



**UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS**

Facultad del Medio Ambiente y Recursos Naturales
Ingeniería Ambiental

**“APLICACIÓN DE TÉCNICAS DE HOMOGENIZACIÓN DE
SERIES DE TIEMPO DE VARIABLES OCEANOGRÁFICAS,
ESTUDIO DE CASO SUBREGIÓN DEL PACÍFICO
COLOMBIANO, COMO APORTE AL TEMA DE MONITOREO
AMBIENTAL”**

DIANA PAULINA CASTAÑEDA RODRÍGUEZ

20121180004

BOGOTÁ D.C

2017



**UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS**

Facultad del Medio Ambiente y Recursos Naturales
Ingeniería Ambiental

COMITÉ TÉCNICO NACIONAL DE COORDINACIÓN DE DATOS E INFORMACIÓN OCEÁNICA (CTN Diocean)

**Grupo de Trabajo en Mejores Prácticas en Gestión de Datos
GT-MPGD**

Semillero Grupo Ambiental de Investigación Avanzada G.A.I.A.

PROYECTO DE INVESTIGACIÓN

**“APLICACIÓN DE TÉCNICAS DE HOMOGENIZACIÓN DE SERIES DE
TIEMPO DE VARIABLES OCEANOGRÁFICAS, ESTUDIO DE CASO
SUBREGIÓN DEL PACÍFICO COLOMBIANO, COMO APORTE AL TEMA
DE MONITOREO AMBIENTAL”**

Estudiante tesista: Diana Paulina Castañeda Rodríguez

**Trabajo de grado en la modalidad de investigación para optar al título
de Ingeniera Ambiental**

Docente director: Néstor Ricardo Bernal Suárez

Docente asesora interna: Martha Cecilia Gutiérrez Sarmiento

**Asesora Externa de la Dirección General Marítima - Centro
Colombiano de Datos Oceanográficos: Ruby Viviana Ortiz Martínez**

BOGOTÁ D.C

2017

Nota de aceptación

Director
Néstor Ricardo Bernal Suárez

Asesora interna
Martha Cecilia Gutiérrez Sarmiento

Jurado

Bogotá D.C., 15 de Septiembre de 2017.

NOTA DEDICATORIA

A mis padres Carmen, Miguel y Rosalba por su apoyo incondicional en los momentos cruciales. A Laura y Alejandra por la fuerza, la determinación y los buenos momentos que me han otorgado en este largo camino. A mi incondicional compañero Rufus, por acompañarme en las largas noches de trabajo. A mis docentes Martha y Néstor por los consejos y la guía constante que han sido en este arduo proceso, a Ruby por su colaboración y diligencia en cada paso. A todos mis familiares y amigos sin quienes este logro no hubiese sido posible.

AGRADECIMIENTOS

Al Centro de Investigaciones y Desarrollo Científico (CIDC) de la Universidad Distrital Francisco José de Caldas por el apoyo institucional para la realización del actual proyecto de investigación; a los profesores Carlos Zafra Mejía, director del Grupo de Investigación de Ingeniería Ambiental (GIIAUD); Martha Cecilia Gutiérrez, líder del Grupo Ambiental de Investigación Avanzada (GAIA) y al docente Néstor Ricardo Bernal Suárez por el constante acompañamiento y la guía en todo el proceso de investigación. A la Dirección General Marítima (Dimar) y el Centro Colombiano de Datos Oceanográficos (Cecoldo) por el suministro de información, apoyo logístico y técnico; al Grupo de Trabajo en Mejores Prácticas en Gestión de Datos en el marco del cual se planteó el desarrollo de este proyecto; a Ruby Viviana Ortiz Martínez Coordinadora Técnica Centro Colombiano de Datos Oceanográfico por su apoyo, aportes técnicos y seguimiento constante en el desarrollo de la presente investigación; a Angela Liliana Losada García de gestión de calidad de Cecoldo por la colaboración logística y administrativa; a Milton Puentes Galindo responsable Área de Investigación Científica Marina y Gerente proyecto “Reestructuración y Posicionamiento Estratégico del Cecoldo 2014-2017” de la Subdirección de Desarrollo Marítimo - Área de Investigación Científica Marina de Dimar por las sugerencias hechas respecto a los ajustes en la metodología sobre el uso del Atlas de los Datos Oceanográficos de Colombia 1922-2013 como referencia externa para la comparación evaluación temporal de los datos; al Capitán de Fragata Leonardo Marriaga Rocha Director del Centro de Investigaciones Oceanográficas e Hidrográficas del Pacífico (CCCP), por sus recomendaciones y aclaraciones sobre el comportamiento de parámetros oceanográficos en la Cuenca Pacífica Colombiana y al Contralmirante Juan Manuel Soltau Ospina Secretario Ejecutivo Comisión Colombiana del Océano (CCO).

Tabla de contenido

1.	INTRODUCCIÓN	13
2.	OBJETIVOS	14
3.	JUSTIFICACIÓN	15
4.	MARCO REFERENCIAL.....	16
4.1.	<i>Contexto Investigativo</i>	16
4.2.	<i>Marco Geográfico</i>	17
4.3.	MARCO CONCEPTUAL	18
4.3.1.	<i>Monitoreo Ambiental</i>	18
4.3.1.1.	Observación de los Océanos	21
4.3.2.	<i>Variables oceanográficas y Correlaciones</i>	23
4.3.2.1.	Ciclos biogeoquímicos	23
4.3.2.2.	Temperatura Superficial del Mar (TSM).....	24
4.3.2.3.	Oxígeno Disuelto	26
4.3.2.4.	Nutrientes	27
4.3.2.5.	Instrumentación Oceanográfica	30
4.3.3.	<i>Fenómeno El Niño- La Niña</i>	33
4.3.3.1.	Estudio Regional Fenómeno El Niño ERFEN.....	35
4.3.3.2.	Impactos del Fenómeno El Niña- La Nina en Colombia.....	38
4.3.4.	<i>Calidad de Datos</i>	38
4.3.5.	<i>Procedimientos Estadísticos de Control de Calidad</i>	39
4.3.5.1.	Pruebas de Coherencia	40
4.3.5.2.	Series de Tiempo.....	43
4.3.5.3.	Intervalos de Confianza.....	43
4.3.5.4.	Homogeneidad de Datos.....	46
4.3.6.	<i>Banderas de Calidad (QF)</i>	46
4.3.6.1.	Primer Nivel	47
4.3.6.2.	Segundo Nivel	47
5.	METODOLOGÍA	51
5.1.	<i>Búsqueda bibliográfica</i>	54
5.2.	<i>Inventario de Datos</i>	55
5.3.	<i>Preselección</i>	56
5.4.	<i>Etapa 1: Estandarización de Formatos</i>	57
	Asignación de código ONI	57
5.5.	<i>Etapa 2: Identificación de Datos Atípicos</i>	58

5.6.	<i>Etapa 3: Prueba de Coherencia Temporal</i>	59
5.7.	<i>Etapa 4: Prueba de Coherencia Espacial</i>	60
	Correlograma espacial en PASSAGE	60
5.8.	<i>Etapa 5: Resultados Síntesis de Asignación de Banderas de Calidad</i>	61
5.9.	<i>Etapa 6: Detección de Cambios Entre Períodos Normal vs. El Niño y Normal vs. La Niña.</i>	61
6.	RESULTADOS.....	62
6.1.	<i>Propuesta Metodológica</i>	62
	62
6.2.	<i>Inventario De Datos</i>	63
6.3.	<i>Priorización de Estaciones de Muestreo</i>	70
6.4.	<i>Etapa 1- Estandarización de Formatos</i>	72
6.5.	<i>Etapa 2- Identificación De Datos Atípicos</i>	73
6.6.	<i>Etapa 2- Identificación de Datos Atípicos (Caso de Oxígeno Disuelto)</i>	76
6.7.	<i>Etapa 3- Prueba de Coherencia Temporal</i>	79
6.8.	<i>Etapa 4- Prueba de Coherencia Espacial</i>	88
6.9.	<i>Etapa 5- Banderas de Calidad de Clasificación Final</i>	94
6.10.	<i>Etapa 6- Prueba de Detección de Cambios Entre Períodos Normal Vs. El Niño Y Normal Vs. La Niña</i>	96
6.11.	<i>Resultados Generales</i>	103
	TSM.....	103
	OXIGENO DISUELTO	104
	NUTRIENTES NITRATOS (NO3)	105
7.	ANÁLISIS Y DISCUSIÓN DE RESULTADOS	106
7.1.	<i>Etapa 2- Identificación de Datos Atípicos</i>	106
7.2.	<i>Etapa 3- Prueba de Coherencia Temporal</i>	107
7.3.	<i>Etapa 4- Prueba de Coherencia Espacial</i>	108
7.4.	<i>Etapa 6- Prueba de Detección de Cambios Entre Períodos Normal Vs. El Niño Y Normal Vs. La Niña</i>	109
8.	CONCLUSIONES	111
9.	RECOMENDACIONES	112
10.	BIBLIOGRAFÍA.....	113

LISTA DE TABLAS

Tabla I. Descripción de los datos promedio de temperatura en la Cuenca Pacífica Colombiana.

Tabla II. Banderas de calidad IODE.

Tabla III. Pruebas de control de calidad IODE de nivel secundario.

Tabla IV. Banderas de calidad GTSP.

Tabla V. Banderas de calidad de la OceanSites.

Tabla VI. Síntesis de la metodología.

Tabla VII. Guías metodológicas revisadas.

Tabla VIII. Tipos de Datos suministrados por la Dimar para el período 1970-2016.

Tabla IX. Variables oceanográficas presentes en los datos 1970-2016 suministrados por la Dimar.

Tabla X. Distribución temporal de los datos 1970-2016 suministrados por la Dimar.

Tabla XI. Coordenadas de las 21 estaciones prioritarias.

Tabla XII. Información de Atlas de los Datos Oceanográficos de Colombia correspondiente a la estación 25, coordenadas 6°N, -79°W.

Tabla XIII. Índice de Moran mensual para datos de TSM.

Tabla XIV. Ejemplo de clasificación de banderas de calidad para TSM y su equivalencia en banderas IODE.

Tabla XV. Ejemplo de clasificación de banderas de calidad para oxígeno disuelto y nitratos con su equivalencia en banderas IODE.

Tabla XVI. Resultados de la prueba de Mann-Whitney por estación para TSM, oxígeno disuelto y nitratos.

Tabla XVII. Anomalías por estación para TSM, oxígeno disuelto y nitratos para periodos El Niño y La Niña.

Tabla XVIII. Resumen general de las anomalías para TSM, oxígeno disuelto y nitratos para periodos El Niño y La Niña.

Tabla XIX. Resultado de asignación de banderas de calidad por etapas para TSM.

