。因此，我們想要用一片區域的大量新派和老派字音的材料來探索方言測量學的穩定性，假如一片區域中的每個地點分別挑選一位老年、年輕發音人作為該地方言新老派口音的代表，然後我們就得到這片區域新派和老派兩組數據集。Yang & Castro（2008）討論了聲調距離的測量問題，其中把五度制標調和音段一起作為字符串計算語言距離。在使用方言測量學計算新派和老派的語言距離時，我們採用同樣的方法，將漢語方言的五度制標調納入到語言距離的計算中。

我們先用二項式分佈檢驗考察新派、老派原始數據之間的差異，然後用方言測量學方法計算兩組數據集中不同地點之間的語言相似度。我們先用相關分析來探索語言相似度之間的差異，然後確認新派和老派的語言地理分佈有無明顯差異。假如它們的地理分佈差不多，則說明新派和老派的社會變異

不會干擾方言測量學識別底層的方言結構，集成數據分析具備相當的穩定性。假如新派和老派的地理分佈有明顯的差異，則說明方言測量學的穩定性不夠，會受到社會變異的影響。

二、材料與方法

我們首先分別計算新派和老派各個地點之間的語言相似度，然後對新派和老派的語言相似度矩陣做相關分析。相關係數可以從整體上反映語言相似度之間的關係，但是缺失了空間信息。為了更好地比較新派、老派單字音的地理格局的差異，我們不僅採用降維方法呈現各數據集的地理分佈，還設置了普通話、中古音（鄭張 - 潘體系）、中古音（白 - 沙體系）以及上古音（鄭張 - 潘體系）作為參照係，然後計算地圖上各個地點到參照係之間的相似度，並進行可視化。

2.1 材料

本文的研究材料來自語保工程項目^[1]，語保工程項目利用現代化技術手段記錄漢語方言、少數民族語言和口頭語言文化的實態語料，構成大規模的多媒體語言資料庫，並開展語言資源保護研究工作，其中漢語方言數據分為老派和新派兩組數據，部分方言點存在數據缺失。去掉存在缺失數據的方言點，我們得到 1101 個方言點包含 994 個字的新派和老派的字音數據，這些新派數據和老派數據是成對的（pairwise）。該數據是近年來統一組織調查的大型數據庫，布點密度高，字音數量多，我們用它來討論本文的議題是十分合適的。

如上文所述，為了檢驗新派和老派的單字音分佈有何種差異，我們設置了普通話、中古音（鄭張 - 潘體系）、中古音（白沙體系）以及上古音（鄭張 -

潘體系）四套參照系統。我們把四套參考系統納入到後續的語言距離的計算中去，分別給新派、老派數據的 994 個字註上普通話以及中古音（鄭張 - 潘體系）、中古音（白沙體系）以及上古音（鄭張 - 潘體系）擬音。中古音擬音（鄭張 - 潘體系）、上古音擬音（鄭張 - 潘體系）參考鄭張尚芳（2003），中古音擬音（白 - 沙體系）參考 Baxter & Sagart（2014）。

我們採用了五度制作為聲調編碼方式。五度制標音指的就是傳統方言學的五度制標調法（Chao 1930），這裡相當於把五度制標調緊隨記錄音段的國際音標，整個當做一個字符串計算標準化萊氏距離。以北京話的“家”這個字為例，編碼為 tciA55 。語保數據庫中的新派、老派原始數據採用的就是五度制標調法，中古音（鄭張 - 潘體系）、中古音（白 - 沙體系）對聲調調值的構擬採用的也是五度制標調法。上古音（鄭張 - 潘體系）聲調可能是韻尾差異的伴隨特征，因此將韻尾納入語言距離的計算就已經涵蓋了上古聲調的差異，上古平聲是零形式，上聲帶喉塞尾，去聲帶 -s 尾，入聲帶 -p、-t、-k 尾。上古音參照係在新派和老派數據集中都是一樣形式。

