

# به کارگیری روش های خوشه بندی میانگین K، میانگین فازی و گوستافسون کسل در تلفیق نتایج وارون سازی داده های توموگرافی لرزه ای انکساری و مقاومت ویژه الکتریکی برای ارزیابی آبرفت و سنگ بستر

بهمن محمدی، ابوالقاسم کامکار روحانی\*؛

دانشگاه صنعتی شاهرود، دانشکده مهندسی معدن، نفت و ژئوفیزیک

پذیرش ۹۶/۱۱/۲۵

دریافت ۹۵/۰۸/۱۱

## چکیده

در سال های اخیر، الگوریتم های خوشه بندی فازی بیش تر از سایر روش ها برای خوشه بندی نتایج به کار برده شده اند. در خوشه بندی، الگوریتم های فازی می توانند در تلفیق مدل های چندگانه با گروه بندی نمونه ها در یک فضای چندبعدی، به کار روند. مفهوم عضوهای جزئی خوشه، ناهمگنی ساختاری همه مدل های ورودی را تلفیق کرده و آن را در یک مفهوم فازی توصیف می کند. در این پژوهش، بعد از وارون سازی داده های مقاومت ویژه الکتریکی با روش کم ترین مربعات گوس- نیوتن و با استفاده از نرم افزار RES2DINV و نیز محاسبه اولین زمان رسیده ها با استفاده از نرم افزار PickWin و وارون سازی داده های توموگرافی لرزه ای انکساری با استفاده از نرم افزار Geotom CG، داده ها با سه روش میانگین K، میانگین فازی و گوستافسون کسل خوشه بندی شدند. با استفاده از روش شاخص دان برای بهینه سازی تعداد خوشه ها عدد ۱۲ به دست آمد که با توجه به نقشه های به دست آمده برای مقاومت ویژه الکتریکی و توموگرافی لرزه ای انکساری، تعداد خوشه مناسبی است. با توجه به بررسی های انجام شده در بستر سد، محدوده آبرفت و سنگ بستر و همچنین لایه بندی، روش خوشه بندی گوستافسون کسل نتیجه بهتری را نشان داده است. با به کار بردن نتایج میانگین فازی در آغاز الگوریتم گوستافسون-کسل، گام های تکرار کاهش و سرعت همگرایی افزایش می یابد. محاسبات مربوط به الگوریتم های خوشه بندی با کدنویسی در محیط نرم افزار متلب انجام شده است

**واژه های کلیدی:** روش میانگین K، روش میانگین فازی، روش گوستافسون کسل، خوشه بندی، توموگرافی

## مقدمه

طی دهه های اخیر چندین روش عملی برای تلفیق توموگرافی مدل های ژئوفیزیکی در یک مدل چند پارامتری مطرح شده است. انواع روش های تلفیق عبارتند از: مدل منطقی بولین، مدل های منطق فازی، وارون سازی توأم مدل های ژئوفیزیکی، روش تلفیق بصری ساده، از قبیل نمودارهای سه تایی<sup>۱</sup> [۱] یا هم پوشانی مدل ها در فرم لایه های نیمه شفاف<sup>۲</sup> [۲]. روش های پایدار و قدرتمند برای تلفیق مدل های بازسازی شده توموگرافی بر اساس تحلیل آماری چند متغیره<sup>۳</sup> وجود دارد. خصوصاً روش های تحلیل خوشه ای، برای مثال الگوریتم های میانگین K<sup>۴</sup>، میانگین فازی<sup>۵</sup>،

kamkarr@yahoo.com

\*نویسنده مسئول

1. ternary plots
2. semi-transparent layers
3. Multivariate statistical analysis
4. K-Means
5. Fuzzy C-Means

[۳]، [۴]، [۵]، و گوستافسون کسل [۶] برای تلفیق مدل‌های توموگرافی درون‌چاهی بسیار مفید هستند. الگوریتم خوشه‌ای، نمونه‌ها را در یک فضای چندبعدی تحلیل کرده و به زیر مجموعه‌های از پیش تعریف شده تقسیم‌بندی می‌کند. سپس این الگوریتم، مدل‌های زیرسطحی چند پارامتری را بر اساس مدل‌های ورودی چندگانه با روابط نامعلوم و غیرخطی متغیر از نظر فضایی، ایجاد می‌کند [۷]. روابط پیچیده بین پارامترهای فیزیکی مختلف، به طور نوعی وابسته به محیط‌های رسوبی سطحی تحکیم نیافته است [۸]. به هر حال قبل از مشخص کردن یک مدل توموگرافی در تحلیل خوشه‌ای، کیفیت مدل باید به دقت ارزیابی شود. خطاهای ساختاری مدل، مثل خطاهای مصنوعی، ممکن است روی نتایج خوشه‌بندی تأثیر بگذارد [۹].

اپستین و دوگرتی<sup>۱</sup> (۱۹۹۶) [۱۰] از خوشه‌بندی میانگین K برای منطقه‌بندی سفره آبخوان و تخمین مقادیر قابلیت‌گذرده‌ی<sup>۲</sup>، استفاده کردند. فیچنر و دیتریچ<sup>۳</sup> (۱۹۹۷) [۱۱]، دیتریچ و همکاران<sup>۴</sup> (۱۹۹۸) [۱۲]، ترونیکه و همکاران<sup>۵</sup> (۲۰۰۲) [۱۳]، و ترونیکه و همکاران (۲۰۰۴) [۱۴]، از تحلیل خوشه‌ای میانگین K در تلفیق مدل‌های مختلف ژئوفیزیکی مقطعی در بهبود نمایش آبخوان‌های رسوبی و در نتیجه اکتشاف آن‌ها، استفاده کردند. بروچلر و همکاران<sup>۶</sup> (۲۰۱۲) [۱۵]، خوشه‌بندی میانگین K را برای تلفیق نتایج توموگرافی لرزه‌ای و هیدرولیکی (ضریب پخش و ذخیره ویژه) برای بهبود توصیف هیدرولوژی آبخوان رسوبی، به کار گرفتند. پاشه و همکاران<sup>۷</sup> (۲۰۰۶) یک رهیافت را بر اساس تحلیل خوشه‌ای میانگین فازی برای تلفیق مدل‌های مختلف توموگرافی درون‌چاهی به منظور منطقه‌بندی آبخوان، توسعه دادند [۷]. گانتر و همکاران<sup>۸</sup> (۲۰۰۶) از آنالیز خوشه‌ای میانگین فازی در تلفیق مقاومت ویژه الکتریکی دو بعدی و مدل‌های سرعت موج P استفاده کردند [۱۶]. پاشه و همکاران (۲۰۰۷) تحلیل خوشه‌ای میانگین فازی را برای اکتشاف مسیر جریان بر اساس مدل سه‌بعدی مقاومت ویژه، توموگرافی‌های عرضی و هدایت هیدرولیکی، به کار بردند [۱۷]. به هر حال در سال‌های اخیر برخی تلاش‌ها در جهت تلفیق مدل‌های توموگرافی سطحی با استفاده از الگوریتم خوشه‌ای بوده است.

