

## تحلیل تشخیصی خطی فیشر برای داده‌های هواسناسی با استفاده از چارچوب فضا‌های هیلبرت با هسته باز آفرین

بهداد مصطفائی<sup>۱</sup>، محمدرضا فرید روحانی<sup>۲</sup> سید محمد ابراهیم حسینی نسب<sup>۳</sup>

تاریخ دریافت: ۹۹/۸/۳۰

تاریخ پذیرش: ۹۹/۱۲/۲۷

چکیده:

امروزه به موازات رشد روزافزون علم و فناوری، داده‌هایی با ماهیت تابعی به سهولت جمع‌آوری می‌شوند. از این رو تحلیل آماری چنین داده‌هایی از اهمیت ویژه‌ای برخوردار شده است. همانند تحلیل‌های چند متغیره ترکیبات خطی از متغیرهای تصادفی نقشی کلیدی در تحلیل داده‌های تابعی دارند. در این میان نقش نظریه‌ی فضا‌های هیلبرت با هسته‌ی باز آفرین بسیار حائز اهمیت است. در این مقاله، مفهومی عام از تحلیل تشخیصی خطی فیشر که توسط [۴] معرفی شده و تعمیمی از روش چند متغیره کلاسیک برای داده‌های تابعی است، مرور شده است. در این تعمیم یک نگاشت دوسویی که فرایند تصادفی نوع دوم را با فضای هیلبرت با هسته باز آفرین تولید شده به وسیله تابع کوواریانس درون رده‌ای ارتباط می‌دهد، بکار گرفته می‌شود. در نهایت داده‌های هواسناسی کشور در سال ۲۰۰۸ به منظور رده‌بندی اقلیمی مورد تجزیه و تحلیل قرار گرفته‌اند.

**واژه‌های کلیدی:** فضای هیلبرت با هسته باز آفرین، تحلیل تشخیصی خطی، داده‌های تابعی، فرایند تصادفی مرتبه دوم.

### ۱ مقدمه

### ۲ RKHS و فرایندهای تصادفی

در این بخش به معرفی فضای هیلبرت با هسته باز آفرین و ارتباط آن با فرایندهای تصادفی می‌پردازیم. همچنان که در ادامه خواهیم داد هسته‌های همیشه مثبت، فضای هیلبرتی با هسته باز آفرین یکتا تولید می‌کنند و برعکس هسته باز آفرین این فضا نیز یکتاست. بر این اساس چون هسته کوواریانس همیشه مثبت است، فضای هیلبرتی با هسته باز آفرین یکتا تولید می‌کند. این فضا اهمیت ویژه‌ای در این مقاله دارد.

در ادامه با  $RKHS$ <sup>۴</sup> و برخی ویژگی‌های آن آشنا می‌شویم. برای جزئیات بیشتر می‌توان به [۱، ۲، ۳] مراجعه کرد.

**تعریف ۱.۲.** فضای هیلبرت  $\mathcal{H}$  (فضای هیلبرت یک فضای ضرب داخلی کامل است) از توابع تعریف شده روی مجموعه  $T$  با ضرب داخلی  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  را  $RKHS$  با هسته باز آفرین  $K$  گویند هر گاه تابع دومتغیره  $K$  روی  $T \times T$  با خواص زیر موجود باشد:

$$(۱) \text{ برای هر } t \in T \text{ و } f \in \mathcal{H} \text{، } K(\cdot, t) \in \mathcal{H}$$

$$(۲) \text{ برای هر } t \in T \text{ و } f \in \mathcal{H} \text{، } f(t) = \langle f, K(\cdot, t) \rangle_{\mathcal{H}}$$

بسیاری از پدیده‌ها مانند دما و میزان بارندگی، ذاتاً توابع پیوسته‌ای از زمان هستند. ماهیت تابعی این پدیده‌ها موجب می‌شود که این داده‌ها به جای مورد توجه قرار گرفتن در فضا‌های برداری با بعد متناهی، در فضا‌های تابعی در نظر گرفته شوند.