Tabla XX. Resultado de promedio de TSM por épocas Normal, El Niño y La Niña.

Tabla XXI. Resultado de asignación de banderas de calidad por etapas para Oxígeno Disuelto a 0m de profundidad.

Tabla XXII. Resultado de promedio de oxígeno disuelto a 0m por épocas Normal, El Niño y La Niña.

Tabla XXIII. Resultado de asignación de banderas de calidad por etapas para Oxígeno Disuelto a 0m de profundidad.

Tabla XXIV. Resultado de promedio de oxígeno disuelto a 0m por épocas Normal y La Niña.

LISTA DE FIGURAS

Figura 1. DIMAR (2016) Sistema de Medición de Parámetros Oceanográficos y de Meteorología Marina - SMPOMM.

Figura 2. Ciclo del oxígeno.

Figura 3. Ciclo del nitrógeno en los ambientes marinos.

Figura 4. Las cinco etapas del ciclo del nitrógeno.

Figura 5. Perfilador de ejemplo de Conductividad, Temperatura y Presión (CTD) SEA-BIRD. Modelo SBE 19 plus V2. 600 m.

Figura 6. Termómetro de cazoleta.

Figura 7. Botella Niskin.

Figura 8. Umbrales para identificar la magnitud de El niño/La Niña, según la NOAA en la región Pacífico tropical Niño 3.4.

Figura 9. Mecanismos institucionales Protocolo ERFEN.

Figura 10. Malla ERFEN Cuenca Pacífica Colombiana (CPC).

Figura 11. Gráfico de la distribución normal.

Figura 12. Ejemplo de formato de datos suministrados por Dimar.

Figura 13. Ejemplo de documento con asignación de código ONI Etapa 1.

Figura 14. Valores del Oceanic Niño Index por meses y años.

Figura 15. Explicación partes gráfico de caja.

Figura 16. Metodología general propuesta para el control de calidad de datos de TSM.

Figura 17. Metodología ajustada para el control de calidad de oxígeno disuelto.

Figura 18. Malla ERFEN Cuenca Pacífica Colombiana.

Figura 19. Estaciones prioritarias (señaladas en rojo) para las variables de temperatura y oxígeno disuelto.

Figura 20. Estaciones prioritarias (señaladas en rojo) para la variable de nutrientes.

Figura 21. Ejemplo gráfico de caja para una estación con datos de TSM. Etapa 2 parte 1.

Figura 22. Equivalencias de banderas de calidad GTSP.

Figura 23. Equivalencias de banderas de calidad OceanSites Etapa 1.

Figura 24. Ejemplo gráfico de caja para una estación sin presencia de datos atípicos de TSM. Etapa 2 parte 2.

Figura 25. Ejemplo gráfico de caja para una estación con datos de oxígeno disuelto. Etapa 2 parte 1

Figura 26. Equivalencias de banderas de calidad OceanSites para la Etapa 1.

Figura 27. Ejemplo gráfico de caja para una estación sin presencia de datos atípicos para oxígeno disuelto. Etapa 2 parte 2

Figura 28. Ejemplo Intervalos de confianza t-Student para una estación.

Figura 29. Equivalencias de banderas de calidad GTSP para Etapa 3.

Figura 30. Distribución de las isothermas para media mensual de TSM en el mes de enero para la CPC.

Figura 31. Distribución de las isothermas para media mensual de TSM en el mes de febrero para la CPC.

Figura 32. Distribución de las isothermas para media mensual de TSM en el mes de marzo para la CPC.

Figura 33. Distribución de las isothermas para media mensual de TSM en el mes de abril para la CPC.

Figura 34. Distribución de las isothermas para media mensual de TSM en el mes de mayo para la CPC.

Figura 35. Distribución de las isothermas para media mensual de TSM en el mes de junio para la CPC.

Figura 36. Distribución de las isothermas para media mensual de TSM en el mes de julio para la CPC.

Figura 37. Distribución de las isothermas para media mensual de TSM en el mes de agosto para la CPC.

Figura 38. Distribución de las isothermas para media mensual de TSM en el mes de septiembre para la CPC.

Figura 39. Distribución de las isothermas para media mensual de TSM en el mes de octubre para la CPC.

Figura 40. Distribución de las isothermas para media mensual de TSM en el mes de noviembre para la CPC.

Figura 41. Distribución de las isothermas para media mensual de TSM en el mes de diciembre para la CPC.

Figura 42. Correlograma de datos de mensuales TSM en software *PASSAGE*.

Figura 43. Estimación de rango de distancia de datos mensuales de TSM por Correlograma en software *PASSAGE*.

Figura 44. Rango de distancia de estaciones vecinas para TSM mensual de una estación, mediante el software ArcGis.

Figura 45. Estimación de rango de distancia de datos mensuales de oxígeno disuelto por Correlograma en software *PASSAGE*.

Figura 46. Rango de distancia de estaciones vecinas para oxígeno disuelto mensual de una estación, mediante el software ArcGis.

Figura 47. Estaciones seleccionadas para distribución homogénea de datos de nitratos para ensayo en software Pasage.

Figura 48. Estimación de rango de distancia de datos mensuales de nitratos con coordenadas homogéneas por Correlograma en software *PASSAGE*.

Figura 49. Rango de distancia de estaciones vecinas para nitratos de una estación, mediante el software ArcGis.

Figura 50. Equivalencias de banderas de calidad GTSP para Etapa 5.

Figura 51. Equivalencias de banderas de calidad OceanSites para Etapa 5.

Figura 52. Distribución espacial de las anomalías de TSM por estación para la CPC.

Figura 53. Valor de las anomalías de TSM por estación para época El Niño.

Figura 54. Valor de las anomalías de TSM por estación para época La Niña.

Figura 55. Valor de las anomalías de oxígeno disuelto (0m) por estación para época El Niño.

Figura 56. Valor de las anomalías de oxígeno disuelto (0m) por estación para época La Niña.

Figura 57. Valor de las anomalías de nitratos (0m) por estación para época La Niña.

LISTA DE ANEXOS

Anexo 1: Marco Referencial.

Anexo 2: SCRIPT_co_tesisdimar_UD_R_temperatura_0m_etapa2.

Anexo 3: SCRIPT_co_tesisdimar_UD_R_temperatura_0m_etapa3.

Anexo 4: SCRIPT_co_tesisdimar_UD_R_temperatura_0m_etapa4.

Anexo 5: SCRIPT_co_tesisdimar_UD_R_temperatura_0m_etapa6.

Anexo 6: SCRIPT_co_tesisdimar_UD_R_OD_0m_etapa2.

Anexo 7: SCRIPT_co_tesisdimar_UD_R_OD_0m_etapa4.

Anexo 8: SCRIPT_co_tesisdimar_UD_R_OD_0m_etapa6.

Anexo 9: SCRIPT_co_tesisdimar_UD_R_NO3_0m_etapa2.

Anexo 10: SCRIPT_co_tesisdimar_UD_R_NO3_0m_etapa4.

Anexo 11: SCRIPT_co_tesisdimar_UD_R_NO3_0m_etapa6.

Anexo 12: Banco_Imagenes 1970_2016.

Anexo 13: Resultados por Estación de Parametros Oceanográficos Etapa de la 2 a la 5.

Anexo 14: Resultados por Estación de Parametros Oceanográficos Etapa 6 y Series de Tiempo.

Anexo 15: Artículo Colaboración Universidad Distrital Dimar **

NOTA: Los anexos denominados como SCRIPT corresponden a un código escrito en el Software R versión 3.3.1 paquete Rcmdr, utilizado para la ejecución de la metodología y obtención de los resultados producto de la presente investigación. Cada SCRIPT corresponde a una etapa del proceso y parámetro oceanográfico de acuerdo a las técnicas estadísticas aquí descritas, y contiene en su primer ejemplo la explicación de lo que el código significa y los procesos que ejecuta, se encuentra disponible solo en la versión digital que acompaña al presente documento.

** El artículo anexo a este documento en cumplimiento del quinto objetivo del proyecto: Realizar un artículo en colaboración Universidad Distrital y La Dimar es una primera versión sujeta a modificaciones y en proceso de edición.

1. INTRODUCCIÓN

El presente proyecto de investigación hace una aplicación práctica de técnicas de homogenización estadística a series de tiempo de variables oceanográficas (temperatura, oxígeno disuelto y nutrientes) con el propósito de aportar en el desarrollo de un protocolo para el procesamiento de información que permita evaluar la calidad de datos oceanográficos en el marco del monitoreo ambiental para el seguimiento del estado de los recursos naturales marinos.

El protocolo propuesto toma elementos climatológicos, experiencias en homogenización de series de tiempo de variables como la precipitación, y guías e investigaciones similares sobre datos oceanográficos, climatológicos, procesos estadísticos, procedimientos de control de calidad e incluso monitoreo y planificación ambiental.

Con los antecedentes en mención se construyó una metodología para la evaluación de datos oceanográficos a nivel temático, espacial y temporal; y la posterior asignación de banderas de calidad, que permite contar con datos confiables para su aplicación en análisis sobre la variabilidad y la correlación entre variables físicas y químicas (temperatura, oxígeno disuelto y nutrientes, con énfasis en nitratos) como base para la búsqueda de soluciones acertadas respecto a la gestión de los recursos naturales marinos. También se abordó la variación de los parámetros oceanográficos frente a los cambios ambientales considerados normales, y en períodos con ocurrencia de los fenómenos climáticos de los últimos años (El Niño y La Niña) como un aporte al monitoreo ambiental de los océanos.

El proceso aplicado a los datos oceanográficos incluye técnicas de homogenización como una pauta clave para la obtención de datos confiables. Este proceso consta de (i) la detección de datos atípicos, (ii) aplicación de pruebas de coherencia para la homogenización de las series, (iii) detección de diferencias en el promedio de las series y (iv) clasificación de datos según el esquema de banderas de calidad de datos (*Quality Flags*) del programa para el Intercambio Internacional de Datos e Información Oceanográfica (IODE por sus siglas en inglés), de la Comisión Oceanográfica Intergubernamental (Unesco-COI).

Los datos utilizados para la aplicación de la metodología propuesta, corresponde al registro de variables de oceanografía física y química, suministrados por la Dirección General Marítima (Dimar) desde 1970 hasta 2016 procedentes del Pacífico Colombiano como zona objeto de estudio.

2. OBJETIVOS

General

Aplicar técnicas de homogeneización de series de tiempo de variables oceanográficas como aporte al tema de monitoreo ambiental.

Específicos

1. Realizar un análisis exploratorio y descriptivo de las variables oceanográficas tanto físicas como químicas.
2. Analizar los casos de datos faltantes de series de tiempo de variables oceanográficas.
3. Evaluar la calidad de los datos oceanográficos.
4. Explorar la relación entre variables físicas y químicas, como un aporte al tema de monitoreo ambiental.
5. Realizar un artículo en colaboración Universidad Distrital y La Dimar.

