



بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

شیوه نامه کنترل کیفیت داده ها

ا پژوهشگاه ملی اقیانوس شناسی و علوم جوی
مرکز ملی داده های اقیانوسی و دریایی

فروردین ۱۳۹۷

مقدمه:

روزانه در دنیا حجم زیادی داده‌های دریایی توسط ابزارها و تجهیزات مختلف، تولید می‌شود که این داده‌ها یکی از نیازهای اساسی در کار محققین دریایی و تصمیم‌گیری سازمان‌های دریایی می‌باشد. این در حالی است که در هنگام تولید و انتقال داده به مراکز داده، داده‌ها دارای نقص، ابهام و عیب هستند. یکی از اجزای اصلی در فرایند مدیریت داده، کنترل کیفیت داده می‌باشد. جهت رسیدن به داده با کیفیت، باید پایش مستمر بر جمع‌آوری داده، تلاش برای تصحیح مشکلات ایجاد شده و جلوگیری از توسعه داده‌های غیر صحیح صورت گیرد.

کیفیت داده‌ها اهمیت بسیار زیادی دارد چرا که تعیین‌کننده قابل استفاده بودن داده‌ها و همچنین تعیین‌کننده کیفیت تصمیمات گرفته شده بر مبنای آنها است. در موارد بسیاری داده‌هایی که حتی استاندارد نیز شده‌اند ممکن است نادرست، ناقص و یا مبهم باشند. کیفیت داده‌ها سال‌هاست که به علت اهمیت آن در تصمیم‌گیری و تبادل داده‌ها مورد توجه است. به طور کلی دو روش کنترل کیفیت علمی و خودکار برای داده‌های اقیانوسی در نظر گرفته می‌شود. در روش اول نیاز به متخصص آن رشته و علم را می‌طلبد و در روش دوم می‌توان الگوریتم‌های موجود را به ابزار نرم‌افزاری تبدیل نمود که در مرکز داده‌ها با توجه به تخصص افراد موجود در این مرکز، تمرکز بر روی روش دوم بوده است. در حال حاضر روش و استاندارد واحدی برای تمام انواع داده‌های اقیانوسی وجود ندارد ولیکن استانداردهای مختلفی برای برخی از ابزارها و انواع داده، توسط چند برنامه بین‌المللی مانند برنامه بین‌المللی تبادل داده و اطلاعات اقیانوسی (IODE¹)، مجمع بین‌المللی اکتشاف دریا (ICES²)، برنامه جهانی نیم رخ دما و شوری (GTSP³) و همچنین اتحادیه زیرساخت اروپایی مدیریت داده اقیانوسی و دریایی (SeaDataNet) تعریف شده است. لازم به ذکر است بخشی از عملیات کنترل کیفیت، مربوط به تولیدکننده داده و بخشی از آن از وظایف مراکز داده می‌باشد. در این شیوه نامه، دو فرایند برای کنترل کیفیت داده‌ها در نظر گرفته شده است: ۱- فرایند اول می‌بایست توسط تولیدکننده داده انجام شود. مرکز داده‌ها، استانداردهای علمی موجود در مورد کنترل کیفیت داده‌های اقیانوسی و دستورالعمل‌های کیفی سازی داده‌های حاصل از ابزارهای مختلف برداشت داده را توسط برنامه کاربردی فراداده‌ها که توسط مرکز داده‌ها، توسعه داده شده است در اختیار تولیدکنندگان داده قرار داده می‌دهد و تولیدکننده داده متناسب با نیاز خود آنها را انتخاب می‌کند. ۲- فرایند دوم می‌بایست توسط مرکز داده‌ها انجام شود. این مرکز ابزارهای نرم‌افزاری کنترل کیفیت را مطابق با الگوریتم‌های برنامه GTSP، برای کیفی سازی خودکار داده‌ها با بررسی صحت آنها از جمله تاریخ، زمان، مختصات جغرافیایی، عمق یا فشار، بررسی داده‌های گمشده یا پرت و غیره توسعه داده است. این مرکز، این الگوریتم‌ها را بر روی ستونهای مجموعه داده‌های مرکز داده‌ها به اجرا در می‌آورد و ستونها مطابق با نتیجه اجرای الگوریتم‌ها، نشانه گذاری⁴ می‌شوند.

جهت تسهیل در به‌کارگیری و اعمال موارد مطرح شده، مرکز ملی داده‌ها اقدام به تهیه این شیوه نامه در سه ماده نموده است.

¹ International Oceanographic Data and Information Exchange

² International Council for the Exploration of the Sea

³ Global Temperature and Salinity Profile Programme

⁴ Flag

شیوه نامه کنترل کیفیت داده ها

ماده یک: تعاریف

تضمین کیفیت^۵: رویه هایی است که قبل از استقرار دستگاه انجام می شود برای اینکه به بهترین داده با کیفیت ممکن دسترسی داشته باشیم.