2.2 新派、老派數據集的差異檢驗

我們需要事先確認原始數據中新派、老派數據到底存在多大的差異。假如新派、老派的數據不存在顯著差異，那麼方言測量學得到的結果也不存在差異，就是理所當然的，並不能說明方言測量學在識別漢語方言地理面貌上具有抵抗社會變異的穩定性。如果新派和老派數據集的差異是顯著的，如果方言測量學的結果看不到明顯差異，能獲得類似的地理語言底層模式，才能說方言測量學的算法具有穩定性。因此，我們下面採用二項分佈參數檢驗來看新派和老派原始數據之間是否具有顯著差異。

我們匹配新派和老派數據，在所有地點的

1094394 個字音中有 270724 個字音是存在差異的，相異字音佔比為 24.7%。二項分佈參數檢驗的零假設是新派和老派相異的概率為 0，p 值為 2.2e-16，結果說明新派和老派數據集是顯著相異的。因此，我們可以用新老派數據檢驗方言測量學的穩定性。

我們知道，字音的地理分佈存在一定的差異，我們想要綜合所有字音，在宏觀視角下計算這些方言的語言距離，然後基於綜合的語言距離呈現方言的地理分佈。我們的數據是每個地點一大批單字音的國際音標形式，十分適用於計算萊氏距離。萊氏距離，又叫萊文施坦距離（Levenshtein Distance），又叫編輯距離（Edit Distance），最早是由蘇聯數學家萊文施坦開發，用於計算兩個字符串之間的距離（Levenshtein 1965）。這種方法將每個方言點的某個字音的國際音標形式當做一個字符串，在計算兩個方言點某個字音之間的距離時，考察一個形式轉化為另一個形式的最小操作步數。萊氏距離只允許三種基本的操作：刪除、增音和替換。假如有兩個語言的〈馬〉的語音形式分別為 horse 和 ros，它們之間的萊氏距離等於 horse 轉化為 ros 的最小操作步數。從 horse 到 ros 之間的操作步數存在多個解，取決於音段之間的匹配方式，假如把 horse 裡的 r 和 ros 開頭的 r 匹配，如表 2，得到兩者轉換的操作步數為 4。倘若把兩個詞形按照表 3 裡的方式匹配和轉換，得到的操作步數為 3。因此，不同的音段匹配方式會產生不同的結果。要獲得最小操作步數，有兩種常用的方法：（1）遍歷兩個詞形的所有匹配方案，並計算出每種匹配方案下的操作步數，然後取其中的最小值；（2）採用二維表可以快速計算出最小操作步數。如表 4，二維表將兩個詞形分別寫於表格的橫行和縱列，並從 0 開始依次編號。每個格子的值等於 $\min(\text{左格} + \text{插入}, \text{上格} + \text{刪除}, \text{對角格} + \text{替換})$ ，

即每個格子裡的數值是這個格子左邊格子數值加上插入步數，上方格子數值加上刪除步數，對角格子的數值加上替換步數三個數值中的最小值，如表 3 中 X1 的數值為 $\min(1+1, 1+1, 0+1)=1$ ，以此類推，X2、X3、X4 分別為 2、2、1。按照以上計算規則算出表格中所有格子的數值，如表 5，表 5 右下角格子的數值即為最小操作步數 3。

表 2 第一種匹配方案下 horse 與 ros 的萊氏距離

	語音形式					操作	操作步數
	h	o	r	s	e		
輸入：詞形 1						刪除 h	1
						刪除 o	1
						將 s 替換為 o	1
						將 e 替換為 s	1
輸出：詞形 2							4

表 3 第二種匹配方案下 horse 與 ros 的萊氏距離

	語音形式					操作	操作步數
	h	o	r	s	e		
輸入：詞形 1						將 h 替換為 r	1
						刪除第二個 r	1
						刪除 e	1
輸出：詞形 2							3