3. JUSTIFICACIÓN

La gestión de datos oceanográficos constituye un reto institucional que requiere de la colaboración internacional de todos los países que comparten el recurso oceánico y desde donde existe una preocupación por el estado y evolución del mismo. Tal labor incluye la recolección, control y análisis de una gran cantidad de información sobre parámetros oceanográficos, con el objetivo de emitir decisiones acertadas sobre la gestión de este recurso. Razón por la cual la calidad de los datos como base para cualquier análisis o deducción toma tanta relevancia.

Con el propósito de estandarizar los procesos a los que son sometidos los datos primarios, se han planteado múltiples metodologías, guías y protocolos sobre control de calidad de datos oceanográficos; cada uno responde a las necesidades y criterios de quien lo emite, y aborda la calidad de datos desde perspectivas particulares; tanto así que cuando alguno se intenta aplicar para Colombia, se evidencian elementos que no aplican para las condiciones propias de la región, mientras que carecen de otros criterios a los que se requiere prestar especial atención.

Es por ello que el presente trabajo de investigación pretende abordar en el desarrollo de un protocolo para el procesamiento de datos oceanográficos con base en técnicas estadísticas y procesos de asignación de banderas de calidad, que permita de una parte aplicar mejores prácticas internacionales, y de otra, en la medida de lo posible, responder a las singularidades propias del recurso marino del territorio colombiano y la región, como base para la gestión del recurso oceánico sobre las tareas que conciernen a su seguimiento y monitoreo.

4. MARCO REFERENCIAL

En presenta proyecto de investigación se desarrolló en el marco del Comité Técnico Nacional de Coordinación de Datos e Información Oceánica (CNC- DIOCEAN) y el Grupo de Trabajo en Mejores Prácticas de Gestión de Datos (GTMPG), según el Plan de trabajo 2015-2020 en cuyas actividades para “recomendar mejores prácticas de gestión de datos oceánicos” la Universidad Distrital Francisco José de Caldas participa y aporta con los resultados aquí presentados. El comité es coordinado por la Dirección General Marítima proveedora de los datos aquí trabajados y en convenio interinstitucional con la Universidad Distrital y la Facultad de Medio Ambiente y Recursos Naturales, con el apoyo del Grupo de Investigación en Ingeniería Ambiental (GUIIAUD) y el Semillero G.A.I.A. (Grupo Ambiental de Investigación Avanzada). Los detalles y la profundización sobre el marco referencial se encuentran en el ANEXO 1.

4.1. Contexto Investigativo

Las mejores prácticas de procesamiento de datos oceánicos propuestas desde la Universidad Distrital Francisco José de Caldas comprenden una serie de técnicas estadísticas aplicadas a datos y evaluación de su comportamiento para un análisis. Existen varias investigaciones previas que se han desarrollado sobre procedimientos estadísticos de parámetros ambientales, gestión y procesamiento de datos oceanográficos de las cuales se tomaron varios aportes que consolidan la base del presente proyecto de investigación.

A nivel nacional a este proyecto de investigación le preceden estudios sobre calidad de datos oceanográficos del Pacífico colombiano, tales como “Estudio Piloto de los Procedimientos de Control de Calidad de Datos Oceanográficos y Meteorológicos Producidos por el Centro Control Contaminación del Pacífico” (L. A. Rodríguez & Ortiz Martínez, 2006); “Diseño e Implementación de una Herramienta Computacional para el Control de Calidad y Validación de Datos Oceanográficos” (Sánchez Escobar, 2006); “Desarrollo de un Sistema Electrónico para el Almacenamiento, Procesamiento y Control de Calidad de Datos en Tiempo Real Proveniente de Mareógrafos para el Centro de Alerta de Tsunamis” (García Torres, 2009) e “Implementación de Estándares Internacionales para el Control de Calidad de Datos Oceanográficos Provenientes de una Boya de Oleaje Direccional Triaxys” (Rúa, 2010).

En la Universidad Distrital Francisco José de Caldas existen aportes importantes como:

- El artículo “Propuesta Metodológica para la Homogenización de Series de Tiempo de Precipitación Mensual y su Utilidad en Procesos de Toma de Decisiones, Estudio de Caso Región Climatológica del Bajo Magdalena” (Bernal, Barrios, Ramos, Velásquez, & Ibarra, 2012), que puntualiza en la homogenización de las series de tiempo de precipitación enfocado hacia la disponibilidad de series con criterios de calidad que permitan avanzar en análisis climáticos de tipo regional, teniendo en cuenta la variabilidad climática interanual, eventos El Niño y La Niña vinculados con las anomalías de precipitación en éstos períodos de ocurrencia.
- El proyecto titulado “Aporte al Proceso de Toma de Decisiones en la Planificación Ambiental de la Región Climática del Bajo Magdalena con Base en Análisis Estadístico Empleando una Metodología para la Homogenización de Series Mensuales de Precipitación” (Barrios Moreno, 2013), como una propuesta metodológica para la homogenización de series mensuales de precipitación, en respuesta a la heterogeneidad de alguna series de precipitación atribuibles a factores no climáticos que afectan su calidad y utilidad en productos operativos y de investigación.

Estos avances se enmarcan dentro del desarrollo del Proyecto de Investigación titulado: “Homogenización de Series de Tiempo Mensuales de Precipitación y su Utilidad en Estudios Climáticos y Procesos de Toma de Decisiones” financiado por el Centro de Investigación y Desarrollo Científico (CIDC) de la Universidad Distrital Francisco José de Caldas desarrollado por los grupos de investigación GIIAUD de Ingeniería Ambiental e INDESOS de la Maestría en Desarrollo Sustentable y Gestión Ambiental (GIIAUD e INDESOS, 2011), y el apoyo técnico del IDEAM – Subdirección de Meteorología. También se ha consolidado también la “Guía Para el Procesamiento de Series de Tiempo de Precipitación y Temperatura: Estimación de Datos Faltantes, Detección de Cambios y Homogenización - Procesamiento De Datos en Excel-Spss- Tsw-Anclim” (Lombana González, Bernal Suárez, & Barrios Moreno, n.d.) la cual se encuentra en proceso de publicación.

En este mismo contexto de avances y que se tuvieron en cuenta en la Universidad Distrital para los trabajos referidos anteriormente, corresponden a los desarrollos de Martínez *et al.*, 1996; Nieto & Ruíz, 2002 y Bernal *et. al.*, 2011, estos tres aportes se lograron con el apoyo de la Subdirección de Meteorología del IDEAM.

Las metodologías y avances hechos por las investigaciones ya mencionadas aportan desde la estadística, la oceanografía, las tecnologías de la información, la ingeniería electrónica y la Ingeniería Ambiental para la calidad de datos y la evaluación de parámetros ambientales que permiten hacer más riguroso y confiable los procesos de toma de decisiones sobre los recursos naturales. Procesos que pueden adaptarse y aplicarse al recurso oceánico como lo plantea el desarrollo del presente proyecto de investigación.

4.2. Marco Geográfico

La **Cuenca Pacífica Colombiana (CPC)**, está situada en el extremo occidental del país entre los paralelos 01°30' hasta los 07°10' de latitud Norte y entre los 77°40' y 82°00' de longitud Oeste. Al Norte limita con las aguas del golfo de Panamá; al Este se haya el litoral occidental de Colombia, integrado por los departamentos de Chocó, Valle del Cauca, Cauca y Nariño; al Sur limita con las aguas de la costa ecuatoriana y la cordillera submarina de Carnegie (1350 km longitud, 300 km de ancho) y al Oeste se abre al océano Pacífico, las aguas territoriales de la República de Panamá (30 km) y la cordillera de Cocos (aproximadamente 200 km de ancho, 1000 km de longitud). Adicional a las fronteras políticas y las barreras geográficas terrestres, la Cuenca Pacífica Colombiana se encuentra enmarcada por dos cordilleras submarinas que se unen en la plataforma de Galápagos, constituyendo uno de los 20 puntos calientes del planeta y elevándose desde más de 3700 metros hasta 1000 metros de profundidad, lo que impide procesos de intercambio de aguas profundas. Por lo que se puede considerar que cuenta con fronteras tanto sólidas como líquidas (CCCP & DIMAR, 2002).

A nivel geológico la CPC se localiza en el margen occidental de la placa continental de Sudamérica de aproximadamente 200-300 km de espesor, la cual choca con la placa oceánica de Nazca (aproximadamente 50 km de espesor), que posteriormente se hunde ocasionando en la fosa Peruana-Chilena una zona de subducción de 200-300 km, a una velocidad media de 6 cm/año, se extiende a lo largo de las costas occidentales del continente suramericano. En esta área se encuentra la zona de fractura de Panamá, con la presencia de un centro litosférico en expansión con dos dorsales asísmicas: Malpelo y Coiba (Gansser, 1950 y Scientific American, 1976) (CCCP & DIMAR, 2002).

En la geomorfología de la CPC sus elementos más importantes son la isla de Gorgona (2°55' y 3°00' de latitud Norte y 78°09' y 78°14' de longitud Oeste), y Malpelo (3°51' de latitud Norte y 81°35' de longitud Oeste) ambas de origen volcánico.

Del Suroeste proviene el oleaje oceánico, el viento y la corriente predominante en la CPC. Los trenes de olas de amplio período y escasas alturas procedentes del océano abierto, van en

incremento de su altura sobre fondos someros y se refractan perdiendo su energía y cambiando la dirección. La alta actividad tectónica de la zona hace muy probable la ocurrencia de tsunamis o maremotos. La CPC cuenta con aproximadamente 80000 km² de superficie atribuidos a la región costera oriental del país (CCCP, 2002).

La oceanografía y climatología de esta región se conoce hace relativamente poco tiempo, gracias a algunas expediciones internacionales que produjeron los informes de Forsbergh (1969) y Wyrtsky (1965); y más recientemente por los informes de los cruceros oceanográficos de la Dimar, los cuales han presentado datos oceanográficos y meteorológicos de zonas alejadas a la costa y de regiones costeras, publicados por el Cccp (CCCP, 2002).

4.3. MARCO CONCEPTUAL

4.3.1. Monitoreo Ambiental

El entendimiento del mundo natural es una labor que ha constituido para los investigadores y naturalistas un gran reto. Aún hoy día se logra evidenciar los avances científicos y tecnológicos, sin embargo se requiere profundizar en el entendimiento de los procesos naturales. Desde la concepción de los fenómenos naturales como eventos aislados unos de otros hasta la visión holística del medio ambiente como un continuo de procesos físicos, químicos y biológicos que no pueden separarse fácilmente entre sí (Artiola, Pepper, & Brusseau, 2004) el estudio y observación del medio ambiente ha evolucionado a la vez que lo hace la sociedad humana. Por ello hoy día aún sin terminar de entender del todo la naturaleza de la que dependemos, las actividades antrópicas empiezan a modificarla de maneras que generan una gran incertidumbre para las generaciones presentes y futuras; y es en este punto, en el que monitoreo del medio ambiente toma tanta relevancia.