کنترل کیفیت^۶: رویه هایی است که روی داده هایی که از دستگاه خارج شده انجام می شود. رویه های کنترل کیفیت داده می بایست توسط دو گروه از افراد به شرح زیر انجام شوند:

- روش های بکارگیری شده توسط تولید کننده داده به منظور بهبود سازگاری داده در یک مجموعه داده
- روش های بکارگیری شده توسط مدیر داده روی یک مجموعه داده چند منبعی

تولیدکننده داده مسئولیت های زیر را بعهده دارد:

- بکارگیری روش ها و تجهیزات اندازه گیری استاندارد پیشنهاد شده بین المللی
- روش ها و ابزار ملی و بین المللی واسنجی
- اعتبارسنجی داده مطابق با نتایج واسنجی و مقایسه با روش های استاندارد
- تست کردن محدوده های ثابت و محاسبه شده، شکاف ها و مقادیر ثابت
- کشف، اصلاح و نشانه گذاری داده های پرت
- کشف، اصلاح و نشانه گذاری خطاها در مکان و زمان
- مستندسازی فرایند نمونه برداری و اعتبارسنجی داده ها شامل الگوریتم های بکارگیری شده
- مستندسازی بررسی های کنترل کیفیت و نتایج آن

مدیر داده مسئولیت های زیر را بعهده دارد:

- تست کردن فرمت کدگذاری
- بررسی مجموعه داده های ورودی در مقابل مکان و خطاهای شناسایی شده
- تست های محدوده های ثابت و محاسبه شده
- تست های مرتبط با استانداردهای اقلیم شناسی مانند لویتوس، اشویل
- بررسی بصری
- چک کردن تکرارها
- پایش پارامترها
- ارزیابی اقیانوس شناسی و هواشناسی

⁵ Quality Assurance

⁶ Quality Control

ماده ۱:

الزامات زیر می‌بایست توسط تولیدکنندگان داده در فایل‌های فراداده برای کنترل کیفیت در نظر گرفته شود:
الف- برنامه کاربردی تولید فراداده، از وبسایت مرکز داده‌ها به آدرس زیر دانلود شود. کاربر بایستی فایل دانلود شده را از حالت فشرده خارج کرده و سپس فایل اجرایی MetadataTool.exe را اجرا نماید. توصیه می‌شود قبل از دانلود و اجرای برنامه کاربردی، به موارد مطرح شده در همین آدرس توجه شود.

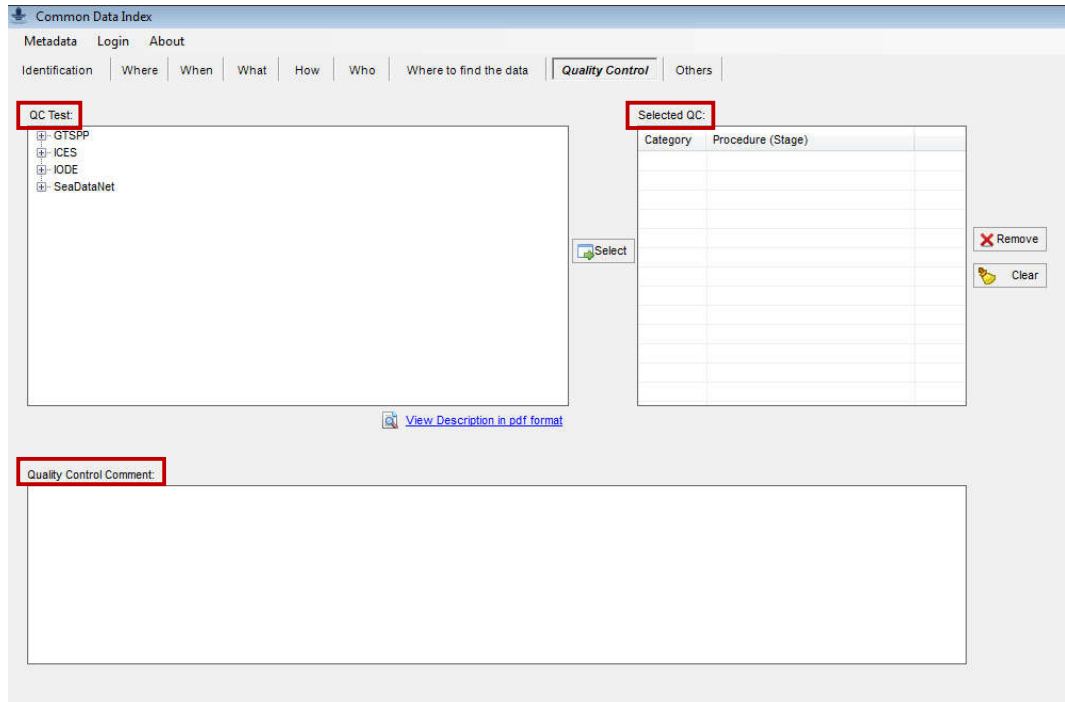
<http://incod.inio.ac.ir/Software/Metadata-Tool>

ب- در صورتی که در رابطه با داده‌های جمع‌آوری شده پروسه (های) کنترل کیفیت خاصی انجام شده باشد، در این قسمت اطلاعات مربوط به آن وارد می‌شود. این Tab از دو قسمت تشکیل شده است که در ادامه توضیح داده شده است. شکل ۱ آیتم‌های مختلف آن را نشان می‌دهد.

QC Test: در این قسمت، مستندات مربوط به کنترل کیفیت که توسط برنامه‌های مرجع و عمده در حوزه اقیانوس‌شناسی مانند IODE، ICES، GTSP و SeaDataNet تولید شده‌اند، برای سهولت کار، در اختیار کاربران قرار گرفته شده است. کاربر از میان لیستی که در اختیار دارد، روند کنترل کیفیت استاندارد مورد نظر خود را انتخاب می‌نماید. در صورت ابهام، کاربر می‌تواند با انتخاب یک روند کنترل کیفیت و سپس کلیک بر روی View Description in PDF Format توضیحی را در آن رابطه مشاهده نماید. برای انتخاب یک رویه خاص، روی گزینه Select کلیک می‌کنیم. سمت راست این قسمت، شامل لیستی از فرایندهای انتخاب شده کنترل کیفیت است که کاربر می‌تواند آن‌ها را پاک (با کلیک روی Clear) یا گزینه‌ای را حذف نماید (با کلیک بر روی Remove).