表 4 二維表初始樣態

		r	o	s
	0	1	2	3
h	1	X1	X3	
o	2	X2	X4	
r	3			
s	4			
e	5			

表 5 二維表結果樣態

		r	o	s
	0	1	2	3
h	1	1	2	3
o	2	2	1	2
r	3	2	2	2
s	4	3	3	2
e	5	4	4	3

用這種方法計算得到的萊氏距離存在一個問題，通常來講，越長的詞之間轉換所需的操作步數越多，得到的萊氏距離也就越大。因此，我們需要對萊氏距離進行標準化處理，得到標準化萊氏距離 (Standardized Levenshtein Distance)。標準化的方式有多種，最常見的方式是將計算得到的最小操作步數除以兩個詞中較長一方的詞長。還是以上面 horse 和 ros 的距離計算為例，兩者的最小操作步數為 3，horse 比 ros 長，horse 的長度為 5，3 除以 5，兩個詞的標準化萊氏距離為 0.6。這樣計算得到的標準化萊氏距離始終位於 0 到 1 之間，距離越接近 1，表示兩個詞的距離越大，反之亦然。萊氏距離的操作過程可以看做是對音變過程的模擬，它假設從一個形式演變為另一個形式遵循最短路徑。

我們來看漢語的例子，我們以北京話、成都話、陽江話的“牛”的讀音為例來說明標準化萊氏距離的計算原理。三種方言“牛”的字音依次為 niou35、niəu21、ŋeu43。我們首先來看北京話和成都話字音之間的萊氏距離。如表 6，從北京音到成都音之間需要 4 步操作，然後除以字符長度 6，得到標準化萊氏距離為 0.667。

表 6 北京話和成都話“牛”的萊氏距離

	語音形式					操作	操作步數	
輸入：北京音	n	i	o	u	3	5	將 n 替換為 ɲ	1
	ɲ	i	o	u	3	5	將 o 替換為 ə	1
	ɲ	i		u	3	5	將 3 替換為 2	1
	ɲ	i		u	2	5	將 5 替換為 1	1
輸出：成都音	ɲ	i		u	2	1		4

然後看北京音和陽江音之間的萊氏距離。如表 7，從北京音到陽江音之間需要 5 步操作，然後除以兩者字符中長度較長的長度 6，得到標準化萊氏距離為 0.833。

表 7 北京話和陽江話“牛”的萊氏距離

	語音形式					操作	操作步數	
輸入：北京音	n	i	o	u	3	5	將 n 替換為 ɲ	1
	ɲ	i	o	u	3	5	將 o 替換為 ə	1
	ɲ	i	ə	u	3	5	將 3 替換為 4	1
	ɲ	i	ə	u	4	5	將 5 替換為 3	1
	ɲ	i	ə	u	4	3	刪除 i	1
輸出：陽江音	ɲ		ə	u	4	3		5

成都話和陽江話之間的萊氏距離。如表 8，從成都音到陽江音之間需要 5 步操作，然後除以兩者字符中長度較長的長度 6，得到標準化萊氏距離為 0.833。

表 8 成都話和廣州話“牛”的萊氏距離

	語音形式					操作	操作步數	
輸入：成都音	ɲ	i	ə	u	2	1	將 ɲ 替換為 ɲ	1
	ɲ	i	ə	u	2	1	將 ə 替換為 ə	1
	ɲ	i	ə	u	2	1	將 2 替換為 4	1
	ɲ	i	ə	u	4	1	將 1 替換為 3	1
	ɲ	i	ə	u	4	3	刪除 i	1
輸出：陽江音	ɲ		ə	u	4	3		5