El monitoreo ambiental es el estudio y observación del medio ambiente (Artiola et al., 2004) y se compone de una serie temporal de mediciones de variables físicas, químicas y / o biológicas, diseñadas para responder preguntas sobre cambios ambientales (Lovett et al., 2007).

El monitoreo ofrece la posibilidad de hacer un seguimiento temporal y espacial de los cambios que ocurren en el medio ambiente, perceptibles a través de las variaciones en los parámetros objeto de observación como indicadores del estado de los elementos que lo componen. El monitoreo ambiental es una herramienta que ofrece múltiples beneficios, ya que permite: la protección de los suministros públicos de agua, la gestión adecuada de residuos, garantizar la calidad del aire, evaluar la degradación de la tierra y de los suelos, formular pronósticos climatológicos, hacer ordenamiento del territorio, adecuada asignación y explotación de recursos, diagnóstico de ecosistemas, caracterización de especies amenazadas y biodiversidad, seguimiento de cambios climáticos globales, entre otros (Artiola et al., 2004).

Entre los múltiples componentes de la atmósfera, la criósfera, la hidrósfera, la litósfera y la biósfera sería deseable tener un registro y seguimiento completo de todos sus elementos; sin embargo por cuestiones logísticas y principalmente económicas eso no es posible. Por ello, para priorizar los elementos bióticos y abióticos que han de ser monitoreados se diseñan e implementan los programas de monitoreo como un plan de acción que permite gestionar el seguimiento del estado de los recursos naturales.

El monitoreo ambiental usa una variedad de equipos y técnicas dependiendo del enfoque que sea requerido. Por ejemplo, el monitoreo de la calidad del agua superficial puede medirse utilizando instrumentos desplegados a distancia, o con instrumentos *in situ* en la mano o mediante la aplicación de biomonitoreo en la evaluación de la comunidad bentónica de macro invertebrados (Weston, 2011) tanto así pueden variar los enfoques y formas de diseñar un programa de monitoreo ambiental para un mismo parámetro, que son muchos los errores que se

pueden cometer en el proceso. El objetivo de un buen programa de monitoreo es aplicar los métodos y recursos necesarios de forma eficiente y efectiva para reunir una cantidad de datos representativos que permitan responder preguntas sobre los cambios ambientales de interés. Artiola *et al* (2004) asemeja un buen programa de monitoreo a un buen proceso de investigación, ya que posee una fuerte base en el método científico y este parte de una hipótesis o pregunta inicial hacia la cual van dirigidos todos los esfuerzos con tal de resolverla, sin olvidar que en esencia el monitoreo ha de tener un fin práctico.

Artiola *et al* (2004) describe a los programas de monitoreo altamente efectivos como aquellos que se (i) diseñan en torno a una pregunta que es científicamente clara y convincente (ii) incluyen un diseño que pueda ser revisado, retroalimentado y adaptado (iii) usan indicadores cuidadosamente seleccionados (iv) procuran por la calidad y consistencia de los datos (v) planifican a futuro la accesibilidad y transferencia de la información (vi) examinan e interpretan continuamente los datos e (vii) integran programas de investigación en el monitoreo.

Los programas de monitoreo están fuertemente determinados por la escala espacial y temporal que se les asigna al momento de su diseño. Cuando se pretende hacer seguimiento de los recursos naturales se debe tener presente que los patrones biogeográficos, ecológicos y sus procesos asociados, son usualmente dependientes de la escala y muchas veces están estructurados de una forma jerárquica (N. Rodríguez, Armenteras, Bernal Suárez, & Rincón, 2007). Muchas investigaciones corresponden a escalas locales, sin embargo, las necesidades actuales requieren responder preguntas a escalas regionales. Preguntas que requieren de una base sólida de información primaria para ser respondidas. No siempre se cuenta con datos de amplias zonas geográficas y de una larga duración de tiempo, en especial para períodos en los que el seguimiento de los recursos naturales no era la principal preocupación para las autoridades territoriales y académicas.

Para propósitos del desarrollo de este trabajo se considera la escala del monitoreo desde el estudio de los ecosistemas como la dimensión física de un objeto o proceso ecológico en el espacio y que se puede descomponer en tres dimensiones: ecológica, de muestreo y analítica. La escala ecológica expresa la dimensión real de los fenómenos ecológicos, la escala de muestreo hace referencia a la extensión del área de observación y sus características espaciales y la escala analítica que refleja las características de las unidades de muestreo en términos de análisis (N. Rodríguez et al., 2007). Así, se adaptan estos conceptos al estudio en cuestión donde la escala ecológica comprende la observación de parámetros oceanográficos, la escala de muestreo se traducen en una gran área geográfica que comprende la Cuenca Pacífica Colombiana (CPC) y la escala analítica se refleja como las unidades de medición de los instrumentos y métodos para asignar valores a los datos. Finalmente la escala temporal, que es a veces la que más limitantes supone; en este caso se cuenta con décadas de muestreos que atribuyen a los datos un amplio rango de tiempo que puede ser analizado.

En un principio los programas de monitoreo han sido, a menudo, impulsados para prevenir problemas de salud humana en lugar del deterioro ambiental. Por lo que hoy día existe una fuerte necesidad de integración entre los programas de monitoreo y las agencias públicas y privadas para reducir la redundancia y los costos (Lovett et al., 2007). Los datos recolectados originalmente para propósitos de salud pública pueden resultar útiles ahora para responder preguntas sobre el estado y variación de ciertos recursos naturales. Mediciones de varios años o incluso décadas se convierten en una fuente valiosa que proporciona un vistazo a las condiciones ambientales en tiempos pasados e incluso en áreas geográficas diferentes a las que son monitoreadas hoy día.

Representa un reto para los investigadores el hacer uso de estas mediciones que no fueron contempladas en un principio como parte de un programa de monitoreo e integrarlas a las bases de datos, tanto nacionales como internacionales, de manera tal que la valiosa información no se

pierda si no que contribuya al entendimiento y análisis de los recursos naturales que hoy día se pretenden conservar. Así es como los programas de monitoreo pueden integrar proyectos de investigación en sus esquemas, desde la arqueología de datos, pasando por los controles de calidad hasta el análisis final de los mismos. Asociando el monitoreo tanto conceptual como operativamente como un medio para medir la eficacia de las acciones de manejo de los ecosistemas a nivel global, regional y local (N. Rodríguez et al., 2007). Una base de datos primaria sólida, que brinde información confiable para tomar decisiones apropiadas relacionadas con el manejo del medio ambiente es una de las responsabilidades y fines prácticos del monitoreo ambiental, que permite adoptar decisiones y planificar acciones frente a los cambios ecológicos en etapas anticipadas a procesos de disminución de recursos naturales y llevar a una intervención antes que las pérdidas ambientales sean irreversibles (N. Rodríguez et al., 2007).

El monitoreo ambiental surge de la preocupación por la preservación de los recursos naturales, y el recurso oceánico no es la excepción, dada su escala global, el seguimiento de este recurso es un reto para cualquier país con territorio marítimo. Como la importancia del recurso oceánico lo demanda, existen iniciativas internacionales que procuran por el seguimiento del estado de los océanos en el mundo.

En Colombia en cuanto al seguimiento de estado de los océanos, los esfuerzos se multiplican por dos, al presentar dos cuencas oceánicas en su territorio. Para la Cuenca Pacífica Colombiana (CPC) los primeros estudios oceanográficos datan de 1941 con la expedición Askoy; ya para 1965 hay registros de los principales esfuerzos de la Armada Nacional por entender la dinámica de las diferentes variables oceanográficas y meteorológicas sobre la CPC; cuando la Comisión Interamericana del Atún Tropical (CIAT) adelantó la expedición oceanográfica The Augmented Colombian El Niño Tuna Oceanographic (Acento), a bordo del Buque Bocas de Ceniza, en la que se pretendía estudiar la variabilidad estacional de la circulación y distribución de propiedades físicas, químicas y biológicas en aguas del Pacífico colombiano (Forsbergh, 1969). Posteriormente, a partir de 1967, se desarrollaron los cruceros Eastropac, patrocinados también por la CIAT, como un esfuerzo cooperativo hacia el entendimiento de la oceanografía en el Pacífico este tropical.

Al inicio de la década de los ochenta la Dirección General Marítima (Dimar), con su Centro de Investigaciones Oceanográficas e Hidrográficas (CIOH), comenzó a planear y desarrollar sus propias expediciones oceanográficas tanto en el Mar Caribe como en las aguas jurisdiccionales del Océano Pacífico colombiano; tarea que posteriormente fue compartida con el Centro Control Contaminación del Pacífico (CCCP) desde 1984, año de su fundación. Desde entonces, se han completado 54 cruceros oceanográficos que han proporcionado información importante para la determinación de zonas homogéneas del Pacífico colombiano en la capa superficial, investigación sobre la distribución de una sustancia pasiva según el movimiento de corrientes superficiales de la CPC, la determinación de una primera aproximación a los cambios interanuales de temperatura y salinidad en la CPC, la determinación de cambios interanuales de la presión atmosférica sobre la CPC, la evolución de las zonas de surgencia de la CPC en la capa 0 – 100 metros y la variación del contenido de calor de la capa activa de la CPC (CCCP & DIMAR, 2016).

Desde 1972 los cruceros oceanográficos enviados al Pacífico ejercen labores de oceanografía descriptiva como etapa inicial para la elaboración de un Plan de Investigaciones Marinas para Colombia (DIMAR & CECOLDO, 1973), para la década de 1990 se especifican las labores de monitoreo de las condiciones oceanográficas y meteorológicas de la CPC (DIMAR, CECOLDO, & CIOH, 1993), a través de mediciones de parámetros físicos, químicos, biológicos y meteorológicos; los cuales poco a poco han contribuido a la caracterización océano-atmosférica de la CPC, para el desarrollo de diversas actividades marinas y proporcionando información para la investigación aplicada en el aprovechamiento de recursos o en el desarrollo de obras de ingeniería oceánica (DIMAR & CECOLDO, 1992).