خوشه‌بندی از ابزارهای متداول داده‌کاوی است که به استخراج دسته‌هایی با حداکثر شباهت بین عناصر داخل دسته و حداقل شباهت با عناصر سایر دسته‌ها می‌پردازد. خوشه‌بندی فازی به کشف مدل‌های فازی از داده‌ها می‌پردازد. یکی از اولین روش‌های خوشه‌بندی فازی که بر مبنای تابع هدف و استفاده از فاصله اقلیدسی بنا شده بود در سال ۱۹۷۴ دان ارائه کرد و سپس به وسیله بزدک تعمیم داده شد [۱۸]، [۱۹]، [۲۰]. پس از آن یانگ یک بررسی اجمالی روی روش‌های خوشه‌بندی فازی انجام داد [۲۱]. سپس گوستافسون و کسل در سال ۱۹۷۹ الگوریتم خوشه‌بندی فازی با استفاده از ماتریس کوواریانس فازی را ارائه کردند [۲۲]. کاربرد روی کرد امکان در خوشه‌بندی فازی نخستین بار توسط کیم و کریشناپورام در سال ۱۹۹۳ صورت گرفت [۲۳]. پس از آن نیز اصلاحات و بهبودهای بسیاری روی الگوریتم‌های ارائه شده صورت گرفته است.

1. Eppstein and Dougherty
2. Transmissivity
3. Fechner and Dietrich
4. Dietrich et al
5. Tronick et al
6. Brauchler et al
7. Paasche et al
8. Günther et al

## مواد و روش‌ها

### روش خوشه‌بندی میانگین K

این الگوریتم پارامتر  $k$  را به عنوان ورودی گرفته و مجموعه  $n$  شیء را به  $k$  خوشه افراز می‌کند. به طوری که سطح شباهت داخلی خوشه‌ها زیاد است و سطح شباهت اشیاء بیرون خوشه‌ها کم باشد. شباهت هر خوشه نسبت به متوسط اشیاء آن خوشه سنجیده شده که این متوسط مرکز خوشه نامیده می‌شود. این الگوریتم بدین صورت کار می‌کند:

ورودی:  $k$ ، تعداد خوشه‌ها است. پایگاه داده  $X$  شامل  $n$  شیء،

$$X = \{x_1, x_2, \dots, x_n\}, x_i \in R^m \quad (1)$$

خروجی: یک مجموعه از  $k$  خوشه که معیار مربع خطا را حداقل می‌کند.

مراحل اجرای الگوریتم خوشه‌بندی میانگین K بدین صورت است:

۱. به صورت تصادفی،  $k$  داده دلخواه را به عنوان مراکز خوشه‌های ابتدایی انتخاب می‌کنیم
۲. هر داده را با توجه به نزدیکی آن به مراکز خوشه‌ها، به خوشه‌ها تخصیص می‌دهیم. معیار نزدیکی را فاصله اقلیدوسی در نظر می‌گیریم:

$$d(x_i, m_j) = \|x_i - m_j\| = \sqrt{\sum_{i=1}^m (x_{ik} - m_{jk})^2} \quad (2)$$

که در آن  $x_i$ ،  $i$ امین داده ورودی و  $m_j$  مرکز (میانگین) خوشه  $j$ ام است. داده مذکور در خوشه‌ای قرار می‌گیرد که کم‌ترین فاصله را با مرکز آن خوشه داشته باشد. مرکز خوشه نیز میانگین حسابی داده‌های آن خوشه است.

۱. مراکز خوشه‌ها را به روز می‌کنیم یعنی برای هر خوشه میانگین اعضای خوشه را به دست می‌آوریم.
۲. با توجه به مراکز جدید خوشه‌ها به مرحله ۲ برمی‌گردیم و فرآیند فوق را تا جایی ادامه می‌دهیم که هیچ تغییری در خوشه‌ها رخ ندهد (در این حالت الگوریتم پایان یافته است) [۲۴].

در واقع هدف مینیمم کردن رابطه (۳) است،

برای مراکز خوشه‌ها  $\{\mu_j, j = 1, \dots, k\}$  داریم [۲۵]:

$$k_j = \min \sum_{j=1}^k S_j \quad (3)$$

که  $k_j$  خوشه  $j$ ام و مجموع مربعات برای خوشه  $j$  به صورت رابطه (۴) است [۲۵]:

$$S_j = \sum_{i=1}^n z_{ji} |x_i - \mu_j|^2 \quad (4)$$

که در آن  $z_{ji} = 1$  است، اگر در گروه  $j$  باشد و در غیر این صورت برابر صفر است. پارامتر  $\mu_j$  برابر میانگین گروه  $j$  است [۲۵]:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^n z_{ji} x_i \quad (5)$$

از مزایای این الگوریتم سادگی و همگرایی سریع در رسیدن به جواب نهایی است. از معایب آن نیز این است چون در ابتدا تعداد خوشه‌ها به الگوریتم داده می‌شود بهینه بودن جواب به دست آمده را نمی‌توان تضمین کرد، چون که

تعداد خوشه‌ها به صورت دستی است و اشتباه در تعیین این عدد در جواب نهایی تأثیر زیادی دارد. کارایی الگوریتم نیز بسیار وابسته به نحوه تعیین مراکز اولیه برای خوشه‌هاست [۲۴].

### خوشه‌بندی میانگین فازی

یکی از مهم‌ترین و پرکاربردترین الگوریتم‌های خوشه‌بندی، الگوریتم میانگین فازی است. در این الگوریتم نمونه‌ها به  $C$  خوشه تقسیم می‌شوند و تعداد  $C$  از قبل مشخص شده است. در نسخه فازی این الگوریتم نیز تعداد خوشه‌ها ( $C$ ) از قبل مشخص شده است (در فرمول‌های زیر تعداد خوشه‌ها  $g$  در نظر گرفته شده است). الگوریتم میانگین فازی برای پارامترهای  $y_{ji}$  ( $i = 1, \dots, n; j = 1, \dots, g$ ) به صورت رابطه (۶) قابل حل است (تابع هدف) [۲۵]:

$$J_r = \sum_{i=1}^n \sum_{j=1}^g y_{ji}^r |x_i - m_j|^2 \quad (6)$$

که تحت شرایط زیر مینیمم می‌شود (از روی  $y_{ji}$  می‌توان یک ماتریس  $y$  تعریف کرد که دارای  $n$  سطر و  $g$  ستون است و مؤلفه‌های آن هر مقداری بین صفر تا یک را می‌توانند اختیار کنند. اگر تمامی مؤلفه‌های ماتریس  $y$  به صورت صفر و یا یک باشند الگوریتم مشابه میانگین فازی کلاسیک خواهد بود. با این که مؤلفه‌های ماتریس  $y$  می‌توانند هر مقداری بین صفر تا یک را اختیار کنند اما مجموع مؤلفه‌های هر یک از ستون‌ها باید برابر یک باشد) [۲۵]:

$$\sum_{j=1}^g y_{ji} = 1 \quad 1 \leq i \leq n$$

$$y_{ji} \geq 0 \quad i = 1, \dots, n; \quad j = 1, \dots, g \quad (7)$$

معنای این شرط این است که مجموع تعلق هر نمونه به  $C$  خوشه باید برابر یک باشد. پارامتر  $y_{ji}$  درجه تابع عضویت  $i$ امین شیء با  $j$ امین خوشه را نشان می‌دهد. پارامتر  $r$  یک عدد حقیقی است که فازی بودن را کنترل می‌کند ( $r \geq 1$ ). اگر  $r=1$  قرار دهیم تابع هدف خوشه‌بندی میانگین غیر فازی به دست می‌آید. پارامتر  $x_i$  نمونه  $i$ ام و پارامتر  $m_j$  مرکز  $j$ امین خوشه است و  $n$  تعداد نمونه‌ها (اشیاء) است.