هر چند که با توجه به روش جمع‌آوری این گونه داده‌ها استفاده از روش چندمتغیره کلاسیک برای تحلیل آن‌ها امکان‌پذیر است ولی این تحلیل با محدودیت‌هایی مواجه است. در تحلیل داده‌های تابعی، داده‌ها با در نظر گرفتن ماهیت واقعی آن‌ها که توابعی پیوسته از زمان هستند مورد تحلیل قرار گرفته و اثر زمان در کل بازه زمانی و نه لزوماً در زمان‌های گسسته یا فواصل زمانی برابر، مورد توجه قرار می‌گیرد. بنابراین استفاده از تحلیل داده‌های تابعی در بررسی و تفسیر داده‌های تابعی صحیح‌ترین روش در پیش روی تحلیل گران است.

<sup>۱</sup>گروه آمار، دانشگاه محقق اردبیلی

<sup>۲</sup>گروه آمار، دانشگاه شهید بهشتی (نویسنده مسئول) m.faridrohani@gmail.com

<sup>۳</sup>گروه آمار، دانشگاه شهید بهشتی

مجموعه همه متغیرهای تصادفی به صورت  $\sum_{i=1}^n a_i X_i$  به ازای یک  $n$  ثابت‌های  $a_1, \dots, a_n \in \mathbb{R}$  و نقاط  $t_1, \dots, t_n \in T$  است را در نظر می‌گیریم. این فضا را با  $L_X^1$  نشان می‌دهیم و مشاهده می‌شود که این فضا یک فضای هیلبرت با ضرب داخلی

$$\langle U, V \rangle_{L_X^1} = E(UV) \quad U, V \in L_X^1$$

است. بنابراین  $L_X^1$  شامل ترکیبات خطی با بعد متناهی از متغیرهای تصادفی  $\{X(t) : t \in T\}$  بوده یا حد چنین ترکیبات خطی است. از آنجایی که  $R$  متقارن و همیشه مثبت است،  $RKHS$ ، متناظر با آن  $\mathcal{H}(R)$  وجود دارد که به صورت  $\mathcal{H}(R) = \overline{\text{span}}\{R(\cdot, t) : t \in T\}$  است. اما چون برای هر  $s, t \in T$

$$\langle R(\cdot, s), R(\cdot, t) \rangle_{\mathcal{H}(R)} = E(X(s)X(t)) = \langle X(s), X(t) \rangle_{L_X^1}$$

بنابر قضیه یکجوری پایه ([۲]) یکجوری  $\psi$  از  $\mathcal{H}(R)$  به  $L_X^1$  وجود دارد به طوری که

$$\psi(R(\cdot, t)) = X(t)$$

و  $\mathcal{H}(R)$  و  $L_X^1$  یکجور هستند.

تابع کوواریانس  $X$  نیز  $\mathcal{H}(K)$ ، را به صورت  $\mathcal{H}(K) = \overline{\text{span}}\{K(\cdot, t) : t \in T\}$  ایجاد می‌کند و ممکن است کسی  $\mathcal{H}(K)$  را برای ساخت یک نمایش برای تابع تصادفی  $X$  بکار گیرد. اما فضای هیلبرت  $L_X^1$  ممکن است برای همه مقادیر  $\mu$  یکسان نباشد چرا که ضرب داخلی به  $\mu$  بستگی دارد. این موضوع زمانی که  $\mu$  به  $\mathcal{H}(K)$  تعلق دارد می‌تواند به طور رضایت بخش برطرف شود. [۲] نشان داده است در این حالت که فضای هیلبرت  $L_X^1$  برای هر  $\mu$  یکسان باشد، مجموعه عناصر  $\mathcal{H}(K)$  مساوی مجموعه عناصر  $\mathcal{H}(R)$  است اگرچه دو فضا به نرم‌های متفاوتی مجهز هستند.