因此，我們可以得到三個地點“牛”這個字音的標準化萊氏距離矩陣，如表9：

表9 三個地點“牛”的字音的標準化萊氏距離矩陣

	北京話	成都話	陽江話
北京話	0	0.667	0.833
成都話	0.667	0	0.833
陽江話	0.833	0.833	0

對任意兩個地點j和k，計算出所有的字音距離，然後取算術平均值，便得到地點j和k之間的方言距離。計算出的結果儲存為“地點×地點”的語言距離矩陣（distance matrix）。如表9，該矩陣是無向的，即j地與k地的語言距離等於k地與j地的語言距離，矩陣中的語言距離沿對角線對稱。因為對角線是某個地點和自己的距離，始終為0。

只需要用1減去萊氏距離，就可以很容易轉化為萊氏相似度。在本文的計算中，我們把普通話、中古音（鄭張-潘體系）、中古音（白沙體系）以及上古音（鄭張-潘體系）四套參照系統也看做四個語言系統納入到語言距離的計算中去，並轉化為相似度。這樣就可以得到每個地點的方言到四套參照系統的語言相似度。經過計算，我們得到新派和老派兩個語言距離矩陣，以及新派、老派各個地點到四套參照系統的語言相似度。

2.4 相關分析（Correlation analysis）

我們採用皮爾遜積矩相關係數（Pearson product-moment correlation coefficient）計算新派和老派語言相似度矩陣之間的相關係數，並檢驗相關性的顯著程度。相關分析返回一個相關係數r值，以及p值。r值介於-1至1之間，正負表示正相關或負相關，絕對值表示相關性的大小，r值接近0表示不相關。p值小於0.05表示相關性顯著，小於0.01表示相關性極為顯著。

2.5 多維縮放（Multidimensional Scaling，

MDS）

多維縮放是方言測量學常用的降維技術，多維縮放的目的是找到一個低維空間，將不同的方言點表示為空間中的點，從而使這些點在低維空間中的距離盡可能精確地反映地點之間的語言距離。兩個方言越相似，它們在多維縮放空間中的距離就越近，反之亦然Cox & Cox（2001）。多維縮放在方言測量學中已經有了較多的應用，用於識別隱藏在高維方言數據底層的方言地理模式（Heeringa & Nerbonne 2001; Nerbonne & Kleiweg 2003）。本文基於新派、老派的語言距離矩陣做多維縮放，分別提取前三個維度。

2.6 多維縮放地圖（MDS map）和參照點地圖（Reference map）

多維縮放返回每個地點在前三個維度中的坐標值，三個維度可以分別可視化為三幅多維縮放地圖。多維縮放地圖以中國國境線為輪廓，以調查點所在位置生成泰森多邊形，將方言點的點數據（point）轉化為面數據（polygon），這樣做的好處是可以通過多邊形填色表示該方言點屬性數值大小，清晰而美觀地呈現各個維度所反映的空間結構。參照點地圖是以上述參照系統為參照點，用地圖呈現每個方言點和參照系統的語言相似度的大小，和多維縮放地圖一樣，也用泰森多邊形進行可視化。

三、結果

3.1 相關分析

相關分析的結果表明，新派、老派語言相似度矩陣之間的相關係數為0.96，p值<0.01。該結果說明，新派和老派的語言相似度矩陣之間呈現顯著的正相關關係，且有着較高的相關性。讓我們回顧2.2小節中新派和老派原始數據的差異檢驗，所有地點的1094394個字音中，有270724個字音是存在差異

的，相異字音佔比為 24.7%。二項分佈參數檢驗表明新派和老派的原始數據之間存在顯著的差異。但是，通過方言測量學的集成分析之後，新派和老派的語言相似度矩陣之間的相關係數達到 0.96。正如 Nerbonne (2010) 所言，方言測量學的集成分析可以在相當程度上“通過對語言變項做集成分析以加強信號”。相關係數忽悠了新派和老派的語言距離矩陣中的空間信息，因此，單憑較高的相關係數無法說明新派和老派的語言距離的地理分佈是一致的。