برای به دست آوردن فرمول‌های مربوط به  $y_{ji}$  و  $m_j$  باید تابع هدف تعریف شده را مینیمم کنیم. با استفاده از شرط فوق و برابر صفر قرار دادن مشتق تابع هدف داریم [۲۵]:

$$y_{ji} = \frac{1}{\sum_{k=1}^g \left( \frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{r-1}}} \quad (8)$$

$$m_j = \frac{\sum_{i=1}^n y_{ji}^r x_i}{\sum_{i=1}^n y_{ji}^r} \quad (9)$$

با استفاده از دو فرمول محاسبه شده الگوریتم خوشه‌بندی میانگین فازی و مراحل اجرای آن بدین صورت است:

۱. انتخاب  $r$  ( $1 < r < \infty$ )، مقدار دهی اولیه برای مقادیر تابع عضویت  $y_{ji}$  ( $i = 1, \dots, n; j = 1, \dots, g$ )
۲. محاسبه مراکز خوشه‌ها (محاسبه  $m_j$ ها).

۳. محاسبه  $d_{ij} = |x_i - m_j|$  که  $d_{ij}, i = 1, \dots, n; j = 1, \dots, g$

۴. محاسبه تابع عضویت: اگر  $d_{il} = 0$  باشد برای مقداری از  $l$  آن‌گاه  $y_{li} = 1$  است و  $y_{ji} = 0$  است برای همه  $j, j \neq l$ ، در غیر این صورت با توجه به رابطه (۸) حل می‌شود.

۵. اگر همگرا نبود، به مرحله ۲ برگرد.

یکی از شرط‌های توقف الگوریتم زمانی است که تغییرات نسبی در مقادیر مرکز کوچک می‌شود [۲۵]:

$$D_z = \left\{ \sum_{j=1}^g |m_j(k) - m_j(k-1)|^2 \right\}^{\frac{1}{2}} < \varepsilon \quad (10)$$

که  $m_j(k)$  مقدار ژامین مرکز روی  $k$ امین تکرار و  $\varepsilon$  حد مشخص شده است. شرط‌های توقف دیگر بر اساس مقادیر تابع عضویت،  $y_{ji}$  یا تابع هزینه،  $J_r$  است. شرط دیگر بر اساس بهینه‌سازی محلی تابع هزینه که سلیم و اسماعیل<sup>۱</sup> (۱۹۸۶) [۲۶] مطرح کرده‌اند. در اجرای این الگوریتم زمانی توقف داریم که:

$$\max_{1 \leq i \leq n} \alpha_i < \varepsilon \quad (11)$$

که

$$\alpha_i = \max_{1 \leq j \leq g} y_{ji}^{r-1} |x_i - m_j|^{r-1} - \min_{1 \leq j \leq g} y_{ji}^{r-1} |x_i - m_j|^{r-1} \quad (12)$$

تا این که در یک مینیمم محلی  $\alpha_i = 0, i = 1, \dots, n$  در این حالت همگرایی حاصل می‌شود.

### الگوریتم گوستافسون کسل

الگوریتم گوستافسون کسل فازی تعمیم الگوریتم میانگین فازی با تغییری از فاصله اقلیدوسی به صورت رابطه (۱۳)

است:

$$d_{GK}^2 = (x_k - v_i)^T A_i (x_k - v_i) \quad (13)$$

که  $A_i$  با استفاده از وارون ماتریس کوواریانس فازی برای هر خوشه محاسبه می‌شود.

الگوریتم گوستافسون کسل فازی با معیار رابطه (۱۴) مینیمم می‌شود:

$$Q = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m d_{GK}^2 = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^m (x_k - v_i)^T A_i (x_k - v_i) \quad (14)$$

بهینه‌سازی این معیار تابع محدودیت‌های استاندارد (۱۵) است:

$$\forall i, k \quad u_{ik} \in [0, 1], \quad \forall i \quad 0 < \sum_{k=1}^N u_{ik} < N, \text{ and } \forall k \quad \sum_{i=1}^c u_{ik} = 1. \quad (15)$$

ماتریس  $A_i$  بر اساس وارون ماتریس کوواریانس فازی  $C_i$  محاسبه می‌شود:

$$A_i = (\rho_i |C_i|)^{1/d} C_i^{-1}, \quad C_i = \frac{\sum_{k=1}^N u_{ik}^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^N u_{ik}^m} \quad (16)$$

بدیهی است که  $C_i$  باید وارون پذیر باشد. در غیر این صورت الگوریتم هیچ جوابی را نمی دهد [۲۷]. با جایگزین کردن فاصله اقلیدسی با ماهولونوبیس (که توسط یک ماتریس متقارن، معین و مثبت ایجاد شده) در الگوریتم میانگین فازی می توان به شناسایی خوشه های بیضوی نیز دست یافت. این پیشنهاد را گوستافسون و کسل در سال ۱۹۷۹ [۲۲] ارائه کردند.

### وارون سازی توموگرافی

مدل اولیه مربوط به ساختار بررسی شده، با استفاده از شواهد زمین شناسی و داده های ژئوفیزیک به دست می آید و زمان سیر محاسبه شده و مشاهده شده با یکدیگر مقایسه می شوند. هدف از وارون سازی به حداقل رساندن اختلاف زمان سیر حاصل از مدل اولیه و زمان سیر مربوط به داده های واقعی (برداشت شده) طی یک فرایند تکراری است [۲۸]، [۲۹]. زمان سیر از رابطه (۱۷) به دست می آید:

$$t_i = \int_{l_i(z)} s \cdot dl_i(z) \quad (17)$$

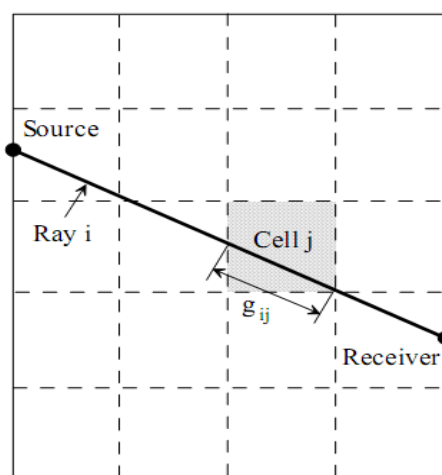
که در آن  $t_i$  زمان سیر مربوط به پرتو  $i$ ام،  $s$  مقدار کندی یا عکس سرعت و  $l$  طول پرتو  $i$ ام است که با عمق  $z$  تغییر می کند.

### توموگرافی اولین زمان رسید

توموگرافی اولین زمان رسید با استفاده از اولین زمان رسیدهای ثبت شده انجام می شود. در این روش مدل اولیه به تعدادی سلول تقسیم می شود و در نهایت حل مسئله وارون با استفاده از یک الگوریتم مانند الگوریتم بازسازی تکرار هم زمان یا  $SIRT^1$  و با استفاده از رابطه (۱۸) انجام می شود:

$$dt_i = M \cdot ds \quad (18)$$

روش توموگرافی لرزه ای قادر به تصویر سازی از ساختمان داخلی زمین بر اساس سرعت سیر امواج است که می تواند به صورت گالری به گالری، گمانه به گمانه و سطح زمین به گمانه، باشد. شکل ۱ نمایش شماتیک از یک مقطع توموگرافی شامل چشمه و گیرنده امواج لرزه ای، سلول بندی مقطع و پرتو لرزه ای است.