ج- Quality Control Comment: اگر روش کنترل کیفیتی که متخصص انجام داده و در بند ب موجود نباشد، تکنیک‌های کالیبراسیون و پردازش داده همچنین جزئیات تجهیزات و سنسورها در قسمت توضیح^۷ کنترل کیفیت، ذکر شود.

⁷ Comment



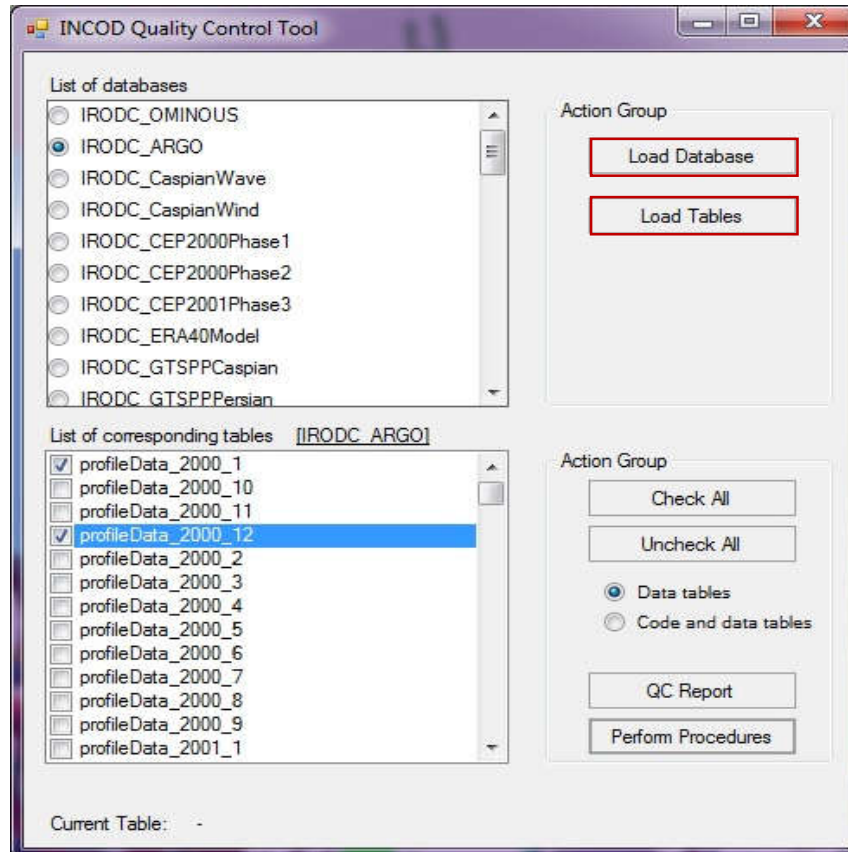
The screenshot shows the 'Common Data Index' web application. The 'Quality Control' tab is active. On the left, under 'QC Test:', there is a list of test categories: GTSP, ICES, IODE, and SeaDataNet. A 'Select' button is located between the list and the 'Selected QC' table. The 'Selected QC' table has two columns: 'Category' and 'Procedure (Stage)'. To the right of the table are 'Remove' and 'Clear' buttons. At the bottom, there is a 'Quality Control Comment:' text area. A 'View Description in pdf format' link is also visible.

شکل ۱: موارد مرتبط با کنترل کیفیت اعمال شده روی مجموعه داده در قسمت QualityControl

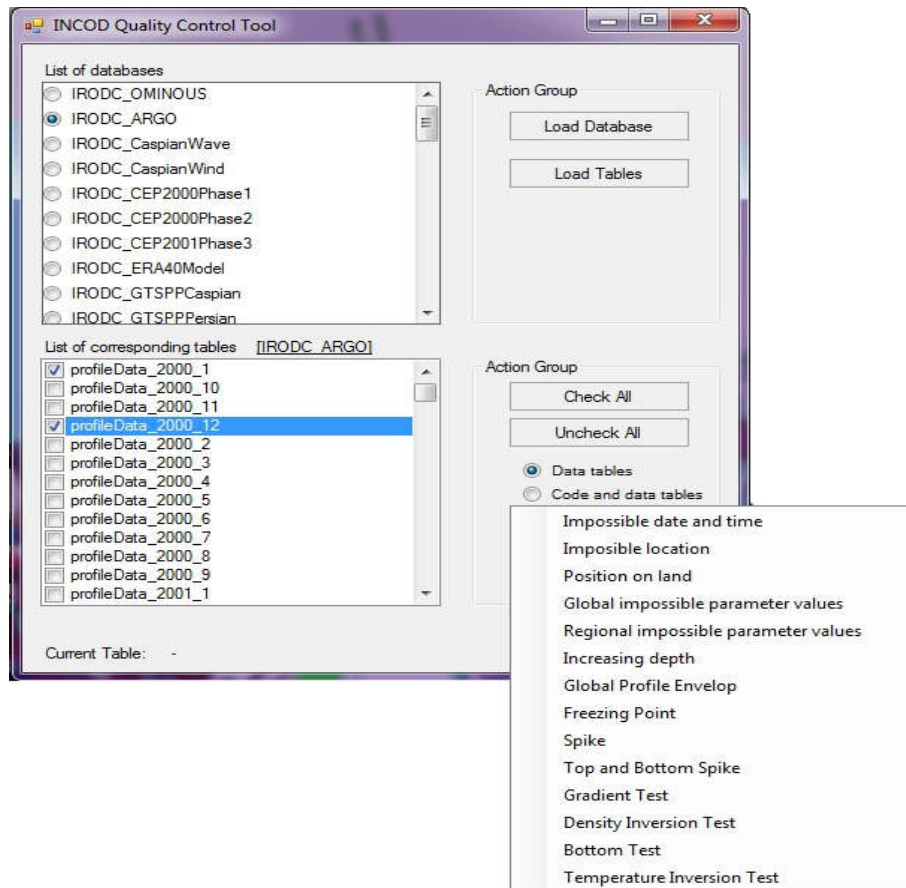
د- لازم بذکر است، برنامه کاربردی اشاره شده در ماده ۳-الف در تاریخ ۹۵/۸/۴ توسط مرکز داده ها به اعضای محترم هیئت علمی و کارشناسان پژوهشی، آموزش داده شده است.