4.3.1.1. Observación de los Océanos

4.3.1.1.1. Contexto Internacional

Sistema Mundial de Observación de los Océanos

El **Sistema Mundial de Observación de los Océanos** (GOOS por sus siglas en inglés) es un programa ejecutado por la Comisión Oceanográfica Intergubernamental (COI) de la UNESCO, que cuenta con la contribución coordinada de varias personas y organizaciones en todo el mundo. Este ha logrado coordinar un sistema colaborativo de observaciones sostenidas. El GOOS, con su condición única dentro del sistema de las Naciones Unidas, puede reunir los recursos de los Estados Miembros de la COI de la UNESCO para crear una red en torno a elementos de observación financiados por los gobiernos independientes (satélites, boyas, científicos, entre otros) como por ejemplo redes de observación, monitoreo de estado y normas y mejores prácticas.

Un ejemplo de los proyectos puestos en marcha por el GOOS es el **Sistema de Observación del Pacífico Tropical** (TPOS) que para 2020 pretende enfocarse en lograr una mayor efectividad para todas las partes interesadas, informando el desarrollo y los requisitos de los modelos de predicción operacional para los principales usuarios de los datos TPOS. El proyecto abarca la integración de diversas tecnologías de muestreo, con un enfoque deliberado en la robustez y la sostenibilidad, y que pretende proporcionar un legado de mejores mecanismos de gobernanza, coordinación y apoyo que contribuyan al GOOS. Particularmente el proyecto TPOS 2020 se financia y gestiona independientemente del GOOS, pero informa al Comité Directivo del GOOS para asegurar la integración de su legado con el GOOS. Es un modelo para otros proyectos de desarrollo que están ampliando el alcance de las ideas del Marco para la Observación del Océano y energizando y ampliando las capacidades del GOOS (COI & UNESCO, n.d.).

Programa Global de Perfil de Temperatura y Salinidad (GTSP)

El Programa Global de Perfil de Temperatura y Salinidad (GTSP, por sus siglas en inglés) es un programa internacional cooperativo para desarrollar y mantener el recurso global de Temperatura-Salinidad con datos actualizados y de la más alta calidad. Se trata de un programa conjunto de la Organización Meteorológica Mundial (OMM) y de la Comisión Oceanográfica Intergubernamental (COI).

El GTSP se desarrolló como una respuesta a las necesidades cambiantes de las comunidades operacionales marinas y de ciencia oceánica. Los científicos que realizan investigaciones sobre los procesos oceánicos y atmosféricos que están abordando problemas globales y necesitan conjuntos de datos multidisciplinarios para sus experimentos. Tanto las comunidades científicas como operacionales están aplicando el conocimiento de los procesos oceánicos obtenidos en cuatro décadas de investigación. Los datos se están asimilando en modelos que hacen predicciones que abordan el desarrollo sostenible, el cambio climático y la seguridad humana y ambiental.

Para satisfacer estas necesidades los programas de recopilación de datos son cada vez más complejos y producen grandes volúmenes de datos. El problema del control de calidad es más difícil debido al aumento del volumen de complejidad de los datos. Los medios mejorados de comunicaciones de datos de alta velocidad (NOAA, 2015b).

4.6.1.1.2. Contexto Nacional

Sistema de Medición de Parámetros Oceanográficos y de Meteorología Marina - SMPOMM

El SMPOMM es coordinado y mantenido por la Dimar; permite obtener en tiempo real datos relacionados con las condiciones oceánicas y atmosféricas del “territorio marítimo colombiano”, con el fin de ofrecer información necesaria para garantizar y salvaguardar la vida humana en las distintas actividades marítimas y costeras que se realizan en el país. Está interconectado con estaciones meteorológicas, boyas de oleaje direccional, boyas metocean y boya Tsunami, ubicadas en diferentes puntos marítimos nacionales cercanas a los Puertos y principales ciudades del litoral, que registran datos como: dirección y velocidad del viento, temperatura ambiente, humedad relativa, presión atmosférica, radiación solar y sobre el estado del mar: dirección y velocidad de corrientes hasta 50 m de profundidad; periodo, altura y dirección de la ola, temperatura superficial del mar, salinidad superficial (DIMAR, 2016).

Con la información obtenida por el SMPOMM, los Centros de Investigaciones CIOH del Caribe y Pacífico colombiano elaboran pronósticos del estado del tiempo y del mar y Alertas Tempranas por ocurrencia de eventos extremos de origen marino para la toma de decisiones por parte de las autoridades competentes (DIMAR, 2016).

El SMPOMM tiene a disposición un aplicativo web que permite acceso a la información sobre registros actuales dependiendo de la ubicación geográfica, tal como se aprecia en la Figura 2..



Figura 1. DIMAR (2016) Sistema de Medición de Parámetros Oceanográficos y de Meteorología Marina - SMPOMM. Fuente: DIMAR (2016).

Sistema de Información de la Red de Vigilancia de la Calidad Ambiental Marina de Colombia (Redcam)

REDCAM es un programa nacional e interinstitucional integrado por las Entidades encargadas de la gestión, vigilancia, control e investigación de la calidad ambiental en los espacios marinos y costeros de Colombia trabajando en conjunto desde el año 2001 y se constituye en un instrumento para la gestión del recurso hídrico marino y costero del País (INVEMAR, 2017).

Es una red a cargo del INVEMAR enfocada a la gestión y toma de decisiones del Ministerios de Ambiente y Desarrollo Sostenible y de las Corporaciones Autónomas Regionales (CAR) con injerencia en las zonas costeras, y, en general, del SINA en temas de contaminación marina-costera. Así, el MADS, el Centro de Investigaciones Oceanográficas e Hidrográficas del Caribe (CIOH) de la Dimar, las doce CAR costeras y el Establecimiento Público Ambiental (EPA) de Cartagena son nodos de la REDCAM que desde 2001 aportan datos al sistema de monitoreo adicional a la información proimaria colectada durante los monitoreos de la REDCAM liderados por el INVEMAR (CECOLDO & DIMAR, 2016).

El programa nacional REDCAM comprende una serie de actividades técnicas, académicas y administrativas como monitoreo, capacitación, fortalecimiento, asesoría, actualización del sistema de información y gestión de convenios para la ejecución de las actividades conjuntas entre el INVEMAR y otros nodos del programa nacional, con el propósito de contribuir con información básica sobre temas relacionados con la calidad ambiental marina colombiana, y mantener en funcionamiento la REDCAM, como una herramienta que soporte la gestión ambiental y toma de decisiones por parte de las autoridades competentes, que posibilite el manejo integrado y aprovechamiento sostenible del recurso hídrico marino y la biota asociada (INVEMAR, 2016).

4.3.2. Variables oceanográficas y Correlaciones

4.3.2.1. Ciclos biogeoquímicos

Los Ciclos Biogeoquímicos describen el movimiento cíclico de los elementos que forman los organismos biológicos (bio) y el ambiente geológico (geo) e interviene un cambio químico. Estos permiten que los elementos se encuentren disponibles para ser usados una y otra vez por otros organismos. Son procesos naturales que reciclan elementos en diferentes formas químicas desde el medio ambiente hacia los organismos, y luego a la inversa. Agua, carbono, oxígeno, nitrógeno, fósforo y otros elementos recorren estos ciclos, conectando los componentes vivos y no vivos de la Tierra (Centro Internacional para la Investigación del Fenómeno de El Niño, 2016).

Cada elemento posee su propio ciclo y dependiendo de las reacciones químicas y procesos que intervienen en los diferentes niveles pueden unos ser más complejos que otros. Aunque aparentemente los ciclos parecen operar de forma independiente, los factores ambientales intervienen significativamente en los procesos de cambio. El entendimiento del medio ambiente como un todo requiere de la comprensión del sistema natural como un todo interconectado, donde los desequilibrios en una parte del sistema, ya sean de causas naturales o antrópicas, provocan cambios en otras para compensar los flujos de materia y energía alterados (Minster, 2004).

El océano en su composición aparentemente sencilla en comparación con la terrestre funciona como una enorme máquina térmica; cuando el Sol lo recalienta en las zonas tropicales el calor queda almacenado en el agua y acaba por restituirse a la atmósfera en las latitudes más altas, provocando su enriamiento. De ese modo se generan las corriente oceánicas, responsables de transportar el calor desde el ecuador hacia los polos, y al mismo tiempo transporta las sustancias disueltas en el agua en forma de sales (Minster, 2004).

La dinámica de los océanos es una ciencia compleja que estudia múltiples factores en la física de los movimientos oceánicos, esta nos permiten entender los cambios que ocurren en el perfil de la columna de agua. Sin embargo para propósitos del este trabajo se hace énfasis en los cambios que ocurren a nivel superficial, es decir donde el océano está en permanente contacto

con la atmósfera. Se trata de un sistema océano-atmosférico, en donde los cambios en uno provocan reacciones en el otro.

El océano intercambia energía con la atmósfera en forma de calor y en forma mecánica por obra del arrastre del viento. En relación con las mareas, también hay que tomar en cuenta las fuerzas gravitacionales. A nivel superficial los intercambios de calor con atmósfera se llevan a cabo, en primer lugar, mediante radiación infrarroja. También ocurren intercambios en forma de calor latente cuando la evaporación del agua absorbe calor del océano y lo libera en la atmósfera cuando tiene lugar la condensación del vapor de agua (Minster, 2004). La temperatura superficial del mar varía en solo unos pocos grados entre estaciones y latitudes, sin embargo hay que considerar que cuando una masa tan grande de agua gana o pierde energía se traducen en cambios significativos en el equilibrio energético de la atmósfera.

El océano absorbe más calor en el ecuador que en los polos. La distribución detallada de las temperaturas revela lo complejas que son. A lo largo del ecuador se advierte una lengua de agua más fría al este de los océanos Atlántico y Pacífico. Paralelamente la temperatura de las aguas oceánicas disminuye notablemente en razón directa con la profundidad. El agua profunda es fría y relativamente homogénea. Al realizar un corte en dirección Norte-Sur que revelará las variaciones de temperatura se observarían aguas calientes centradas en las latitudes tropicales en superficie, mientras que las capas frías profundas se extienden con continuidad desde las aguas de la superficie de las latitudes más altas (Minster, 2004).

Seguido a la temperatura, la salinidad es un parámetro fundamental para la caracterización de las aguas marinas; que presentan variaciones aparentemente ligeras, pero de gran importancia. El 50% del agua marina posee una salinidad comprendida entre 34,6 y 34,7‰ (36,6 miligramos de sal por cada gramo de agua). En general en superficie se encuentran aguas más saladas en las zonas tropicales (hasta 35‰) y aguas menos saladas en las altas latitudes (aprox. 33‰). Aunque la salinidad no es objeto de estudio en la presente investigación si es importante entender su papel en la dinámica del océano para análisis posteriores (Minster, 2004).