شکل ۱: نمایش شماتیک از یک مقطع توموگرافی شامل چشمه و گیرنده امواج لرزه ای، سلول بندی مقطع و پرتو لرزه ای - ای [۳۰]. محور افقی و عمودی، بسته به نوع برداشت می تواند  $xy$  و  $xz$  یا  $yz$  انتخاب شود

1. Simultaneous Iterative Reconstruction Technique

در این آزمایش زمان سیر موج و مختصات مکانی چشمه‌ها و گیرنده‌های امواج به‌عنوان داده‌های اولیه در نظر گرفته می‌شود و مقطع تصویر شده نحوه توزیع سرعت را مشخص می‌کند. به‌طور بسیار ساده شده روش محاسباتی مبتنی بر حل معادله ماتریسی (۱۹) است [۳۰]:

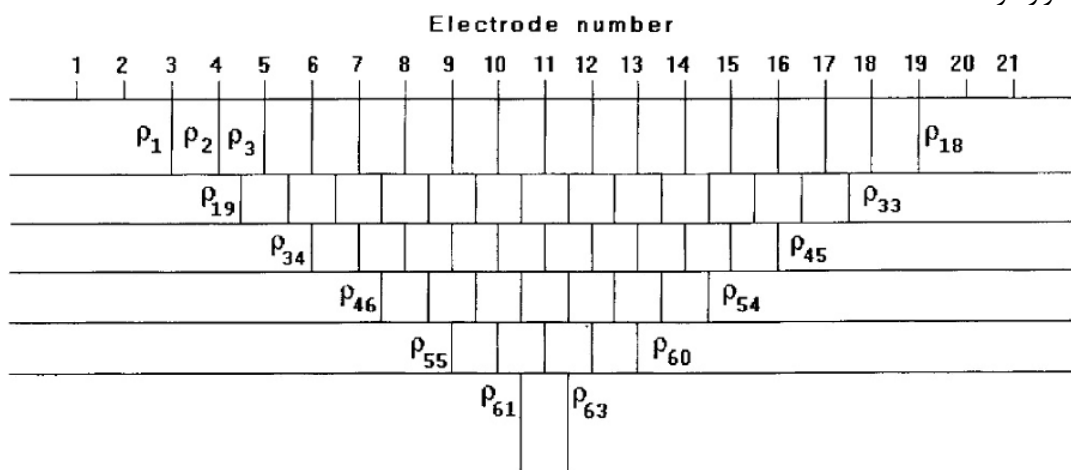
$$t_j = \sum_{i=1,2,\dots,n} a_{ij} \cdot x_i \quad j = 1, 2, \dots, m \quad (19)$$

که در آن  $a_{ij}$  عبارت از طول  $j$  امین پرتو در  $i$  امین سلول،  $x_i$  عکس سرعت موج ( $1/V$ ) در  $i$  امین سلول و  $t_j$  زمان سیر  $j$  امین پرتو است. برای حل این معادله ماتریسی چندین روش وجود دارد که بهترین روش استفاده از الگوریتم تکرار است.

### مدل‌سازی وارون مقاومت ویژه

در روش وارون‌سازی بر خلاف مدل‌سازی مستقیم یا پیش‌رو با استفاده از داده‌های به‌دست آمده، مدل زمین تخمین زده می‌شود. مراحل معمول در فرآیند وارون‌سازی شامل قرائت اولیه داده‌های مقاومت ویژه، تصحیح پارامترهای وارون‌سازی در صورت لزوم و انجام وارون‌سازی است. در مدل‌سازی قبل از هر چیز به مدل فرضی اولیه برای انجام مدل‌سازی نیاز است [۳۱]. روش مدل‌سازی وارون معمولاً از یک تکنیک بهینه‌سازی غیرخطی برای بهبود بخشیدن به یک مدل ساده اولیه استفاده می‌کنند [۳۲].

در مدل‌سازی دوبعدی (2D) داده‌های مقاومت ویژه از نرم‌افزار RES2DINV استفاده می‌شود. در این نرم‌افزار مدل استفاده شده به‌عنوان مدل اولیه شامل یک سری از بلوک‌های مستطیلی دوبعدی نشان داده شده است (شکل ۲). در این مدل تعداد بلوک‌ها با تعداد نقاط داده‌ها در شبه مقطع مقاومت ویژه ظاهری برابر است و به شیوه‌ای مشابه مرتب شده‌اند. همچنین عمق مراکز بلوک‌ها در عمق میانگین تحت بررسی برای بازه‌های الکترونی مختلف استفاده شده قرار گرفته‌اند [۳۱].



شکل ۲. ترتیب بلوک‌های مستطیلی استفاده شده در مدل‌سازی سه بعدی [۳۱]

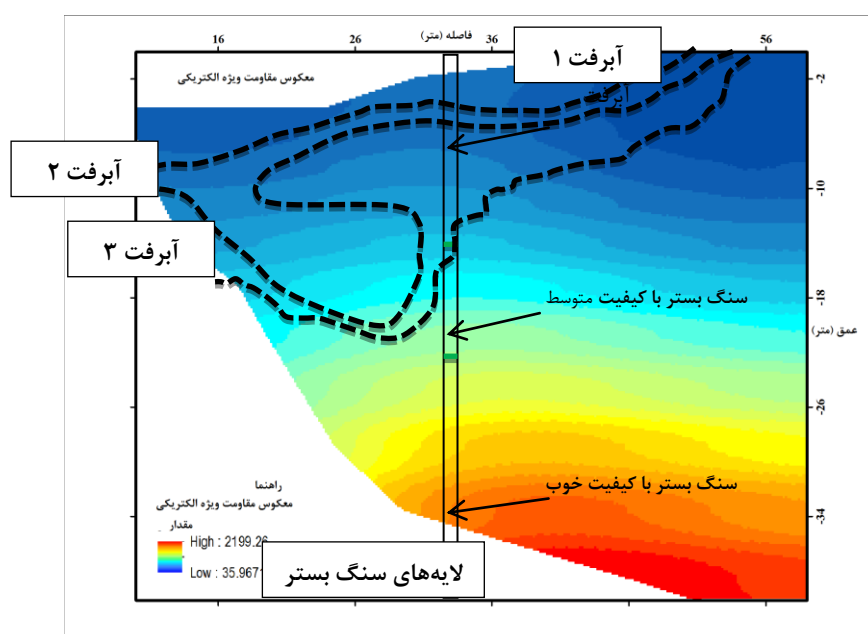
پس از تخمین مدل اولیه نوبت به تصحیح این مدل با هدف رسیدن به انطباق قابل قبول با داده‌های صحرایی می‌رسد. برای این منظور از دو روش کم‌ترین مربعات گوس-نیوتن<sup>۱</sup> و روش شبه نیوتن<sup>۲</sup> استفاده می‌شود.