ماده ۲:

الزامات زیر می بایست توسط مرکز داده ها برای کنترل کیفیت داده ها در نظر گرفته شود.::
 مرکز داده ها اقدام به توسعه ابزارهای نرم افزاری مطابق با الگوریتم های برنامه GTSP نموده است. این مرکز می بایست پس از دریافت داده، متناسب با نوع داده، الگوریتم های فوق را روی فایل ها اجرا نموده و ستونها را نشانه گذاری نماید.
 برای اجرای الگوریتم های GTSP، ابزاری مطابق با فلوجارتهای آن توسعه داده شده است که برای سهولت اجرا از طریق رابط کاربری در اختیار پرسنل مرکز داده ها قرار می گیرد. ابتدا کاربر با انتخاب گزینه Load Database مطابق شکل ۲ می تواند لیستی از پایگاه های داده موجود روی سرور را مشاهده نماید. با انتخاب هر کدام از آنها و با کلیک بر روی گزینه Load tables همه جداول موجود در پایگاه داده انتخابی، قابل مشاهده است. برای شروع کار برنامه، ابتدا کاربر باید یک یا چند جدول را انتخاب نماید. با کلیک بر روی Perform Procedures مطابق شکل ۳ لیستی از الگوریتم های کنترل کیفیت قابل دسترس خواهد بود که کاربر با کلیک روی هر کدام، می تواند آن را روی جدول یا جداول انتخابی، اعمال نماید.



شکل ۲: فهرست بانک‌های اطلاعاتی و جداول در رابط کاربری کنترل کیفیت



شکل ۳: فهرست الگوریتم‌های GTSPG در رابط کاربری کنترل کیفیت

فهرست الگوریتم‌های کنترل کیفیت GTSPG که می‌بایسب بر روی مجموعه داده‌ها اجرا شوند به شرح زیر هستند:

۱- تست تاریخ/زمان غیر ممکن

در این الگوریتم درست بودن فرمت تاریخ و زمان هر رکورد داده، یک به یک چک می‌شود. در مورد تاریخ، فرمت مورد نظر از ۱۰ کاراکتر تشکیل می‌شود. ۴ کاراکتر برای تاریخ، ۲ کاراکتر برای ماه و ۲ کاراکتر برای روز. برای جداکننده‌ها هم باید ۲ کاراکتر در نظر گرفته شود. یک کلیشه برای این کار عبارت است از yyyy-mm-dd. برنامه‌ای که برای این کار نوشته شده است ابتدا تعداد کاراکترهای بکار رفته را بررسی کرده سپس به کنترل اجزای آن می‌نماید. مثلاً نمی‌توانیم ماه سیزدهم داشته باشیم یا روز چهل و سوم. در مورد زمان هم به همین شکل نیازمند به ۵ (بدون احتساب ثانیه) یا ۸ (با احتساب ثانیه) کاراکتر می‌باشیم و از فرمت hh:mm:ss استفاده کرد. حد و حدود هم بدیهی است که مثلاً نمی‌توانیم دقیقه ۸۰ داشته باشیم.

خروجی این الگوریتم، فیلدی با نام DateAndTimeQC را به جدول داده‌ها مطابق با شکل ۴ اضافه می‌نماید.

Code	PRESCX01	TEMP	PSAL	Date	DateAndTimeQC
DE080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	69.4	26.759	34.315	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	79.2	25.656	34.337	2000-01-02 07:44:59.000	1

شکل ۴: فیلد DateAndTimeQC در جدول داده‌ها

۲- تست مکان غیرممکن

با توجه به محدودیت‌های منطقی در مورد طول و عرض جغرافیایی، این الگوریتم مقادیر موجود را مورد بازبینی قرار می‌دهد. مثلاً عرض جغرافیایی باید مقداری عددی بین ۹۰- تا ۹۰+ باشد و نیز طول جغرافیایی هم باید مقداری عددی مابین ۱۸۰- تا ۱۸۰+ باشد.

خروجی این الگوریتم، فیلدی با نام LocationQC را به جدول داده‌ها مطابق با شکل ۵ اضافه می‌نماید.