### ۳ تحلیل تشخیصی فیشر برای فرایندهای تصادفی

روش‌های تشخیصی برای طبقه‌بندی داده‌ها یکی از روش‌های آماری پرکاربرد در زمینه‌های گوناگون است. با فراهم آمدن امکان جمع‌آوری داده‌هایی با ماهیت تابعی، مسئله رده‌بندی این گونه داده‌ها نیز مورد توجه آمارشناسان قرار گرفته است. هدف این بخش ارائه‌ی ساختار تشخیص خطی است که می‌تواند برای داده‌های نوع عام، که شامل داده‌های تابعی نیز می‌شود، بکار رود.

$RKHS$  از سه مؤلفه اساسی تشکیل شده است: مجموعه اندیس‌گذار  $T$ ، فضای هیلبرت  $\mathcal{H}$  از توابع روی  $T$  و تابع هسته با خاصیت بازآفرینی مشخص شده در (۲). برای یک  $RKHS$  با هسته بازآفرین  $K$  نماد  $\mathcal{H}(K)$  را بکار می‌بریم.

برخی ویژگی‌های جالب توجه  $RKHS$  را در ادامه مرور می‌کنیم.

**گزاره ۲.۲.** برای فضای هیلبرت  $\mathcal{H}$  از توابع تعریف شده روی  $T$  هسته بازآفرین  $K$  وجود دارد اگر و تنها اگر برای هر  $y \in T$  تکامل تابعی خطی  $E_y : \mathcal{H} \rightarrow \mathbb{R}$  که برای هر  $f \in \mathcal{H}$  به صورت  $E_y(f) = f(y)$  تعریف می‌شود، تابعی خطی کران‌دار روی  $\mathcal{H}$  باشد. یعنی

$$\|E_y\| = \sup\{|E_y(f)| / \|f\| : f \neq 0\} < \infty.$$

**گزاره ۳.۲.** (چگال بودن)

$$\mathcal{H}(K) = \overline{\text{span}}\{K(\cdot, t) : t \in T\}.$$

نماد  $\text{span}$  اشاره به مجموعه تمامی ترکیبات خطی از عناصر مجموعه مورد نظر دارد.

**گزاره ۴.۲.** پیوستگی نقطه به نقطه فرض کنید  $\mathcal{H}$ ،  $RKHS$  روی  $T$  باشد و  $\{f_n\}_{n \geq 1} \subseteq \mathcal{H}$  اگر  $\lim_{n \rightarrow \infty} \|f_n - f\|_{\mathcal{H}} = 0$  آنگاه  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  برای هر  $x \in T$

**گزاره ۵.۲.** (همواری) اگر  $K$  تابع دومتغیره پیوسته روی  $T \times T$  باشد آنگاه هر  $f \in \mathcal{H}$  روی  $T$  پیوسته است.

**گزاره ۶.۲.** (یکتابی) اگر  $K$  هسته بازآفرین یک  $RKHS$  باشد آنگاه  $K$  معین نامنفی و یکتاست. برعکس اگر  $K$  تابع دومتغیره همیشه مثبت روی  $T \times T$  باشد،  $RKHS$  یکتا از توابع حقیقی مقدار روی  $T$  وجود دارد به طوری که  $K$  هسته بازآفرین متناظر آن باشد.

$RKHS$  ابزار مهمی در استنباط مرتبط با فرایندهای تصادفی نوع دوم بشمار می‌رود. در حقیقت بین فضای هیلبرت توسعه شده به وسیله یک فرایند تصادفی و  $RKHS$  تولید شده توسط تابع کوواریانس متناظر آن ارتباط وجود دارد که این ارتباط در استنباط این گونه فرایندها بکار گرفته می‌شود.