1. Gauss-Newton  
2. Quasi-Newton

## تحلیل نتایج

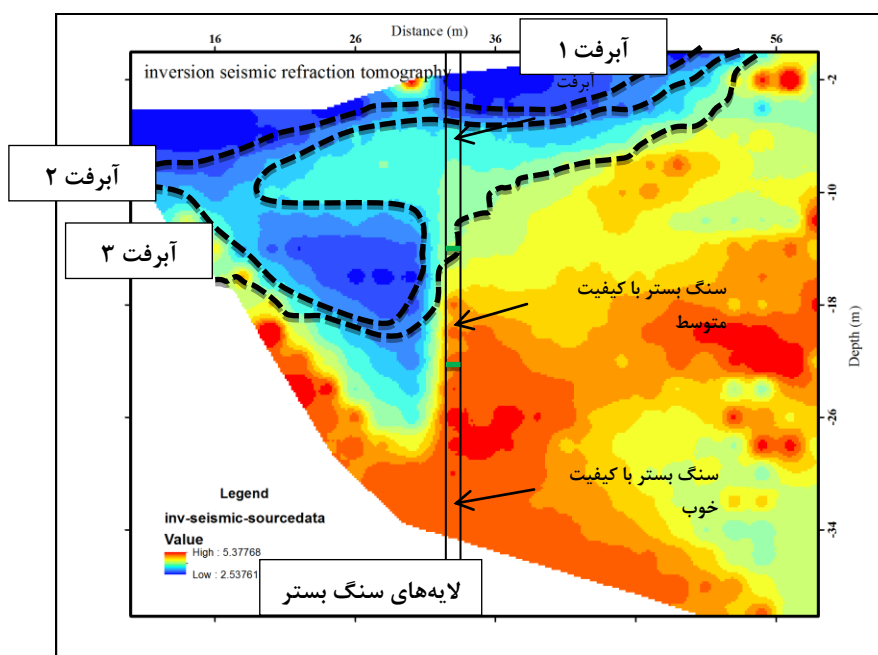
## وارون سازی داده های ورودی

داده های توموگرافی لرزه ای و مقاومت ویژه الکتریکی از ساختگاه یکی از پروژه های سد کوهرنگ در حال اجرا، در ایران برداشت شده است. با توجه به نتایج حاصل از وارون سازی داده های مقاومت ویژه الکتریکی لایه های زیرسطحی مطابق شکل ۳ به دست آمد. وارون سازی با روش کمترین مربعات گوس- نیوتن و با استفاده از نرم افزار RES2DINV محاسبه شده است. برای تلفیق دو نقشه حاصل از وارون سازی داده های مقاومت ویژه و توموگرافی لرزه انکساری، داده ها از عمق ۱۴ متر به پایین و زیر سطح ایستابی در نظر گرفته شدند. عمق صفر در نقشه برابر عمق ۱۴ متر در واقعیت است. لایه های به دست آمده نشان می دهد که مقاومت ویژه به طرف عمق به تدریج افزایش می یابد.



شکل ۳. نتایج وارون سازی مقاومت ویژه الکتریکی برای اعماق پایین تر از ۱۴ متری سطح زمین

روش توموگرافی لرزه ای انکساری از توانایی چشم گیری در تشخیص ویژگی های لرزه شناسی لایه های زمین دارد. ابتدا با استفاده از نرم افزار PickWin اولین زمان رسیده ها به دست آمده و بعد با استفاده از نرم افزار Geotom CG وارون سازی داده های توموگرافی لرزه ای انکساری انجام شد. با توجه به نتایج به دست آمده از این پژوهش، لایه آبرفتی و هوازده سرعتی معادل ۳ کیلومتر بر ثانیه و سنگ بستر سرعتی بالای ۴ کیلومتر بر ثانیه دارد. با کیفیت ترین قسمت های توده سنگ با سرعت های عمده بین ۴ تا بیش از ۴/۵ کیلومتر متر بر ثانیه هستند. قسمت های با سرعت پایین، نواحی با کیفیت کم و یا آبرفتی هستند (شکل ۴).



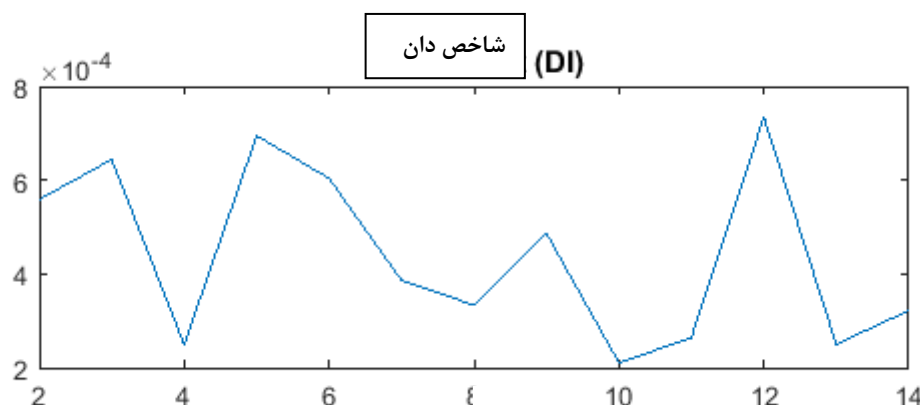
شکل ۴. نتایج وارون‌سازی توموگرافی لرزه‌ای انکساری برای اعماق پایین‌تر از ۱۴ متری سطح زمین

#### خوشه‌بندی داده‌ها

با استفاده از روش شاخص دان<sup>۱</sup> برای بهینه‌سازی تعداد خوشه‌ها عدد ۱۲ به‌دست آمد (شکل ۵) که با توجه به نقشه‌های به‌دست آمده برای مقاومت ویژه الکتریکی و توموگرافی لرزه‌ای انکساری (شکل‌های ۳ و ۴)، تعداد خوشه مناسبی است. روش‌های مختلفی برای بهینه‌سازی تعداد خوشه‌ها نظیر شاخص دان، دیویس-بولدین<sup>۲</sup>، بال-هال<sup>۳</sup>، کالینسکی-هاراباسز<sup>۴</sup> وجود دارد. شاخص دان بر پایه ایده‌ای که درون خوشه‌ها به هم نزدیک‌تر و خوشه‌ها از هم جدا باشند، بیان شده است، بنابراین این شاخص، انتخاب و به‌کار برده شد. با به‌کار بردن فرمول‌های هر یک از الگوریتم‌های خوشه‌بندی (در بخش مواد و روش‌ها آمده است) در محیط متلب<sup>۵</sup>، کد محاسباتی برای آن‌ها نوشته شده و نقشه مربوطه ترسیم شد (شکل‌های ۶، ۷ و ۸). در مقایسه با میانگین فازی و میانگین K در روش گوستافسون کسل هر خوشه علاوه بر مرکز خوشه با یک ماتریس متقارن، معین و مثبت مشخص می‌شود. این ماتریس برای هر خوشه یک مقیاس ایجاد می‌کند. باید این موضوع را هم در نظر گرفت که با انتخاب دلخواه و اختیاری ماتریس‌ها، فاصله‌ها می‌توانند به‌طور دلخواه کوچک شوند. برای اجتناب از حداقل سازی تابع هدف با ماتریس‌های با ورودی‌های تقریباً صفر، نیاز به مقداری ثابت برای خوشه‌ها با ماتریسی با دترمینان یک داریم. که در این صورت شکل خوشه‌ها متغیر و اندازه‌هایشان ثابت است. اما گوستافسون و کسل شکل‌ها متفاوت برای خوشه‌ها را نیز مقدور ساختند. بدین ترتیب که یک مقدار ثابت e برای هر ماتریس A معرفی کردند و در حالت کلی باید  $\det(A)=e$  باشد [۲۲]. اگرچه انتخاب ثابت‌ها نیز نیاز به یک دانش پیشین در مورد خوشه‌ها دارد. شکل‌های بیضوی حاصل از روش خوشه‌بندی گوستافسون

1. Dunn index
2. Davies Bouldin
3. Ball Hall
4. Calinski Harabasz
5. Matlab

کسل در مورد مثال‌های یکسان با میانگین فازی که شکل‌های کروی ارائه دهد، خوشه‌بندی بهتری ارائه دهد. البته موضوع