Code	PRESCX01	TEMP	PSAL	Date	LocationQC
DE080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۵: فیلد LocationQC در جدول داده‌ها

۳- تست موقعیت روی زمین

در ابتدا فایل Etopo مربوط باید بارگزاری شود سپس موقعیت جغرافیایی یک به یک نقاط مربوط به هر رکورد داده باید در این فایل مورد جستجو قرار گیرد. هرگاه چنین نقطه‌ای یافت شد، این بدان معناست که نقطه مورد نظر در دریا قرار دارد، در غیر این صورت نقطه مزبور در خشکی واقع شده است. لازم به توضیح است که برخی نقاط برداشت داده، که خیلی به ساحل نزدیک است ممکن است که خشکی در نظر گرفته شود. تصحیح چنین مواردی که بسیار به ندرت هم اتفاق می‌افتد، غیرممکن به نظر می‌رسد.

لازم به ذکر است، برای پیاده‌سازی از الگوریتم‌های Convex Hull و Concave Hull هم استفاده شده بود که بعد از بررسی‌های به عمل آمده، روش کار به کمک فایل Etopo کارآمدتر و دقیق‌تر تشخیص داده شد.

خروجی این الگوریتم، فیلدی با نام PositionOnLandQC را به جدول داده‌ها مطابق با شکل ۶ اضافه می‌نماید.

Code	PRESCX01	TEMP	PSAL	Date	PositionOnLandQC
DE080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	69.4	26.759	34.315	2000-01-02 07:44:59.000	1

شیوه نامه کنترل کیفیت داده ها

شکل ۶: فیلد PositionOnLandQC در جدول داده‌ها

۴- تست مقادیر پارامترهای غیر ممکن جهانی

در اینجا مقادیر پارامترها از این نظر که در محدوده معتبری قرار گرفته باشند مورد سنجش و بازبینی قرار می‌گیرد. در ابتدا فایلی حاوی این محدوده بارگزاری شده و سپس مقادیر موجود در هر رکورد داده با این مقادیر مورد مقایسه قرار می‌گیرد. البته از قبل باید پارامتر مورد نظر در فایل بارگزاری شده پیدا شود. این مقادیر در جدول ۱-۳ ذکر شده‌اند. در مواردی نادر که ممکن است واحد اندازه گیری مطابق این جدول نباشد، عمل تبدیل واحد هم لازم به انجام است.

خروجی این الگوریتم ستونی با فرمت "نام پارامتر + QC" را به جدول داده‌ها اضافه می‌نماید. به عنوان مثال ستون PSALQC برای پارامتر PSAL مطابق شکل ۷ اضافه می‌شود.

Code	PRESCX01	TEMP	PSAL	Date	PSALQC
DE080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۷: فیلد PSALQC در جدول داده‌ها

۵- تست مقادیر پارامترهای غیر ممکن منطقه‌ای

پیاده سازی این الگوریتم مانند الگوریتم ۴ می‌باشد، اما در اینجا ناحیه مورد نظر از دیدگاه جهانی و عمومی مورد بررسی قرار نمی‌گیرد، بلکه به طور منطقه‌ای و خاص مورد ارزیابی قرار خواهد گرفت. در اسناد GTSPP، مقادیر و محدوده‌های مورد نیاز این الگوریتم تنها برای دو پارامتر (به علاوه عمق)، دما و شوری مورد بررسی قرار گرفته است که البته شامل مناطق دریایی ایران نمی‌باشد. ما برای پیاده سازی، از مراجع معتبر دیگری مانند بانک اطلاعاتی جهانی^۸ و نیز برخی مقالات معتبر در این زمینه و همچنین تجارب متخصص فیزیک دریای همکار طرح، استفاده کردیم که در جدول ۱ نشان داده شده است. لازم به ذکر است مقادیر مختلف برای یک پارامتر ممکن است در نواحی مختلف متفاوت باشد. به عنوان مثال مقدار پارامتر DOXY در ناحیه دریای خزر باید عددی بین ۴.۲ تا ۱۱ باشد ولی در ناحیه خلیج فارس همین مقدار باید بین ۰ تا ۱۲ باشد. در این جدول برخی پارامترها ممکن است ذکر نشده باشد که به دلیل عدم وجود منبع قابل اعتماد برای آنها می‌باشد. به عنوان مثال نتوانستیم منبع معتبری برای محدوده عددی پارامتر DOXY را در دریای عمان پیدا کنیم. بنابراین در این ناحیه این پارامتر در جدول وجود ندارد؛ به عبارتی، کنترل کیفیتی روی این پارامتر در آن ناحیه انجام نشده است.

جدول ۱: مقادیر پارامترهای غیر ممکن در مناطق دریایی ایران

منطقه	مقدار	مقدار	شرح پارامتر	نام استاندارد
-------	-------	-------	-------------	---------------