3.2 多維縮放地圖

圖 1 是新派和老派的多維縮放地圖，左列三幅圖為新派 (Young)，右列三幅圖為老派 (Old)，從上至下依次是多維縮放的第一、二、三維度的地圖。採用紅色和藍色表示每個維度的兩極，中間數值用白色表示。我們發現，從整體上講，第一維度體現了從官話至東南方言的龐大的地理連續體，該地理連續體可能是由自古以來從北向南的移民導致的。東南沿海方言較為保守，而北方官話帶有更多的創新形式。其中尤為矚目的是“秦嶺 - 淮河”一線似乎存在較為陡峭的分界線，該線以北是北方官話，有着更深的藍色，而該線以南的南方官話顏色更淺，這說明南北官話在單字音層面存在較大的差異。第二維度的地圖中值得注意的是紅色一極存在兩個中心，一個是晉語，一個是吳語，這說明就單字音層面，晉語和吳語存在諸多共性。第三維度的藍色區域主要是長江流域，但是中游位置幾乎被白色和紅色區域阻斷，大致可以視為長江流域與非長江流域的對立，體現了長江作為中國南方最為重要的水系對塑造漢語方言的地理格局起到了重要的作用。長江中游位置幾乎被白色和紅色阻斷，這些區域主要集中在江西、安徽一帶，這是歷史上北方移民南下遷徙的主要通路，攜帶北方方言特征的移民衝擊這塊區域，使得這片區域的方言有混雜的方言特征，因此，從地圖上看，長江流域的藍色在這一區域幾乎被其

他顏色阻斷。通過比較同一行左右兩幅地圖，我們會發現，新派、老派三個維度的地圖都十分相似。這說明新派和老派的語言距離的主要維度是一致的，都能反映漢語方言底層的音韻地理結構。

3.2 參照點地圖

圖 2 是新派和老派的參照點地圖，左列四幅地圖反映新派 (Young) 的情況，右列四幅地圖反映的是老派 (Old) 的情況，從上至下依次是以普通話 (Putonghua)、中古音鄭張 - 潘擬音 (Middle Chinese_zz)、中古音“白 - 沙”擬音 (Middle Chinese_bs)、上古漢語 (Old Chinese) 為參照點，呈現每個方言點與四個參照系統的語言相似度，顏色越偏向於深藍色表示該方言和參照系統的語言相似度越低，顏色越偏向於黃色，表示該方言和參照系統的語言相似度越高。不管是從哪種參照系統來看，新派和老派方言在單字音層面的差異看都不是很明顯。

以普通話為參照系統的地圖呈現了較為清晰的地理格局，北京以及北京周圍的方言呈現黃色，說明和普通話的相似度很高，其次是東北方言，東南方言和普通話的相似度很低，這些都和我們既有的認識相符。值得注意的是晉語、相當一部分西北方言以及山東南部的方言也呈現了較深的藍色，說明這些方言和普通話的語言相似度並不高，這進一步說明了在單字音的層面，這些方言存在諸多存古的特點。

再來看以中古音為參照系統的兩幅地圖，由於這兩幅地圖的參照系統是中古擬音，因此所呈現的地理分佈存在較大的噪音。不過，我們還是可以據此分析整體上的分佈格局，提取有效的信息。“鄭張 - 潘”系統和“白 - 沙”系統是兩個不同的擬音系統，因此兩類地圖存在一定的差異，但是還是能看到兩者具有共性的地方：就單字音層面，和中古音系統差異較大的是晉語區北部和安徽境內的江淮官話，