Finalmente la densidad, es un parámetro que cambia con la salinidad y la temperatura, por lo general se mantiene en un promedio de 1,028 g/ml. Así las aguas más densas se encuentran en el fondo y las menos densas en la superficie. La temperatura, salinidad y densidad son características claves y únicas de cada masa de agua alrededor del mundo, ya que estas se consideran propiedades conservativas porque, una vez que la masa de agua se separa del contacto con la atmósfera, los valores adquiridos sólo se modifican mediante la mezcla con otras masas de agua de características diferentes (Instituto de Tecnologías Educativas, n.d.).

Las principales variables físicas (temperatura, salinidad y densidad) interactúan en el mar junto con parámetros químicos con la concentración de nutrientes, estas pueden cambiar con respecto al tiempo y al espacio donde se presentan. Conocer la periodicidad y extensión de estas modificaciones en el océano, es una labor que permite a los oceanógrafos pronosticar el estado del mar en forma semejante a los pronósticos meteorológicos. Pero ello es necesario repetir durante años, meses, días y horas el registro de datos en la misma zona, mientras que en el caso de cambios en el espacio, es deseable obtener observaciones simultáneas en diferentes zonas del océano, necesitando el uso de 2 o más barcos o unidades de registro. (Cifuentes Lemus, Torres arcía, & Frías M., 1997a), lo cual en conjunto hace parte de las labores de monitoreo de los océanos.

4.3.2.2. Temperatura Superficial del Mar (TSM)

La TSM, es una variable que suministra información sobre la cantidad de energía disponible en la capa superior del océano que está en contacto con la atmósfera. A través de esta variable, es

posible inferir otros procesos físicos, químicos y biológicos debido a que influye en el comportamiento de gran parte de ellos. (IDEAM & Melo Franco, 2005)

A través de la información que proporciona la temperatura superficial del mar es posible realizar seguimiento a procesos de interacción océano-atmósfera; es altamente significativa en los componentes de solubilidad de sales y gases y afecta de manera indirecta los valores de conductividad y pH; sirve para predecir la intensificación de procesos ciclónicos, condiciones de niebla, fenómenos aire-mar, desarrollo de tormentas, formación de hielos marinos; y sirve para explicar de manera directa o indirecta, la distribución de las distintas especies de peces, debido a la estrecha relación entre la temperatura de la superficie del mar (TSM) y la actividad biológica. (IDEAM & Melo Franco, 2005)

Para considerar confiables las mediciones realizadas en campo, en el caso de la TSM que ha de observarse, debe ser representativa de las condiciones que existen en la capa de la primera profundidad óptica por debajo de la superficie libre del océano. La unidad de medida está dada en grados Celsius °C. En la Tabla 1 se aprecia el comportamiento mensual de la TSM esperado para la CPC.

Comportamiento de la TSM en la Cuenca Pacífica Colombiana

Tabla I. Descripción de los datos promedio de temperatura en la Cuenca Pacífica Colombiana (CCCP & DIMAR, 2002)

Descripción de los datos promedio de temperatura en la Cuenca Pacífica Colombiana	
Enero	No existen muchas mediciones para el mes de enero que permita identificar un patrón en el comportamiento de la TSM en la CPC, sin embargo la información presente en el Atlas de los Datos Oceanográficos de Colombia 1922-2013 (Andrade Amaya, Rangel Parra, & Herrera Vásquez, 2015) permite estimar un promedio de 27, 01 °C
Febrero	La temperatura promedio en condiciones homogéneas es de 26.7°C para la costa y 26.9°C para aguas oceánicas, alteradas por pequeños núcleos de altas temperaturas alrededor de los 28° C en algunos puntos de la costa y la región central. Se observa en el límite oeste del área de estudio, sobre el meridiano 82°W, entre latitudes 3 y 5°N, condiciones cálidas que no alcanzan a llegar a los 81°30'W.
Marzo	La temperatura promedio de marzo se caracteriza por presentar en forma generalizada valores bajos, 26.2°C para la costa y 26°C para la sección oceánica. Existe en un pequeño sector con aguas cálidas cercanas a la costa frente a la bahía de Buenaventura con temperaturas de 27.9°C.
Abril	Condiciones de temperatura alteradas por núcleos de aguas cálidas ubicados, indistintamente sobre la CPC, con valores entre los 28 y 29° C. La temperatura promedio de la región oceánica es de 26.6°C y para la costa de 26.8°C, presentándose en la bahía de Buenaventura uno de los núcleos cálidos más grandes.
Mayo	Las condiciones poco definidas por las regiones costera y oceánica. Se presentan temperaturas con promedios superiores a los tres meses anteriores 27.4°C para la costa y 27.2°C para las aguas oceánicas.
Junio	La temperatura en junio presenta valores menores al sur de la CPC; sobre el centro se combinan las temperaturas bajas con las altas y hacia el norte predominan las temperaturas altas. El rango de valores que cubre toda la cuenca se ubica entre los 26 y 29°C, con temperaturas promedio de 27.5° C.

Julio	No existen muchas mediciones para el mes de enero que permita identificar un patrón en el comportamiento de la TSM en la CPC, sin embargo la información presente en el Atlas de los Datos Oceanográficos de Colombia 1922-2013 (Andrade Amaya et al., 2015) permite estimar un promedio de 27, 12 °C
Agosto	La temperatura desciende en forma generalizada para la CPC, quedando solamente pequeñas áreas con temperaturas superiores a 27.5°C ubicadas en la región norte y central. Los valores de TSM de la cuenca se ubican entre 26 y 28° C. El promedio de temperatura para la zona costera es de 27.3°C, mientras que en el océano fue similar a febrero con valores de 26.9°C.
Septiembre	Las condiciones promedio de temperatura muestran un proceso de enfriamiento de las aguas en la CPC a medida que finaliza el año. El promedio para la costa es de 27°C y en el sector oceánico de 26.5°C. Los valores máximos se dan en la parte norte de la cuenca entre 5 y 7° de latitud Norte.
Octubre	Las condiciones muestran una continuación en el descenso de la temperatura en comparación con el mes anterior; presentándose pocas áreas cálidas que se ubican hacia el extremo norte de la zona costera y una pequeña región central entre los 79° y 80° W. Los promedios de temperatura son de 27°C para la costa y de 26.4°C para las aguas oceánicas.
Noviembre	La temperatura en este mes presenta condiciones similares al anterior, con un área mayor en la que se presentan temperaturas entre 27.5 y 28°C. La temperatura promedio para la costa es de 27.3°C y para el sector costero de 26.9°C, evidenciándose un leve calentamiento de la CPC.
Diciembre	La temperatura alcanza registros superiores a los promedios que se presentan en los tres meses anteriores, siendo este comportamiento generalizado para toda la cuenca, registrando valores promedio de 27.7° C

4.3.2.3. Oxígeno Disuelto

El oxígeno es un elemento básico de comportamiento cíclico en el océano, se encuentra como componente de la atmósfera y llega al mar al disolverse en el agua. (OD). A través de la evaluación de oxígeno disuelto, es posible identificar y reconocer la distribución de los organismos en los océanos, para los estudios de descomposición orgánica y productividad primaria de los mismos. Como parámetro oceanográfico se le denomina comúnmente como Oxígeno Disuelto (IDEAM & Melo Franco, 2005).

El **ciclo del oxígeno** a nivel oceánico comienza como un componente de la atmósfera y llega al mar al disolverse en el agua, de donde lo toman los vegetales y los animales para su respiración. Con las corrientes y el oleaje la cantidad de oxígeno en el agua aumenta, pero también contribuyen a incrementar su cantidad los vegetales verdes, que durante el proceso de fotosíntesis, fijan el carbono y desprenden el oxígeno como resultado de las reacciones químicas que efectúan, completándose el ciclo del oxígeno en el mar (Figura 3). La cantidad de oxígeno disuelto en el agua del mar es inversamente proporcional a la profundidad, encontrándose en los grandes fondos marinos áreas carentes de este gas. (Cifuentes Lemus et al., 1997a)

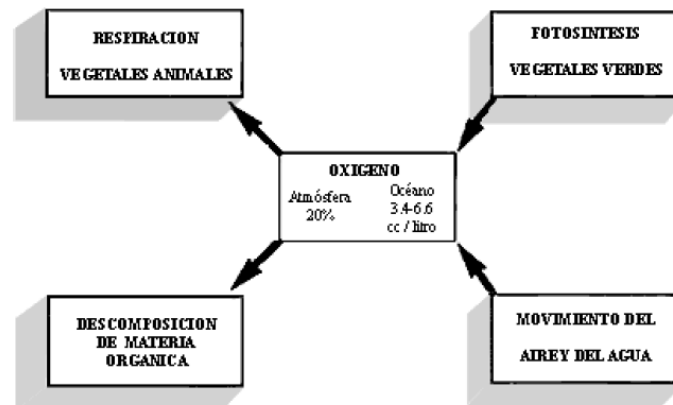


Figura 2. Ciclo del oxígeno. Fuente: Cifuentes Lemus, Torres García, & Frías M (1997b)

En cuanto al **comportamiento del oxígeno disuelto** en la cuenca Pacífica colombiana a nivel superficial presenta en el sector sur la entrada de agua con valores de 5.8 a 6.2 mg/l, entre las latitudes 82° W y 80° W, y se mantiene hasta los 50 m oscilando entre 3.5 y 5.5 mg/l de OD.

Sobre la parte central de la cuenca y longitudinalmente se observa la formación de dos núcleos con valores que fluctúan entre 6 y 6.8 mg/l, que permanecen elevados con casi 6.0 mg/l hasta el nivel de los 50 m. Mientras que en la parte noroeste se presenta una entrada con valores menores de 5.4 a 6.0 mg/l, obteniéndose los registros más bajos de OD en esta zona a medida que se profundizó en la columna de agua hasta llegar a 1.4 mg/l, a partir de los 100 m.

A lo largo de la costa se presentaron valores de 6.1 a 6.3 mg/l en la superficie, donde los mayores registros desde 0 hasta 100 m se observan en las zonas de Tumaco y Buenaventura. (CCCP & DIMAR, 2002).

4.3.2.4. Nutrientes

En el mar existe un conjunto de elementos denominados nutrientes que intervienen de manera básica en la vida de los océanos. Estos nutrientes son indispensables para la producción de alimento hecha por los vegetales. Están compuestos a base de nitrógeno, fósforo, sílice, manganeso, cobre y hierro; en su ciclo intervienen una serie de procesos químicos y biológicos, que determinan su concentración en las diferentes capas del agua del mar (Cifuentes Lemus et al., 1997b).