⁸ World Ocean Database



شیوه نامه کنترل کیفیت داده ها



پارامتر		کمینه	بیشینه	جغرافیا بی
CPWC	Chlorophyll	۱	۸	دریای خزر
LCDAZZ01	Current Direction	۰	۳۶۰	
LCSAZZ01	Current Speed	۰	۱	
DEPTH	-	۰	۱۰۲۵	
DOXY	Oxygen	۴.۲	۱۱	
PHXXZZXX	pH	۸.۲	۸.۸	
PSAL	Salinity	۰	۱۳.۶	
SVEL	Sound velocity	۱۴۲۵	۱۵۷۵	
TEMP	Water Temperature	۰	۲۸	
GTDHZZ01	Significant Wave Height	۰	۱۰	
WVST	Mean Wave Period	۰	۸	
WIND10DF	Wind Direction	۰	۳۶۰	
WIND10SP	Wind Speed	۰	۲۵	
ALKYZZXX	Alkalinity	۲	۲.۸	
NTRI	Nitrite	۰	۱۸	
PHOS	Phosphate	۰	۴.۵	
SLCA	Silicate	۰	۲۰۰	
CPWC	Chlorophyll	۰	۵	
LCDAZZ01	Current Direction	۰	۳۶۰	
LCSAZZ01	Current Speed	۰	۳	
DEPTH	-	۰	۹۳	
DOXY	Oxygen	۰	۱۲	
PHXXZZXX	pH	۶	۹.۳	
PSAL	Salinity	۰	۴۲	
SVEL	Sound velocity	۱۴۲۵	۱۵۷۵	
TEMP	Water Temperature	-۳	۳۵	

GTDHZZ01	Significant Wave Height	۰	۴	
WVST	Mean Wave Period	۰	۱۰	
WIND10DF	Wind Direction	۰	۳۶۰	
WIND10SP	Wind Speed	۰	۲۵	
TEMP	Water Temperature	-۳	۳۵	اقیانوس هند
PSAL	Salinity	۰	۴۰	
DOXY	Oxygen	۰	۱۰	
PHOS	Phosphate	۰	۴.۵	
SLCA	Silicate	۰	۲۵۰	
NTRA	Nitrate	۰	۵۴	
PHXXZZXX	pH	۷.۱	۸.۹	
CPWC	Chlorophyll	۰	۵۰	
ALKYZZXX	Alkalinity	۰	۲.۸	

در برخی از بانک‌های اطلاعاتی موجود بجای عمق از پارامتر فشار استفاده شده است که باید عمل تبدیل صورت می‌گرفت. فرایند این تبدیل به این قرار است. ابتدا باید پارامتر شتاب گرانش (g) را بر حسب عرض جغرافیایی محاسبه کنیم:

$$g \text{ (m/sec}^2\text{)} = 9.780318 * [1.0 + (5.2788 \times 10^{-3} + 2.36 \times 10^{-5} * x) * x] + 1.092 \times 10^{-6} * p$$

که در آن متغیر x به صورت $x = [\sin(\text{latitude} / 57.29578)]^2$ قابل محاسبه می‌باشد. متغیر p هم فشار بر حسب دسی بار است. با از محاسبه g، فرمول بدست آوردن مقدار عمق بر حسب فشار به قرار زیر می‌باشد:

$$\text{Depth (meters)} = \frac{[(((-1.82 \times 10^{-15} * p + 2.279 \times 10^{-10}) * p - 2.2512 \times 10^{-5}) * p + 9.72659) * p]}{g}$$

خروجی این الگوریتم پارامتری با فرمت "نام پارامتر + RegionalQC" را به جدول داده‌ها اضافه می‌نماید. به عنوان مثال فیلد PSALRegionalQC برای پارامتر PSAL مطابق شکل ۸ اضافه می‌شود.

Code	PRESCX01	TEMP	PSAL	Date	PSALRegionalQC
DE080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۸: فیلد PSALRegionalQC در جدول داده‌ها

۶- تست محدوده نیم رخ جهانی

در این الگوریتم مقدار پارامترهای دما و شوری در عمق‌های مختلف بررسی می‌شود تا این اطمینان حاصل گردد که این مقادیر در محدوده درست خود قرار دارند. لازم به ذکر است که الگوریتم فوق برای دیگر پارامترها هم کارایی دارد ولی در اینجا ما منبع قابل اعتمادی پیدا نکردیم و لذا به همان دما و شوری چنانکه در اسناد GTSPG اشاره شده است، اکتفا کردیم. محدوده‌های معتبر برای دما و شوری در اعماق مختلف در جدول ۲ آمده است.

جدول ۲: مقادیر پارامترهای محدوده نیم رخ جهانی

Depth (metres)	Range Temperature (degrees C)	Salinity (psu)
0 to 25	-2.0 to 37	0 to 41
>25 to 50	-2.0 to 36	0 to 41
>50 to 100	-2.0 to 36	1 to 41
>100 to 150	-2.0 to 34	3 to 41
>150 to 200	-2.0 to 33	3 to 41
>200 to 300	-2.0 to 29	3 to 41
>300 to 400	-2.0 to 27	3 to 41
>400 to 1100	-2.0 to 27	10 to 41
>1100 to 3000	-1.5 to 18	22 to 38
>3000 to 5500	-1.5 to 7	33 to 37
>5500 to 12000	-1.5 to 4	33 to 37

خروجی این الگوریتم، ستونی با فرمت "نام پارامتر + GlobalProfileEnvelopeQC" را به جدول داده‌ها اضافه می‌نماید.

به عنوان مثال ستون PSALGlobalProfileEnvelopeQC برای پارامتر PSAL مطابق شکل ۹ اضافه می‌شود.