فرایند تصادفی نوع دوم  $\{X(t) : t \in T\}$   $(E(X(t)^2) < \infty)$  با میانگین  $\mu(t) = E(X(t))$  و تابع کوواریانس  $K(s, t) = \text{cov}(X(s), X(t))$  برای  $s, t \in T$  را در نظر بگیرید. فرض کنید  $R$  تابع گشتاور دوم باشد یعنی  $R(s, t) = E(X(s)X(t))$ . اکنون فرض کنید  $(Q, \mathcal{B}, \nu)$  فضای احتمال متناظر با  $X(\cdot)$  باشد و  $L^2(p)$  مجموعه همه متغیرهای تصادفی روی  $(Q, \mathcal{B}, p)$  که گشتاور دوم متناهی دارند باشد. ما زیرمجموعه‌ای از  $L^2(p)$  که کامل شده

$$E(\psi_W(h)|G = j) = \langle h, \mu_j \rangle_{\mathcal{H}(K_W)}, \quad (7)$$

$$E(\text{var}(\psi_W(h)|G)) = \|h\|_{\mathcal{H}(K_W)}^2. \quad (8)$$

با به کارگیری دوسویی  $\psi_W$  داریم

$$\text{var}_G(E[l|G]) = \text{var}_G(E[\psi_W(h)|G]) \quad (9)$$

$$= \langle h, T_B h \rangle_{\mathcal{H}(K_W)}$$

که در آن برای  $h \in \mathcal{H}(K_W)$

$$(T_B h)(t) = \langle K_B(t, \cdot), h(\cdot) \rangle_{\mathcal{H}(K_W)}. \quad (10)$$

عملگر  $T_B$  دارای تجزیه ی طیفی

$$T_B = \sum_{i=1}^{J-1} \gamma_i \alpha_i \otimes_{\mathcal{H}(K_W)} \alpha_i \quad (11)$$

است که در آن  $\gamma_1 \geq \dots \geq \gamma_{J-1} \geq 0$  و ویژه مقادیر  $T_B$  و  $\alpha_i$ ،  $i = 1, \dots, J-1$  ویژه توابع آن هستند و  $\otimes_{\mathcal{H}(K_W)}$  عملگری است که به صورت

$$(g \otimes_{\mathcal{H}(K_W)} f)h = \langle g, h \rangle_{\mathcal{H}(K_W)} f$$

تعریف می شود. بنابراین با توجه به (۹) و روابط (۲)-(۵) توابع تشخیص خطی به صورت زیر داده می شود

$$l_i = \psi_W(h_i) \quad i = 1, \dots, J-1 \quad (12)$$

که در آن بردارهای ویژه  $T_B$  متناظر با مقادیر ویژه  $\gamma_i$  است. در نتیجه رده بندی با استفاده از فاصله ی توان دوم ماهالانویس بر اساس  $s$  تابع تشخیص خطی اول به دست خواهد آمد. به عبارت دیگر برای جامعه زام رده بندی بر اساس فاصله ی

$$\sum_{k=1}^s (\psi_W(h_k) - \langle h_k, \mu_j \rangle_{\mathcal{H}(K_W)})^2$$

به دست می آید.

## ۴ داده های هواشناسی

با توجه به اینکه دما در هر منطقه تابعی از زمان است و در طول زمان تغییر می کند، این پدیده از ماهیتی تابعی برخوردار است. در این بخش داده های دمای ایستگاه های سینوپتیکی کشور در سال ۲۰۰۸ به منظور رده بندی تابعی نواحی بر اساس دما مورد مطالعه قرار گرفته است. مجموعه ی داده های فوق متشکل از ۱۲ میانگین ماهیانه ی دما در هر ایستگاه است که قبل از تجزیه و تحلیل با استفاده از روش های مناسب هموارسازی به صورت تابعی پیوسته از زمان تبدیل شده اند. از آنجا که بین دمای ثبت شده در ایستگاه های سینوپتیکی همبستگی مکانی وجود دارد، باید ایستگاه هایی در

تحلیل تشخیص خطی فیشر (۰۰۰۰) یکی از ابزارهای مشهور چند متغیره آماری برای مطالعه ی رابطه ی میان یک مجموعه از پیش گوها و یک پاسخ رسته ای است. ۰۰۰۰ همچنین ابزار کاهش بعد را که می تواند برای اهداف رده بندی بکار رود، فراهم می کند. به طور کلی هدف روش فیشر پیدا کردن فضایی با کمترین بعد است به طوری که رده ها (با جوامع) تا حد ممکن تفکیک شوند.