شکل ۵. شاخص دان که برای محاسبه تعداد خوشه‌های بهینه استفاده شده است. در این شاخص حداکثر مقدار به‌دست آمده برای خوشه‌ها، به‌عنوان تعداد خوشه بهینه در نظر گرفته می‌شود. محور افقی تعداد خوشه‌ها و محور عمودی مقدار شاخص را نشان می‌دهد

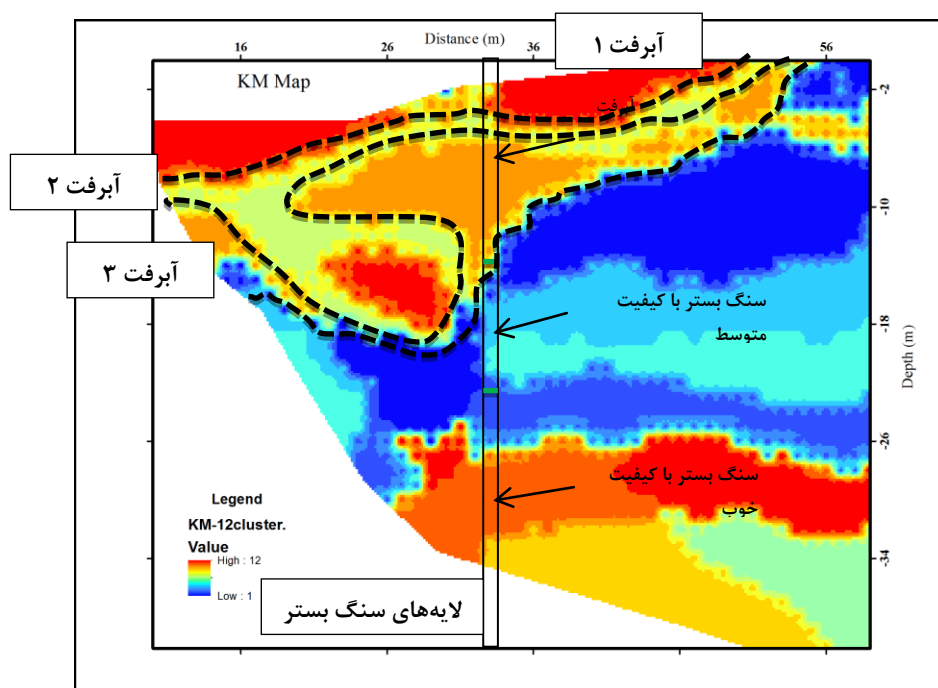
تجمع نقاط در مراکز خوشه‌ها در مورد خوشه‌بندی امکانی با گوستافسون کسل نیز به‌صورت مسئله باقی می‌ماند. اگر داده‌ها با رویکرد امکان خوشه‌بندی شوند، فاکتورهای توسعه برای خوشه‌ها برای تشخیص شکل‌ها در تصاویر استفاده می‌شوند. موقعیت و جهت را از مرکز خوشه و ماتریس می‌توان به‌دست آورد. کیفیت نتیجه حاصل از این روش به‌شدت وابسته به داده‌های در دسترس است. در مقایسه با میانگین فازی هزینه‌های محاسباتی روش خوشه‌بندی گوستافسون کسل بیشتر است. یک روش پیشنهادی مؤثر در کاهش گام‌های تکرار و افزایش سرعت همگرایی، آغاز الگوریتم گوستافسون کسل با نتایج حاصل از یک اجرای میانگین فازی است که با پیاده کردن آن نتایج مطلوبی حاصل شد. با توجه به نتایج حفاری و تحلیل مغزه‌ها سه قسمت آبرفت، سنگ بستر با کیفیت متوسط و سنگ بستر با کیفیت خوب مشخص شد. با تطابق این نتایج با خوشه‌بندی‌های مختلف، این نتیجه حاصل شد که روش خوشه‌بندی گوستافسون کسل فازی نتیجه بهتری را نشان می‌دهد و لایه‌های مختلف را به‌خوبی از هم تفکیک کرده است (شکل ۸). در شکل ۶ که نتایج خوشه‌بندی میانگین K را نشان می‌دهد، خوشه‌های مختلف زمین را به‌خوبی از هم تمایز نمی‌دهد به‌طوری‌که شروع و انتهای لایه، انطباق مناسبی با مرز خوشه‌بندی ایجاد نکرده است. در شکل ۷ که نتایج خوشه‌بندی میانگین فازی را نشان می‌دهد نیز لایه‌ها با مرز خوشه‌بندی هماهنگی مناسبی ندارد اما در شکل ۸ که نتایج خوشه‌بندی گوستافسون کسل است، لایه‌ها با مرز خوشه‌بندی سازگاری تقریباً خوبی را نشان می‌دهد.

### بحث و تحلیل عدم قطعیت

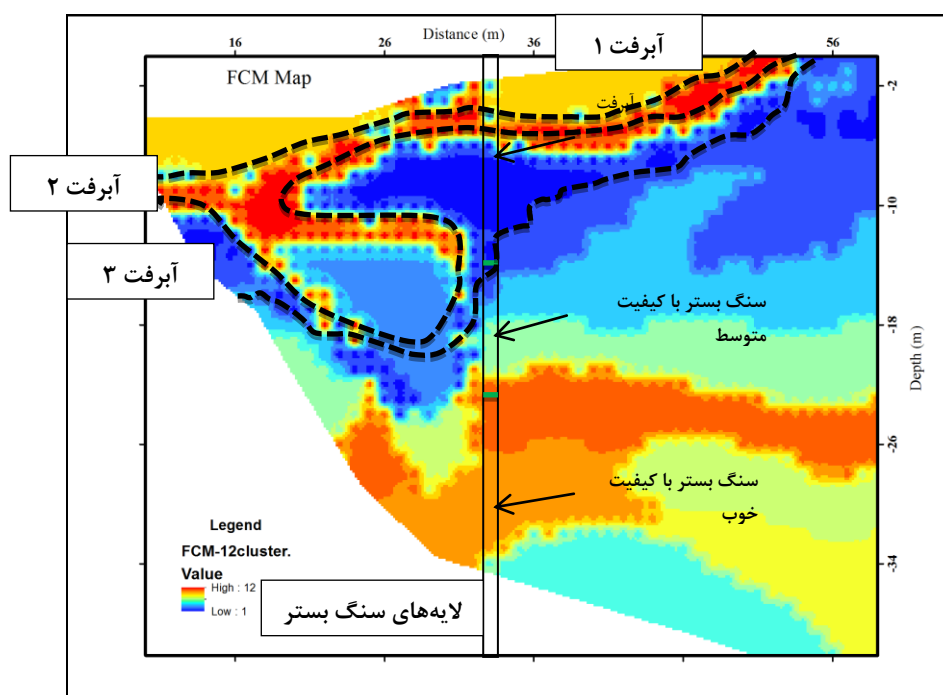
عدم قطعیت، یک پارامتر با نتیجه اندازه‌گیری مرتبط است و پراکندگی مقادیری را مشخص می‌کند که می‌توانند به شکل منطقی به اندازه کمیت اندازه‌گیری شده نسبت داده شود. خطا متفاوت از عدم قطعیت است. خطا به شکل اختلاف بین یک نتیجه تک و مقدار واقعی تعریف می‌شود. در عمل خطای اندازه‌گیری مشاهده شده، اختلاف بین مقدار مشاهده شده و مقدار مرجع است. بدین ترتیب خطا چه تئوری، چه مشاهده شده، یک مقدار تک است. در حالی که عدم قطعیت شکل یک گستره یا بازه را به خود می‌گیرد. عموماً نمی‌توان از عدم قطعیت برای تصحیح یک نتیجه