	PRESCX01	TEMP	PSAL	Date	PSALGlobalProfileEnvelopeQC
2-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
2-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
2-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۹: فیلد PSALGlobalProfileEnvelopeQC در جدول داده‌ها

۷- تست نقطه انجماد

در این الگوریتم با توجه به عمق و مقدار شوری بررسی می‌شود که دمای آب پایین‌تر از نقطه انجماد نباشد. دمای انجماد به وسیله رابطه زیر محاسبه می‌شود:

$$T = -0.0575 * S + 1.710523E - 3S^{3/2} - 2.154996E - 4 * S^2 - 7.53E - 4 * P$$

در فرمول بالا T دمای انجماد، S مقدار شوری در عمقی با فشار P می‌باشد. تبدیل فشار و عمق بر حسب متر در تست ۵ توضیح داده شده است. دمای اندازه گیری شده نباید از مقدار محاسبه شده فوق کمتر باشد، در غیر این صورت احتمال خطا در اندازه‌گیری وجود دارد.

خروجی این الگوریتم، ستونی با نام FreezingPointQC مطابق با شکل ۱۰ به جدول داده‌ها اضافه می‌نماید.

Code	PRESCX01	TEMP	PSAL	Date	FreezingPointQC
4E0BEDE6-3EAC-44C1-9F9C-5E615ABF92AE	8	27.832	34.809	2003-03-01 18:02:37.000	0
4E0BEDE6-3EAC-44C1-9F9C-5E615ABF92AE	9.5	27.838	34.809	2003-03-01 18:02:37.000	0
4E0BEDE6-3EAC-44C1-9F9C-5E615ABF92AE	19.5	27.841	34.809	2003-03-01 18:02:37.000	0

شکل ۱۰: فیلد FreezingPointQC در جدول داده‌ها

۸- تست داده‌های پرت

در این الگوریتم، یک حد آستانه برای اختلاف سه مقدار متوالی از یک پارامتر که بر حسب عمق مرتب شده باشند، تعریف می‌شود. هیچ‌گاه این اختلاف نباید از این حد آستانه تجاوز کند. اسناد GTSP این حد آستانه را تنها برای پارامترهای دما و شوری تعریف کرده است. ما در اینجا منبع قابل اعتمادی را برای یافتن این حد آستانه در مورد پارامترهای دیگر، نیافتیم؛ لذا این الگوریتم فقط بر روی دو پارامتر دما و شوری اعمال شد. این الگوریتم مبتنی بر رابطه زیر می‌باشد:

$$|V2 - (V3 + V1)/2| - |V1 - V3| / 2 > V_THRESHOLD$$

متغیرهای $V1$ ، $V2$ ، و $V3$ عبارت است از مقادیر متوالی (بر حسب عمق) مربوط به یک پارامتر می‌باشد (دما یا شوری). متغیر $V_THRESHOLD$ برای شوری ۰.۳ و برای دما ۲ می‌باشد. اگر مقادیر از این حد آستانه تجاوز کند، مقدار $V2$ احتمالاً غلط اندازه‌گیری شده است.

خروجی این الگوریتم پارامتری با فرمت "نام پارامتر + SpikeQC" را به جدول داده‌ها اضافه می‌نماید. به عنوان مثال ستون PSALSpikeQC برای پارامتر PSAL مطابق شکل ۱۱ اضافه می‌شود.

Code	PRESCX01	TEMP	PSAL	Date	PSALSpikeQC
DE080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
DE080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۱۱: فیلد PSALSpikeQC در جدول داده‌ها

۹- تست داده‌های پرت بالا و پایین

شیوه نامه کنترل کیفیت داده ها

شرح این الگوریتم دقیقاً مانند تست ۷ است و تنها تفاوت در شرط وقوع Spike است. اگر $VDN < (V1 - V2) < VUP$ باشد، یک Top Spike، و اگر $VDN < (V2 - V1) < VUP$ ، یک Bottom Spike رخ داده است. در فرمول فوق $V1$ و $V2$ ، به ترتیب مقادیر دما یا شوری دو عمق متوالی است. و مقادیر VUP و VDN برای پارامترهای دما و شوری به ترتیب زیر می باشد.

Parameter	VDN	VUP
Temperature	-10	+10
Salinity	-5	+5

اجرای این الگوریتم در پایان دو ستون با الگوی "نام پارامتر + BottomSpikeQC" و "نام پارامتر + TopSpikeQC" را به جدول داده‌ها اضافه می‌کند. به عنوان مثال ستون PSALBottomSpikeQC برای پارامتر شوری مطابق شکل ۱۲ اضافه می‌شود.

de	PRESCX01	TEMP	PSAL	Date	PSALBottomSpikeQC
:080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
:080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
:080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1
:080002-0622-405A-A070-C166A818F535	69.4	26.759	34.315	2000-01-02 07:44:59.000	1

شکل ۱۲: فیلد PSALBottomSpikeQC در جدول داده‌ها

۱۰- تست شیب

در این الگوریتم گرادیان دو نقطه مجاور از این نظر که گرادیان آنها از حد معینی تجاوز ننماید مورد تست قرار می‌گیرند. این الگوریتم حداقل به سه نقطه در سه عمق متوالی نیاز دارد. می‌توان این‌طور تصور کرد که در اینجا، اینکه میزان پارامترهای دما یا شوری در دو عمق متوالی، با شیب مشخصی افت یا افزایش یافته است یا خیر مورد آزمون قرار می‌گیرد. با استفاده از فرمول $|V2 - (V1 + V3)/2| > V_GRAD$ - این مورد را می‌توان مشخص نمود. در این فرمول، پارامتر V_GRAD به یک حد آستانه مورد قبول اشاره دارد که برای دما عدد ۱۰ و برای شوری عدد ۵ می‌باشد. پارامترهای $V1$ تا $V3$ مقدار شوری یا دما در اعماق متوالی می‌باشد. در صورتی که شرط مورد نظر در فرمول بالا برقرار باشد، در اندازه گیری مقدار پارامتر $V2$ ، احتمالاً اشتباهی رخ داده است.