فرض کنیم  $G$  نمایانگر رده ی عضویت فرایند از جامعه های شماره گذاری شده ی ۱ تا  $J$  باشد و  $p(G = j) = \pi_j$  و  $E(X(\cdot)|G = j) = \mu_j(\cdot)$  همچنین فرض کنیم تابع کوواریانس درون رده ای برای هر  $j = 1, \dots, J$  برابر و به صورت

$$K_W(s, t) = E[(X(s) - \mu_j(s))(X(t) - \mu_j(t)) | G = j] \quad s, t \in T$$

باشد.

مطابق شین (۲۰۰۸) توابع تشخیصی را به صورت متغیرهای تصادفی  $l \in L_X^J \setminus \{0\}$  به گونه ای تعریف می کنیم که نسبت تغییر پذیری بین رده ای به تغییر پذیری درون رده ای یعنی

$$\frac{\text{var}(E(l|G))}{E(\text{var}(l|G))} \quad (1)$$

ماکسیمم شود، مشروط بر اینکه این مسئله ی بهینه سازی جواب باشد. بنابراین اولین تابع تشخیص خطی فیشر  $l_1$  به صورت زیر تعریف می شود

$$\text{var}(E(l_1|G)) = \sup_{l \in L_X^J} \text{var}(E(l|G)) \quad (2)$$

مشروط به اینکه

$$E(\text{var}(l|G)) = 1 \quad (3)$$

به طور مشابه برای  $i > 1$   $i$  امین تابع تشخیص خطی فیشر  $l_i$  به صورت

$$\text{var}(E(l_i|G)) = \sup_{l \in L_X^J} \text{var}(E(l|G)) \quad (4)$$

که در آن  $l$  مشروط به (۳) و شرط زیر است، تعریف می شود.

$$E(\text{cov}(l, l_k|G)) = 0 \quad k < j \quad (5)$$

شین (۲۰۰۸) نشان داده است که اگر  $\mu_j \in \mathcal{H}(K_W)$  برای هر  $j = 1, \dots, J$  در این صورت نگاشت خطی یک به یکی از  $\mathcal{H}(K_W)$  به  $L_X^J$  وجود دارد به طوری که برای هر  $t \in T$

$$\psi_W(K_W(\cdot, t)) = X(t)$$

و برای هر  $h \in \mathcal{H}(K_W)$

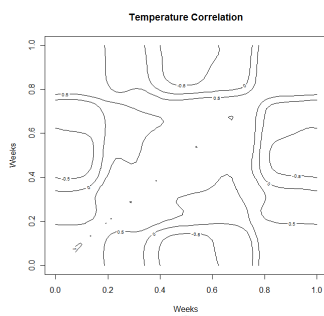
$$E(\psi_W(h)) = \langle h, \mu \rangle_{\mathcal{H}(K_W)}, \quad (6)$$

همچنان که ملاحظه می‌شود میانگین دمای ماهیانه ۵ ناحیه کاملاً از یکدیگر تفکیک شده‌اند. بر این اساس نواحی به ترتیب از سردترین ناحیه (ناحیه ۱) به گرم‌ترین ناحیه (ناحیه ۵) دارای میانگین دمای افزایشی هستند. بعلاوه بیشترین اختلاف میانگین دمای ماهیانه در اوایل بهمن ماه با حدود ۲۰ درجه و کمترین اختلاف میانگین دمای ماهیانه از نیمه تیرماه با حدود ۵/۱۵ درجه بین ناحیه ۱ و ۵ مشاهده می‌شود.