اندازه‌گیری استفاده کرد، در حالی که مقدار خطای معین می‌تواند به‌عنوان یک تصحیح برای نتیجه اندازه‌گیری به کار گرفته شود. پارامترهای عدم قطعیت در این پژوهش می‌تواند زمین‌شناسی و برداشت و همچنین معکوس‌سازی داده‌ها باشد. در بحث برداشت ژئوفیزیکی و زمین‌شناسی می‌توان گفت اگر عملیات برداشت داده چندین بار تکرار شود، مؤلفه نوفه‌های تصادفی موجود در داده‌ها را می‌توان مستقیماً تعیین کرد. اما این حالت به‌ندرت در ژئوفیزیک اتفاق می‌افتد. اصولاً تخمین عدم قطعیت داده‌ها با تحلیل خطای عملیات شروع می‌شود، مانند ثبت نوفه زمینه [۳۳]، اما تحلیل خطا برای هر عملیات ژئوفیزیکی کاملاً پیچیده و شامل تعداد زیادی مجهولات است. هدف از معکوس‌سازی داده‌های ژئوفیزیکی، استنباط کمی اطلاعات درباره زمین از تعداد محدودی مشاهده غیرمستقیم و حاوی نوفه است. در عمل معمولاً با بهینه کردن یک تابع اختلاف تعریف شده برای داده‌ها از عدم قطعیت آن‌ها صرف نظر می‌شود. اما این سؤال پیش خواهد آمد که تابع هدف چه موقع بهینه می‌شود.

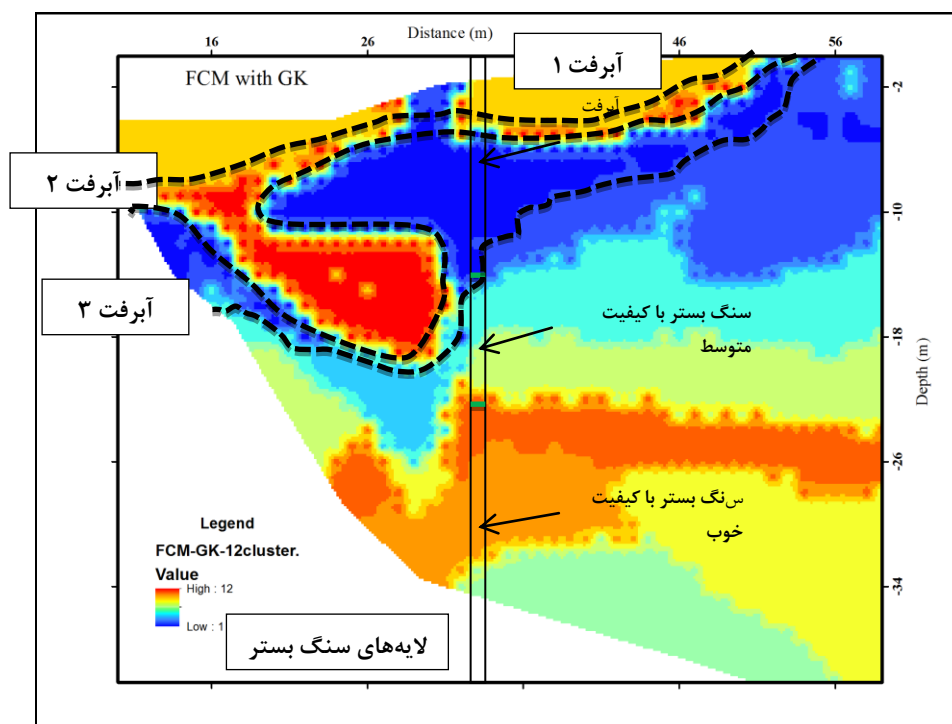
برای تخمین عدم قطعیت داده‌ها و ساخت بازه‌های اطمینان برای پارامترهای مدل در مسائل معکوس خطی ابتدا با تنظیم تیخونف و استفاده از روش منحنی L برای تعیین پارامتر تنظیم [۳۴]، پارامترهای مدل حاصل می‌شوند، سپس بردار باقی‌مانده حاصل از این مدل‌ها برای تخمین عدم قطعیت داده‌ها استفاده می‌شود، این تخمین خطا برای ساخت مجموعه‌ای از مدل‌ها که داده‌ها را پیش‌بینی می‌کنند و همچنین مدل‌هایی که محدودیت‌هایی در بازه پارامترهای مدل دارند، به کار می‌رود. هنگامی که یک مدل معقول به‌دست آمد، از اطلاعات اولیه برای ساخت بازه‌های اطمینان مدل نهایی استفاده می‌شود. روش‌های ارائه شده برای تخمین عدم قطعیت داده‌ها، دقت زیادی دارند.



شکل ۶. مدل منطقه‌بندی به‌دست آمده با الگوریتم میانگین K از مدل‌های معکوس‌سازی شده شکل‌های ۳ و ۴. زون‌ها (طبقه‌ها) با رنگ مشخص شده‌اند



شکل ۷. مدل منطقه بندی به دست آمده با الگوریتم میانگین فازی از مدل های معکوس سازی شده شکل های ۳ و ۴  
زون ها (طبقه ها) با رنگ مشخص شده اند



شکل ۸. مدل منطقه بندی به دست آمده با الگوریتم گوستافسون کسل فازی از مدل های معکوس سازی شده  
شکل های ۳ و ۴. زون ها (طبقه ها) با رنگ مشخص شده اند

### اعتبار سنجی داده‌های حاصل از معکوس‌سازی

بر اساس تحلیل، در عمق ۲۹ متری سرعت حاصل از معکوس‌سازی برابر با ۴,۸۲ کیلومتر بر ثانیه و سرعت آزمایشگاهی بر اساس گزارش مکانیک سنگی شرکت سهامی آب منطقه‌ای اصفهان (طرح سوم سد کوهرنگ، خدمات مهندسی مرحله دوم) [۳۵]، برابر با ۶/۸۵ کیلومتر بر ثانیه است. پس خطای تخمین سرعت معادل ۲۹/۶۴ درصد است.

$$\frac{4.82}{6.85} \times 100 = 70.36\%$$

### نتیجه‌گیری

خوشه‌بندی فازی از تلفیق روی کرد فازی در بحث خوشه‌بندی برای کاربردی‌تر کردن آن و انطباق بیشتر با دنیای واقعی حاصل شده است. با توجه به نتایج مشاهده می‌شود که استفاده هم‌زمان از چندین مجموعه داده نتایج دقیق‌تری را در تفسیر ارائه می‌دهد. با وارون‌سازی داده‌ها، کیفیت مواد آبرفتی و سنگ بستر ارزیابی شد و روش‌های خوشه‌بندی نیز این مرزها را مورد تحلیل قرار داد. با توجه به نتایج حفاری و تحلیل مغزه‌ها و انطباق این نتایج با خوشه‌بندی‌های مختلف، این نتیجه حاصل شد که روش خوشه‌بندی گوستافسون کسل فازی نتیجه بهتری را نشان می‌دهد و لایه‌های مختلف را به خوبی از هم تفکیک می‌کند. با اجرای الگوریتم گوستافسون کسل با نتایج حاصل از میانگین فازی، گام‌های تکرار کاهش و سرعت هم‌گرایی افزایش یافت. با توجه به شکل ۸، لایه‌بندی آبرفت، سنگ بستر با کیفیت متوسط و سنگ بستر با کیفیت خوب با مرز خوشه‌بندی گوستافسون کسل، سازگاری خوبی را نشان می‌دهد. در دو الگوریتم دیگر لایه‌ها با دقت پایین‌تری از هم تفکیک شده‌اند.