اجرای این الگوریتم، ستونی با الگوی "نام پارامتر + GradientTestQC" به جدول داده‌ها اضافه می‌نماید. به عنوان مثال ستون TEMPGradientTestQC برای پارامتر دما مطابق شکل ۱۳ اضافه می‌شود.

de	PRESCX01	TEMP	PSAL	Date	PSALGradientTestQC
:080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
:080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
:080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۱۳: فیلد TEMPGradientTestQC در جدول داده‌ها

۱۱- تست وارونگی چگالی

در این الگوریتم، مقدار پارامتر چگالی در دو عمق متوالی باهم مقایسه می‌شود. مقدار چگالی باید با افزایش عمق افزایش یابد، در غیر این صورت پدیده وارونگی چگالی رخ داده است و مقدار مربوط به آن نامعتبر است. این الگوریتم، ستونی به نام DensityInversionQC مطابق شکل ۱۴ به جدول داده‌ها اضافه می‌نماید.

Id	PRESCX01	TEMP	PSAL	Date	PSALGradientTestQC
080002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
080002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
080002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۱۴: فیلد DensityInversionQC در جدول داده‌ها

۱۲- تست پایین

این الگوریتم نیازمند فایل عمق‌سنجی منطقه مورد تست می‌باشد. مرکز داده‌ها از فایل عمق‌سنجی Etopo1 با دقت یک دقیقه استفاده می‌کند. در این الگوریتم نقطه مورد تست از نظر عمق مجاز مورد آزمون قرار می‌گیرد. در ابتدا نقطه متناظر (یا نزدیک‌ترین نقطه) در فایل عمق‌سنجی یافت شده و به بیشترین عمق در ۱۶ نقطه اطراف آن، ۵۰ واحد می‌افزاییم و سپس عدد به دست آمده را با عمق نقطه مورد آزمون مقایسه می‌نماییم. اگر بزرگ‌تر بود، این اندازه‌گیری نادرست بوده است. به عنوان مثال عمق نقطه‌ای ممکن است ۲۰۰ متر باشد و هدف، آزمون صحت اندازه این عمق است. ابتدا نزدیک‌ترین نقطه متناظر در فایل عمق‌سنجی را می‌یابیم و ۱۶ نقطه اطراف آن را نیز پیدا می‌کنیم. سپس به حداکثر عمق در میان این ۱۶ نقطه ۵۰ واحد می‌افزاییم. فرض می‌کنیم حداکثر عمق یافت شده ۱۷۰ متر است. ۵۰ واحد به ۱۷۰ واحد افزوده و آن را با ۲۰۰ متر عمق نقطه مورد نظر مقایسه می‌کنیم. از آنجایی که برای نقطه مورد نظر حداکثر عمق مجاز عدد ۲۲۰ بدست آمد، عمق ۲۰۰ می‌تواند درست باشد.

فایل عمق‌سنجی مورد نیاز برای اجرای این الگوریتم، یک گرید با دقت ۵ دقیقه می‌باشد و ۱۶ نقطه هم مورد نیاز برای آزمون یک نقطه است. با این حال ما برای اجرای این الگوریتم از یک گرید با دقت ۱ دقیقه و از ۲۵ نقطه برای آزمون یک نقطه استفاده کردیم.

این الگوریتم، ستونی به نام BottomTestQC مطابق شکل ۱۵ به جدول داده‌ها اضافه می‌نماید.

code	Latitude	Longitude	date	BottomTestQC
2	39	50	2000-10-15 00:00:00.000	1
2	39	50	2000-10-15 00:00:00.000	1
3	38.99833	49.795	2000-10-15 00:00:00.000	1
3	38.99833	49.795	2000-10-15 00:00:00.000	1
4	38.99833	49.615	2000-10-15 00:00:00.000	1

شکل ۱۵: فیلد BottomTestQC در جدول داده‌ها

۱۳- تست وارونگی دما

این تست مانند تست وارونگی چگالی عمل می‌نماید و این مهم را که دما با افزایش عمق، کاهش یابد را مورد آزمون قرار می‌دهد. این الگوریتم تنها در عمق‌های بالای ۷۵ متر و برای سطوح مجاور با فاصله کمتر از ۵۰ متر درست اجرا می‌شود. مثلاً هرگاه دمای دو نقطه متوالی اندازه‌گیری شده در اعماق ۲۰ و ۱۰۰ متری باشند، چون فاصله آنها بیش از ۵۰ متر است، این الگوریتم را روی آنها قابل اجرا نیست. علاوه بر این تفاوت دمای بین دو نقطه متوالی می‌تواند کمتر از ۰.۱ باشد. این الگوریتم، ستونی به نام TemperatureInversionQC مطابق شکل ۱۶ به جدول داده‌ها اضافه می‌نماید.

	PRESCX01	TEMP	PSAL	Date	TemperatureInversionQC
0002-0622-405A-A070-C166A818F535	49.8	28.059	34.196	2000-01-02 07:44:59.000	1
0002-0622-405A-A070-C166A818F535	49.8	28.031	34.2	2000-01-02 07:44:59.000	1
0002-0622-405A-A070-C166A818F535	59.8	28.173	34.189	2000-01-02 07:44:59.000	1

شکل ۱۶: فیلد BottomTestQC در جدول داده‌ها