از نقطه نظر انحراف معیار ماهیانه همچنان که در شکل ۲ مشخص است، ناحیه ۲ به طور معنی داری دارای پراکندگی بیشتری نسبت به سایر نواحی است. دلیل اصلی این امر، با توجه به شکل ۱، تفاوت قابل توجه دمای ماهیانه یکی از شهرهای این ناحیه با سایر شهرهای موجود در ناحیه فوق است. این ویژگی باعث می‌شود تا تحلیل تشخیصی در این ناحیه با خطای بالاتری همراه شود. البته با حذف ایستگاه مزبور از این ناحیه پراکندگی دمای ماهیانه سایر نواحی به طور چشم‌گیری کاهش می‌یابد. پس از ناحیه ۲، بیشترین پراکندگی در ناحیه ۵ مشاهده می‌شود که تقریباً در تمامی سال به استثنای فصل بهار میزان پراکندگی آن بالاست.

به طور کلی میانگین دمای ماهیانه ایستگاه‌های موجود در فصل زمستان در تمامی نواحی به استثنای ناحیه ۱، از پراکندگی بیشتری برخوردار است. ایستگاه‌های موجود در ناحیه ۱ در فصل زمستان رفتار همگن‌تری داشته و در نیمه تیرماه پراکندگی بیشتری را نشان می‌دهند.

شکل ۳ خطوط تراز تابع همبستگی دمای ناحیه ۵ را به عنوان نمونه نشان می‌دهد. همچنان که ملاحظه می‌شود در این ناحیه همبستگی هفته‌های فصل زمستان بیشترین همبستگی را با یکدیگر داشته که با گذشت زمان همبستگی این هفته‌ها با هفته‌های منتهی به نیمه خرداد به سطح ۵/۰ کاهش می‌یابد. مجدداً این میزان روند کاهشی خود را تا اواخر مرداد ادامه داده و به ۵/۰- می‌رسد. از آن پس این همبستگی روند افزایشی را می‌پیماید و در اواخر آبان به ۵/۰+ پس از آن مجدداً به ۱- می‌رسد.



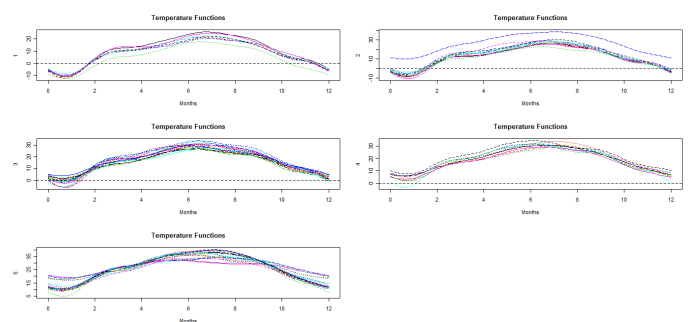
شکل ۳: خطوط تراز تابع همبستگی دما در ناحیه ۵

با استفاده از روش‌های هموارسازی مقدار پارامتر هموارسازی  $U = 1/0.7 \times 10^{-6}$  به دست آمده است. نرخ طبقه‌بندی نادرست ۱۷ از ۶۰ ایستگاه یا ۲۸٪

تحلیل تشخیصی فیشر انتخاب شوند که فاقد این نوع همبستگی باشند. مهم‌ترین عامل در وابستگی مکانی بین ایستگاه‌های یادشده، عرض جغرافیایی و ارتفاع از سطح دریای این ایستگاه‌ها است. از این رو با لحاظ این عوامل از مجموعه ایستگاه‌های سینوپتیکی کشور، ۶۰ ایستگاه جهت انجام تحلیل مورد نظر انتخاب شده‌اند.

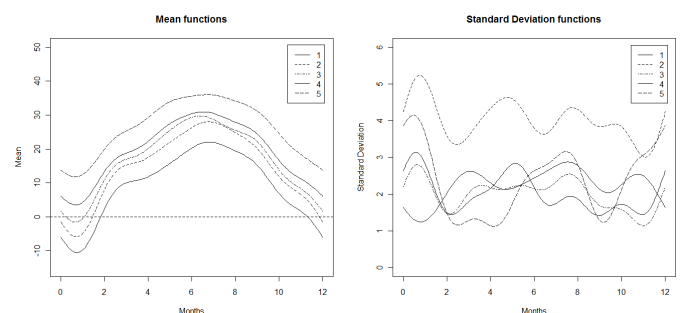
ایران به دلیل گستردگی عرض جغرافیایی و تغییرات قابل توجه در ارتفاعات مناطق مختلف آن دارای تنوع آب‌وهوایی است. صرف نظر از تقسیمات مختلفی که اقلیم‌شناسان در انواع آب‌وهوای حاکم بر کشور داشته‌اند، از نقطه نظر تغییرات دما ایران به ۵ ناحیه تقسیم شده است. بر اساس این تقسیم‌بندی ۷ شهر در ناحیه ۱، ۱۲ شهر در ناحیه ۲، ۱۶ شهر در ناحیه ۳، ۹ شهر در ناحیه ۴ و ۱۶ شهر در ناحیه ۵ قرار دارند (شرکت ملی گاز ایران، ۱۳۸۹).