比較接近中古音系統的是粵語。我們應該認識到一點，保存中古音前的語音特征、發生演變進而偏離中古音都會造成和中古音的相似度變低。

最後來看以上古音為參照系統的地圖，所有方

言和上古音的相似度都比較低，我們可以看到圖例中的數值區間為 0-0.2。其中，相似度最低的是晉語和山東的部分方言點，相似度相對較高是廣東、台灣的客家話。

圖 1 新派和老派的多維縮放地圖

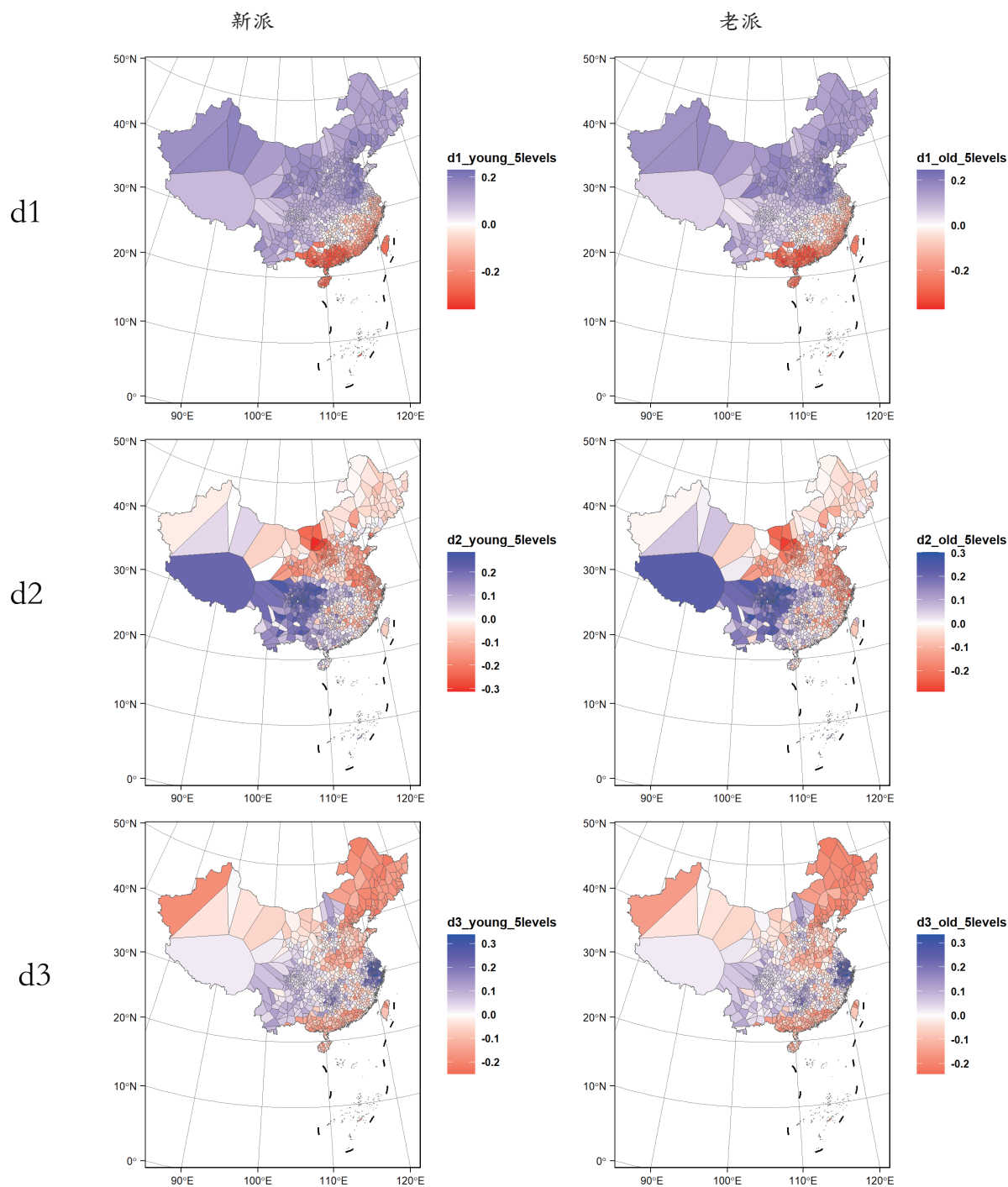
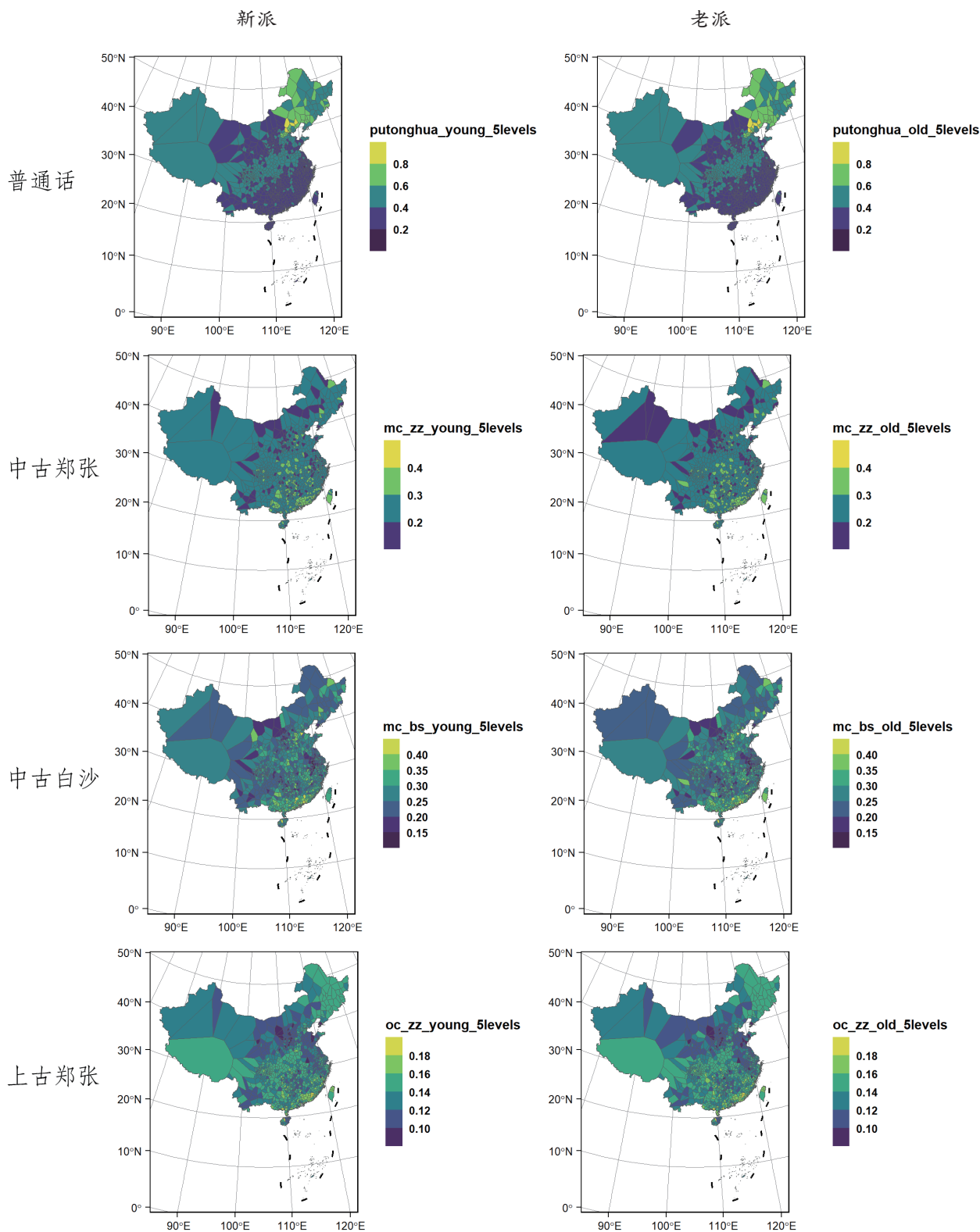