Los nutrientes se encuentran fuertemente relacionados con los organismos vivos que habitan en los océanos. Cuando estos expulsan los desechos de su metabolismo o mueren, la materia orgánica queda en las capas superficiales y se hunde por gravedad hasta las profundidades donde, por acción de las bacterias y de la actividad química, se descompone, liberándose así los nutrientes, los cuales pueden ser utilizados nuevamente por otros vegetales para elaborar nuevas sustancias orgánicas por el proceso de fotosíntesis y pasar a los animales. Las corrientes de aguas ascendentes provocan que los nutrientes salgan nuevamente a la superficie por cambios en la temperatura de las masas de agua, en zonas denominadas de surgencia o afloramiento, al estar nuevamente en superficie y con disponibilidad de radiación solar el ciclo se repite (Cifuentes Lemus et al., 1997b).

Para todos los nutrientes la circulación del agua interviene en la concentración de estos en las diferentes capas del océano y se puede considerar que los nutrientes se encuentran en los primeros 1 000 metros de profundidad y la degradación de la materia orgánica puede ocurrir en toda esta zona; mientras que la fotosíntesis se realiza en los 100 metros de profundidad oceánica en la llamada zona eufótica, que es aquella en que penetra la luz del Sol. De todos los nutrientes

el principal elemento es el **nitrógeno**, cuyo ciclo en el mar es complejo (Cifuentes Lemus et al., 1997b).

El nitrógeno gaseoso en los océanos es aproximadamente 30 veces más abundante que la suma de sus formas inorgánicas (amonio, nitrito, nitrato). Sin embargo, el nitrógeno molecular es relativamente inerte y para ser utilizado por los organismos debe estar en formas disponibles (NH_4^+ , NO_2^- , NO_3^-), los nitratos representan la forma más oxidada del nitrógeno inorgánico y los nitritos son las sustancias intermedias que se presentan durante el proceso de oxidación del amonio a nitratos. Los niveles de amonio ($\text{NH}_3 + \text{NH}_4^+$) se deben a la actividad biológica, principalmente (INVEMAR, 2003). Como se muestra en la Figura 4.

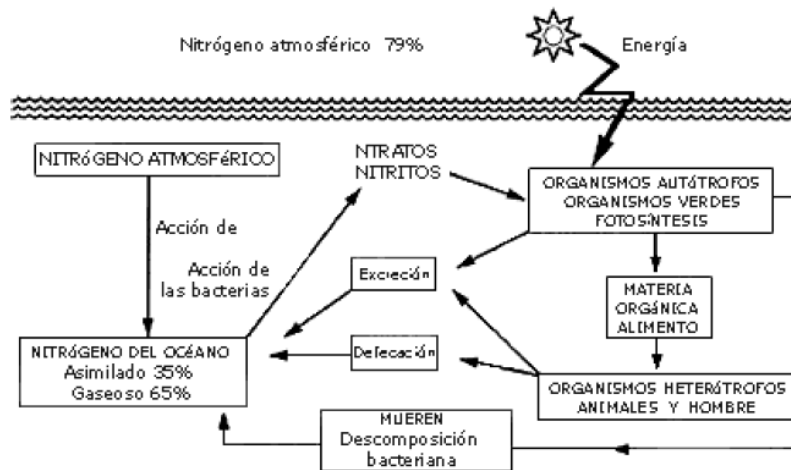


Figura 3. Ciclo del nitrógeno en los ambientes marinos. Fuente: Cifuentes Lemus et al, (1997b)

Los procesos de transformación química que intervienen en el ciclo de nitrógeno (Figura 5) son: (1) fijación (reducción) de nitrógeno molecular a amoniac, (2) asimilación de amoniac, (3) nitrificación, (4) denitrificación, (5) reducción asimilativa de nitrato y (6) amonificación.(Fuentes & Massol-Deyá, 2002b).

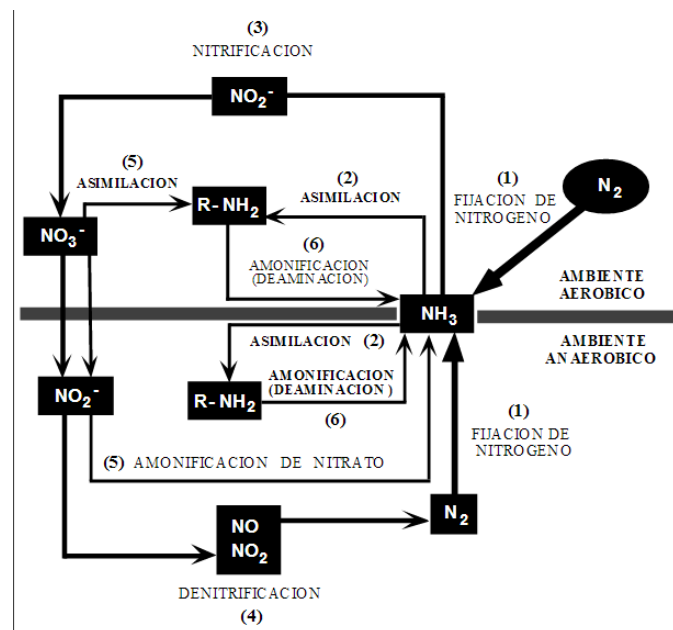


Figura 4. Las cinco etapas del ciclo del nitrógeno. Fuente: Fuentes & Massol-Deyá, (2002)

La **provisión y distribución** del total de nitrógeno se lleva a cabo por los aportes continentales, por las corrientes oceánicas y por las excreciones y muerte de los vegetales y animales marinos. El nitrógeno llega en forma de compuestos nitrogenados no asimilables, pero gracias a la acción de las bacterias y de las algas verde-azules, durante el proceso llamado *nitrificación*, se transforman en nitrógeno soluble, que es fijado por los vegetales verdes para formar nuevamente materia orgánica, la que es aprovechada por ellos o por los animales cuando se los comen, repitiéndose el ciclo (Cifuentes Lemus et al., 1997a).

Para el entendimiento del ciclo del nitrógeno en los océanos, se requiere del conocimiento de sus tres formas disponibles en el agua. Los niveles de amonio en el agua marina se atribuyen principalmente a las excreciones de animales marinos y la descomposición de compuestos orgánicos nitrogenados, provenientes a su vez de organismos muertos. Diversos seres fitoplanctónicos utilizan el amonio y lo convierten nuevamente en compuestos orgánicos nitrogenados, también puede ser oxidado por acción química, fotoquímica o bacteriana a nitrito y luego a nitrato. Posteriormente los nitritos representan una forma intermedia en el ciclo del nitrógeno; pueden estar presentes en las aguas como resultado de la degradación biológica de las proteínas o provenir de otras fuentes. Mientras que la mayor parte del nitrógeno en el mar se halla en forma de iones nitrato, su concentración normalmente varía entre 25 - 500 μg de $\text{N-NO}_3/\text{l}$ (INVEMAR, 2003).

En cuanto al **comportamiento de los compuestos nitrogenado** para la Cuenca Pacífica Colombiana (CPC) ha sido medido en diferentes cruceros oceanográficos con más atención en la última década. Muchos estudios arrojan información sobre el comportamiento registrado por el nitrógeno inorgánico total (NIT), corresponde a la suma de las concentraciones de amonio, nitritos y nitratos; y se asocia estrechamente al comportamiento de dichos constituyentes.

En general para la CPC el **amonio** en superficie muestra concentraciones del orden de los 0 a 0.95 g-at/l hasta alcanzar los registros más altos hacia los 150 m con valores entre 2.0 y 9.1 g-at/l. Es claro observar que las mayores concentraciones se presentan a nivel costero y hacia el noroeste de la CPC, en la zona correspondiente a la isla de Malpelo, ampliándose estos dos núcleos a medida que se van profundizando en la columna de agua, sobre todo a partir de los 100 m (CCCP, 2002).

Los **nitratos** en la CPC a nivel superficial se presentan valores bajos sobre todo hacia la zona costera con registros de 0.02 a 0.21 g-at/l, al sur se presenta una entrada de agua caracterizada por valores más elevados con 2.97 a 3.2 g-at/l. Hacia los 25 m se observan más núcleos con registros mayores en un rango de 0.43 a 5.0 g-at/l, presentándose una entrada al norte y manteniéndose las concentraciones bajas hacia la costa, a excepción de un núcleo que se desarrolló cerca de Tumaco. Entre los 50 y los 150 m las concentraciones van aumentando hasta obtenerse los máximos registros hacia el norte y sur de la cuenca con valores entre 6 y 21 g-at/l (CCCP, 2002).

La identificación de **nitritos** puede en ocasiones resultar más difícil debido a las concentraciones bajas de este nutriente que obedecen a la inestabilidad química del ion asociada a la oxidación rápida a nitrato; conllevando a que en varias ocasiones éste se encuentre por debajo del límite de detección (0,02 μg at N-(NO₂)-/L) de los métodos instrumentales. Por lo general presenta un comportamiento análogo al del amonio y nitratos, con concentraciones promedio de 0,1 μg at N-(NO₂)-/L entre la superficie y los 30 m; y de $0,38 \pm 0,51$ μg at N-(NO₂)-/L ($\mu\text{g}/\text{l}$) y $0,46 \pm 0,45$ μg at N-(NO₂)-/L ($\mu\text{g}/\text{l}$) para los 20 y 30 m, correspondientemente (Cassanova Rosero et al., 2012).

4.3.2.5. Instrumentación Oceanográfica

4.3.2.5.1. Perfilador de Temperatura, Salinidad y Oxígeno (CTD)

El perfilador CTD denominado por sus siglas en inglés *Conductivity- Temperature- Depth* es un equipo que al ser sumergido, almacena en una memoria de estado sólido, datos de diferentes parámetros en medio ambientes marinos. La conductividad, la temperatura y la presión hidrostática son los sensores básicos, y permite incorporar sensores externos para medir pH, oxígeno disuelto, fluorescencia, entre otros. Posee un casco de titanio que resiste profundidades superiores a 1000 m y un casco plástico para profundidades máximas de 600 m. Su peso es de 13.7 Kg aproximadamente (CPPS, 2015b) Las innovaciones tecnológicas han provocado que en los cruceros oceanográficos para la tomas de muestras los métodos análogos sean reemplazados por el CTD, sin embargo en varios casos se efectúan mediciones de parámetros con ambos instrumentos. Los utilizados en los cruceros oceanográficos de Dimar por Sea-Bird (Figura 6).



Figura 5. Perfilador de ejemplo de Conductividad, Temperatura y Presión (CTD) SEA-BIRD. Modelo SBE 19 plus V2. 600 m. Fuente: SEA-BIRD. Fuente: CPPS (2015)

La información registrada por el CDT se guarda y respaldada en una computadora. Los datos obtenidos son procesados con base en una serie de subrutinas, dependiendo del modelo utilizado (CPPS, 2015b).

Al CTD se le puede acoplar un sensor óptico de oxígeno disuelto para la medición de Oxígeno con un rango de 120% de saturación superficial de 120% en todas las aguas naturales ya sean dulces o saladas, y genera mediciones con una precisión de 0.2 $\mu\text{mol/kg}$ (CPPS, 2015b).