شکل ۱ منحنی داده‌های دما مربوط به ۶۰ ایستگاه هواشناسی ایران در سال ۲۰۰۸ به تفکیک ۵ ناحیه‌ی دمایی کشور را نشان می‌دهد.



شکل ۱: منحنی داده‌های دما مربوط به ۶۰ ایستگاه هواشناسی ایران در سال ۲۰۰۸ به تفکیک ۵ ناحیه‌ی دمایی کشور

همچنین توابع میانگین و انحراف معیار برای نواحی نیز در شکل ۲ نشان داده شده است.



شکل ۲: توابع میانگین و انحراف معیار داده‌های دما به تفکیک ۵ ناحیه دمایی کشور

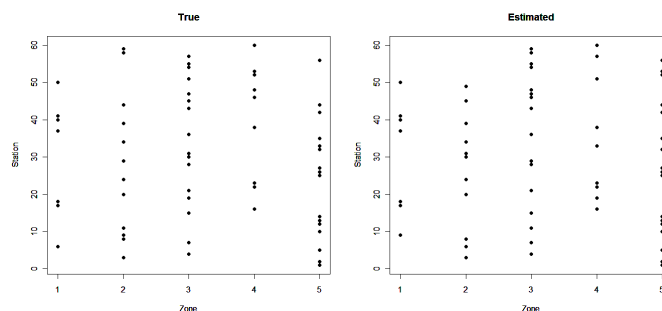
داده شده به هر ناحیه با روش ارائه شده

است. شکل ۴ ایستگاه‌های متعلق به نواحی را در کنار ایستگاه‌های تخصیص داده شده به هر ناحیه با روش ارائه شده، نشان می‌دهد.

## ۵ نتیجه‌گیری

همان‌گونه که در این مقاله تأکید شد،  $RKHS$  ابزار مهمی در استنباط‌های مرتبط با فرایندهای تصادفی نوع دوم بشمار می‌رود. ارتباط بین فرایند تصادفی نوع دوم و  $RKHS$  از طریق نگاشتی دوسویی انجام گرفت. در نتیجه با تعریف صورت مسئله در  $RKHS$  امکان حصول جواب‌های موردنظر میسر گردید.

رویکرد بکار رفته در این مقاله بر پایه‌ی تجزیه‌ی طیفی عملگر خطی تعریف شده روی  $RKHS$  متناظر با فرایند تصادفی نوع دوم است. بر این اساس به‌عنوان کاربردی از روش ارائه شده، داده‌های هواشناسی مورد تجزیه و تحلیل قرار گرفتند که دقت قابل قبولی از این روش را نشان می‌دهد.



شکل ۴: ایستگاه‌های متعلق به نواحی در کنار ایستگاه‌های تخصیص

## مراجع

- [1] Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, **68**(3), 337-404.
- [2] Parzen, E. (1961). An approach to time series analysis. *The Annals of Mathematical Statistics*, **32**(4), 951-989.
- [3] Parzen, E. (1970). Statistical Inference on Time Series by RKHS Methods, (No. TR-14). Stanford Univ. Calif. Dept. of Statistics.
- [4] Shin, H. (2008). An extension of Fisher's discriminant analysis for stochastic processes. *Journal of Multivariate Analysis*, **99**(6), 1191-1216.