圖 2 新派和老派的參照點地圖



四、討論

無論是相關分析還是多維縮放地圖、參照點地圖都指向同一個結論。原始數據存在顯著差異的新派和老派字音，通過方言測量學的集成分析後，其數據的核心信號^[2]通過疊加得到加強，反映到語言相似度上，新派和老派的語言相似度的相關性為0.96。在疊加後加強信號的基礎上做降維分析，提取新派和老派數據的主要維度，我們發現二者的地理分佈如出一轍。無論是新派還是老派的多維縮放地圖，都能夠反映出有意義的空間分佈模式，而不是雜亂無章、難以解釋的數據維度。此外，我們設置了普通話、中古音（兩家擬音系統）、上古音四套參照係，通過觀察地圖上所有方言到參照係的語言相似度的分佈，我們可以看到，新派和老派單字音的地理分佈也沒有太大的差別。這說明，方言測量學可以很好地抵抗社會變異帶來的干擾，能夠通過集成分析疊加大量字音的空間分佈。儘管不同字音的空間分佈並不相同，但是重複的模式會通過不斷疊加，其信號得到大幅度地增強，從而反映出底層的語言地理結構。

註 釋：

[1] 網址：<http://www.chinalanguages.cn/>

[2] 此處與“信號”(signal)這個概念相對的是“噪音”(noise)。信號是對語言學分析有用的部分，反映數據的底層結構，而噪音需要盡可能地過濾掉。

參考文獻：

黃河 2017 社会地理语言学视野下的微观方言研

究——以宜兴方言为个案，北京大學中文系博士學位論文

黃河 2022 “結構化程度”與“各向漸變性”，《中國語文》(3)

鄭張尚芳 2003 《上古音系》，上海：上海教育出版社

Yang Cathryn & Andy Castro 2008 Representing tone in Levenshtein distance. *International Journal of Humanities and Arts Computing* 2 (1-2) 2008, 205-219

Baxter William H. and Laurent Sagart 2014 *Old Chinese: A New Reconstruction*. Oxford University Press

Burnham D. a& E. Francis 1997 Role of linguistic experience in the perception of Thai tones, In A. S. Abramson, ed., *Southeast Asian linguistic studies in honor of Vichin Panupong*. Bangkok.

Chao Yuenren 1930 A system of tone letters, *Le Maitre Phonetique* 30, 24-27.

Cox Trevor F. & Michael A. A. Cox 2001 *Multidimensional Scaling*. Chapman & Hall/CRC, Boca Raton, FL

Goebel Hans 1982 *Dialektometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*. Österreichische Akademie der Wissenschaften, Philosophisch-Historische Klasse, Denkschriften 157. Wien

Heeringa Wilbert & John Nerbonne 2001 *Dialect areas and dialect continua*, *Language Variation and Change*, Vol 13 (3)

Levenshtein, V. I 1965 Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady SSSR*, 163, pp. 845-848

Labov William 2001 *Principles of Linguistic Change*, Volume 2: Social Factors. Wiley-Blackwell

Nerbonne John & Peter Kleiweg 2003 *Lexical Distance in LAMSAS*, *Language Resources and Evaluation* Vol 37 (3)

Nerbonne J., Kleiweg, P., Heeringa, W., and Manni, F. 2008 Projecting dialect distances to geography: Bootstrap clustering vs. noisy clustering.

In Preisach C., H. Burkhardt, L. Schmidt-Thieme, and R. Decker eds. *Data Analysis, Machine Learning and Applications*, Berlin, Heidelberg: Springer, pp. 647-654

Nerbonne. John 2010 *Mapping Aggregate Variation*. In *Language and Space. An International*

Handbook of Linguistic Variation. Vol 1. Theories and Methods, pages 476 – 495. Mouton de Gruyter, Berlin/ New York

Séguy, J. 1973 *La dialectométrie dans l' Atlas linguistique de la Gascogne*, *Revue de linguistique romane*, 37, p 1–24

黃 河 復旦大學現代語言學研究院 huangheku@163.com

宋華強 安徽師範大學文學院 624013365@qq.com