### منابع

1. Gallardo L. A., "Multiple cross-gradient joint inversion for geospectral imaging", *Geophysical Research Letters*, 34 (2007) 1-5.
2. Rumpel H., Binot F., Gabriel G., Hinsby K., Siemon B., Steuer A., Wiederhold H., Cuxhavener rinne, Kirsch, R., Rumpel, H., Scheer, W., and Wiederhold, H., eds., "Groundwater resources in buried valleys- a challenge to geosciences", BurVal Working Group, Leibniz Institute for Applied Geosciences (2006).314.
3. Sun J., Li Y., "Joint inversion of multiple geophysical data using guided fuzzy C-Means clustering", *GEOPHYSICS*, 81 (2016) 37-57.
4. Kaufmann L., Rousseeuw P. J., "Finding groups in data: an introduction to cluster analysis", John Wiley and Sons, 47 (1990) 788.
5. Höppner F., Klawonn F., Kruse R., Runkler T., "Fuzzy cluster analysis: Methods for classification, data analysis and image recognition", John Wiley and Sons, 51 (1999) 769-770.

6. Paasche H., Eberle D., "Automated compilation of pseudo-lithology maps from geophysical data sets: a comparison of Gustafson-Kessel and fuzzy C-means cluster algorithms", *Exploration Geophysics*, 42 (2011) 275-285.
7. Paasche H., Tronicke J., Holliger K., Green A. G., Maurer H., "Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on fuzzy C-means cluster analyses", *Geophysics*, 71 (2006) 33-44.
8. Schön J. H., "physical properties of rocks: Fundamentals and principles of petrophysics", Pergamon Press (1998) 583.
9. Hachmoler B., Paasche H., "Integration of surface-based tomographic models for zonation and multimodel guided extrapolation of sparsely known petrophysical parameters", *GEOPHYSICS*, 78 (2013) 43-53.
10. Eppstein M. J., Dougherty D. E., "Simultaneous estimation of transmissivity values and zonation", *Water Resources Research*, 32 (1996) 3321-3336.
11. Fechner T., Dietrich P., "Lithological inversion of tomographic data: Proceedings of the 3rd Conference on Environmental and Engineering Geophysics Society (EEGS), Expanded Abstracts (1997) 355-358.
12. Dietrich P., Fechner T., Whittacker J., Teutsch G., "An integrated hydrogeophysical approach to subsurface characterization", M. Herbert, and K. Kovar, eds., *Groundwater quality: Remediation and protection*, International association of hydrological sciences, 250 (1998) 513-519.
13. Tronicke J., Dietrich P., Wahlig U., Appel E., "Integrating surface georadar and crosshole radar tomography: A validation experiment in braided stream deposits", *Geophysics*, 67 (2002) 1516-1523.
14. Tronicke J., Holliger K., Barrash W., Knoll M. D., "Multivariate analysis of crosshole georadar velocity and attenuation tomograms for aquifer zonation", *Water Resources Research*, 40 (2004) 1-14.
15. Brauchler R., Doetsch J., Dietrich P., Sauter M., "Derivation of site-specific relationships between hydraulic parameters and P-wave velocities based on hydraulic and seismic tomography", *Water Resources Research*, 48 (2012) 1-14.
16. Günther T., Bentley L. R., Hirsch M., "A new joint inversion algorithm applied to the interpretation of DC resistivity and refraction data", *XVI International Conference on Computational Methods in Water Resources*, Expanded Abstracts, (2006) 1-8.

17. Paasche H., Günther T., Tronicke J., Green A. G., Maurer H. R., Holliger K., "Integrating multi-scale geophysical data for the 3D characterization of an alluvial aquifer", EAGE near surface 2007 -13<sup>th</sup> Conference on Environmental and Engineering Geophysics, Expanded Abstracts, (2007) 11.
18. Dunn J. C., "A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters", Journal. Cybernetics. 3 (1974) 95-104.
19. Bezdek J. C., "Fuzzy mathematics in pattern classification", Ph. D. Thesis, Applied math. Center, Cornell University, Ithaca, (1973) 147.
20. Hathaway R., Bezdek J., "Nerf C-Means: Non-Euclidean Relational Fuzzy clustering", Pattern Recognition, 27 (1994) 429-437.
21. Yang X., Zhao W., Chen Y., Fang X., "Image segmentation with a Fuzzy clustering algorithm based on Ant-Tree", Signal Processing, 88 (2008) 2453- 2462.
22. Gustafson E. E., Kessel W. C., "Fuzzy Clustering With A Fuzzy covariance Matrix", Proceedings of The IEEE Conference on Decision and Control, San Diego, California, (1979) 761-766.
23. Krishnapuram R., Keller J., "A Possibilistic Approach to Clustering", IEEE Transactions on Fuzzy Systems, 1(1993) 98-110.
۲۴. هداوندی. الف، "طراحی یک سیستم خبره فازی بر پایه یک مدل ترکیبی هوش مصنوعی برای مدل‌سازی پیش‌بینی فروش"، پایان نامه کارشناسی ارشد، دانشگاه صنعتی شریف، دانشکده مهندسی صنایع (۱۳۸۸) ۱۵۴.
25. Andrew R. Webb., QinetiQ Ltd., Malvern, UK., "Statistical Pattern Recognition", second edition, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, (2002) 496.
26. Selim S. Z., Ismail M. A., "On the local optimality of the fuzzy isodata clustering algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, 8 (1986) 284-288.
27. Graves D., Pedrycz W., "Fuzzy C-Means, Gustafson-Kessel FCM, and Kernel-Based FCM: a comparative study", Analysis and Design of Intelligent Systems Using Soft Computing Techniques, 41 (2007) 140-149.
28. Zhou H., "Determination of velocities and interfaces by multi-scale tomography", 67th Annual International Meeting: Society of Exploration Geophysicists, Expanded Abstracts, (1997).1877-1880.
29. Zhou H., "Multi-scale travelttime tomography", Geophysics, 68 (2003).1639-1649.

30. Lehmann B., "Seismic travel time tomography for engineering and exploration applications", European Association of Geoscientists and Engineers Publications BV (2007) 14.
31. Loke M. H., Barker R. D., "Rapid least-square inversion of apparent resistivity pseudosections using a quasi-Newton method", *Geophysical Prospecting*, 44 (1996) 131-152
32. Loke M. H., "Tutorial: 2-D and 3-D electrical imaging surveys", *Geotomo Software, Res2dinv 3.5 Software* (2004) 128.
33. Gouveia W., Scales J. A., "Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis", *Journal of Geophysical Research*, 103 (1998) 2759-2779.
34. Tikhonov A. N., Arsenin V. Y., "Solutions of Ill-posed Problems", Winston, John Wiley and Sons, New York (1977) 258.
۳۵. شرکت سهامی آب منطقه ای اصفهان، "گزارش مکانیک سنگی شرکت سهامی آب منطقه ای اصفهان"، طرح سوم سد کوهرنگ، خدمات مهندسی مرحله دوم، شماره گزارش، ۱۵۴۸۲، (۱۳۸۷) ۸۱.