4.3.2.5.2. Otros Instrumentos y Métodos

Para la medición de TSM, también se utilizan balde de lona doble o un balde de plástico o un balde de goma con termómetro fijo; termómetro invertido o termómetro de cazoleta; sonda multiparámetro Horiba U10 o pH-metro MP 120. (IDEAM & Melo Franco, 2005).

El procedimiento de medición *in situ* de la TSM se realiza con un balde de lona doble, el cual debe ser lanzado lo más lejos posible del buque para evitar contacto con el agua que rodea el casco de navío y la cual puede haber sido calentada por el mismo buque y, nunca debe ser arrastrado sobre la superficie del agua, ya que puede presentar una mezcla de temperatura del

agua de mar y temperatura del aire. Este balde debe permanecer por lo menos un minuto bajo el agua. Al retirarlo, se debe sumergir inmediatamente en el balde un termómetro invertido o de cazoleta (Figura 7). Esta operación debe hacerse a la sombra, ya que la radiación directa del sol puede calentar el agua rápidamente y ejercer influencia en el termómetro. Todo el termómetro debe estar recubierto de agua y no debe tocar los lados o el fondo del balde. Debe ser agitado en el agua teniendo en cuenta que el líquido no se caliente con el contacto del observador. Después de una inmersión de un minuto, el termómetro debe ser parcialmente retirado con el objeto de leer la temperatura con una precisión del 0.1°C , cuando todavía el termómetro del depósito este sumergido y la columna de mercurio se haya estabilizado (IDEAM & Melo Franco, 2005).



Figura 6. Termómetro de cazoleta. Fuente: CIOHP & Dirección de Climatología, (2002)

Cuando se utiliza una sonda multiparámetro Horiba U10, se sumerge directamente al cuerpo de agua a una profundidad de 30 cm aproximadamente, en una zona de poca turbulencia y se procede a la lectura; en el caso de que no se pueda debido a la fuerte turbulencia que se esté presentando en el momento de realizar la muestra, se purga el muestreado, se toma una muestra e inmediatamente se mide la variable con la sonda, desechando posteriormente dicha muestra. La precisión con este mecanismo es de aproximadamente 0.3°C (IDEAM & Melo Franco, 2005).

En el caso de la utilización de un pH-metro MP 120, se toma la muestra ya sea en el balde de lona doble o en un balde plástico, se alista el electrodo quitándole el protector y se introduce en el recipiente que tiene la muestra; obviamente teniendo cuidado de no tocar las paredes del recipiente que la contiene, y se espera hasta que la lectura se estabilice (IDEAM & Melo Franco, 2005).

Existen varios métodos analíticos (volumétrico, gasométrico y la cromatografía) para lograr la estimación del oxígeno disuelto en el cuerpo de agua. El método comúnmente utilizado en los cruceros oceanográficos colombianos, es el método volumétrico de Winkler (1888) de muestras obtenidas mediante botella Niskin (Figura 8) y Nasen. Este consiste sencillamente en una muestra que se hace reaccionar con una solución de iones manganosos y una solución yoduro-alcalina, la cual lleva incorporado ácido de sodio, cuya función es la de eliminar interferencias debidas a iones oxidantes como nitritos y materia orgánica presente; al mismo tiempo que se le protege del aire para evitar la oxigenación. En aguas costeras con alto contenido de materia orgánica, se usa ácido de sodio. El método es aplicable a todo tipo de aguas naturales, residuales e industriales especialmente en agua de mar. La exactitud del método, es determinada por la estequiometría de la secuencia de reacciones, así como por las exactitudes de cada uno de los tratamientos y manipulaciones de las muestras. Así mismo, la calibración de los aparatos volumétricos utilizados, es esencial para alcanzar una alta calibración (IDEAM & Melo Franco, 2005).



Figura 7. Botella Niskin. Fuente: CIOHP & Dirección de Climatología, (2002)

La toma de la muestra debe ser realizada antes que otras muestras de la botella muestreadora (Nansen o Niskin), usando una manguera de caucho y evitando introducir burbujas de aire. La botella Winkler en la cual se colecta la muestra, debe enjuagarse por lo menos dos veces con la misma agua. La muestra se trasiega lentamente, introduciendo la manguera de caucho hasta el fondo de la botella y dejando que la muestra reboce. Luego la manguera se saca lentamente. Cuando se trata de muestras de superficie tomadas con un balde, se sumerge la botella Winkler en el balde, sosteniéndola de forma oblicua, dejando que el agua escurra suavemente por las paredes y evitando las turbulencias. Posteriormente las muestras se almacenan en un lugar fresco y preferiblemente oscuro procediéndose a su traslado al laboratorio para el análisis respectivo, el cual debe realizarse en un tiempo no mayor a doce horas. Si se emplean matraces con un embudo, que permiten un sello de agua entorno al tapón. Además las precauciones mencionadas, es posible almacenar la muestra por varios días. Los errores observados son menores que 0.5% La exactitud numérica expresada por la desviación estándar es de ± 0.02 ml/l para concentraciones de oxígeno menores que 2 ml/l o de ± 0.04 ml/l para concentraciones de oxígeno sobre 2 ml/l (IDEAM & Melo Franco, 2005).

Para la determinación de parámetros químicos como el nitrógeno, en el caso de los cruceros oceanográficos colombianos, se recoge las muestras a través de una roseta marca Sea-Bird de 12 botellas tipo Niskin (Figura 8). La botella Niskin está diseñada para usarse en un cable (guaya del winche). En un "lance oceanográfico", varias botellas se unen a la guaya a distancias pre-determinadas y se bajan al mar, abiertas en ambos extremos, tal que el agua entra y sale libremente. Cuando alcanzan la profundidad deseada, un peso metálico ("mensajero") se deja caer deslizándose por la el cable para accionar el mecanismo que cierra la primera botella. El mismo mensajero viaja hacia abajo a lo largo de la guaya para accionar el cierre de la segunda botella, y así sucesivamente, hasta alcanzar la última (CCCP & Dirección de Climatología, 2002).

Posteriormente las muestras son llevadas a laboratorio. Se recomienda efectuar el análisis inmediatamente después de realizada la colección pueden ser en laboratorios a bordo de los cruceros, pero cuando las muestras requieren ser transportadas hasta los centro de investigación, se deben almacenar en un sitio oscuro y congeladas a -20° C. Es preferible el uso de congelación instantánea con CO₂. Antes de proceder al análisis es necesario filtrar la muestra para evitar las interferencias por el material suspendido, luego se aplican las pruebas estequiométricas pertinentes cada parámetro (INVEMAR, 2003).

Para la determinación de **amonio**, existen diversos métodos, sin embargo es más utilizado es la prueba de *azul de indofenol* propuesto por Riley (1953) y modificado por Strickland & Parsons (1968 - 1972), el método es específico para ion amonio y aplicable a todo tipo de aguas naturales. La mínima cantidad de amonio detectable en celda de 10 cm es de 0.1 $\mu\text{g.at.N-NH}_4/\text{l}$.

Para los **nitritos** la prueba se basa en la clásica reacción de Griess (1879) modificada por Llosvay (1889) El método es aplicable a todo tipo de aguas, especialmente de mar, pero puede

presentar problemas con aguas coloreadas. La reacción colorimétrica es específica para iones nitrito, sin embargo, pueden causar interferencias los iones Cu^{2+} en concentraciones mayores de 0.5 mg/l y los iones sulfuro en concentraciones superiores a 60 g de S=l. La mínima cantidad de nitrito detectable por este método, usando celda de 10 cm es de 0.01 $\mu\text{g.at. N NO-2/l}$.

Para los **nitratos** existen varios métodos para evaluación de estos iones en agua de mar los cuales se basan: (a) En la reducción a iones nitrito, y la subsecuente evaluación de éstos por métodos colorimétricos. (b) En la reacción colorimétrica como resultado de las propiedades oxidantes del ácido sulfúrico. (c) En determinación polarográfica; y (d) por espectrometría ultravioleta. El método de reducción a iones nitrito es el más utilizados y aplicable a todo tipo de aguas, la eficiencia de reducción está alrededor del 95%. Los cambios de temperatura entre 10 y 35°C no tienen efecto sobre la reducción. La cantidad mínima detectable usando celdas de 10 cm es de 0.05 $\mu\text{g.at. N-NO}_3/\text{l}$ ($\mu\text{g/l}$) (INVEMAR, 2003).

4.3.3. Fenómeno El Niño- La Niña

“El Niño” es el término utilizado desde finales del siglo pasado por los pescadores de las costas de Perú y Ecuador para describir las corrientes cálidas que recorrían sus lugares habituales de pesca durante la época de Navidad, correspondientes al verano en el hemisferio sur. Debido a que cada año una débil corriente oceánica cálida procedente del sur se desarrolla a lo largo de las costas de Ecuador y Perú. Sin embargo, en períodos no definidos, la temperatura superficial de esta corriente es mucho mayor a lo normal. La corriente de El Niño de carácter impredecible al no tener un ciclo definido de aparición, cuya presencia obligaba a disminuir las rutinas de trabajo, por la reducción de varias de las principales poblaciones de captura (Villalobos Flores & Reneta, 2000).

Fue hasta la década de 1960 que se empezó a observar una relación entre la ya conocida “corriente El Niño” y las variaciones climáticas en distintas partes del globo. Puesto que era necesario ligar tres aspectos primordiales de los ciclos naturales: mecánica de los vientos, corrientes oceánicas y fuertes precipitaciones en diferentes zonas del planeta.

En 1920 el científico inglés Sir Gilbert Walker, tratando de pronosticar las lluvias monzónicas en la India, examinó las condiciones de presión atmosférica de estaciones meteorológicas en Sur América, el norte de Australia e Indonesia. Lo que Walker encontró fue que cuando había bajas presiones en Australia e Indonesia aumentaban en Tahití (Pacífico central) y viceversa. Sir Gilbert Walker concluyó que existía una oscilación de la presión atmosférica a nivel del mar en el Pacífico ecuatorial, de forma tal que cuando sube en el Este disminuye en el Oeste y viceversa. Tales cambios de presión son cuantificados hoy día por el Índice de Oscilación del Sur (IOS) calculado por la diferencia de anomalías de presión atmosférica entre Darwin en Australia y Tahití en el Pacífico Central. Debido a la mecánica de los vientos que van del Pacífico Este hacia el Pacífico Oeste cuando el IOS es bajo, los vientos superficiales del Este se debilitan, mientras que cuando el IOS es alto, se mantienen normalmente fuertes, lo que explica no solo el comportamiento de los vientos en el Pacífico ecuatorial sino también las fuertes precipitaciones en diferentes puntos del planeta (Villalobos Flores & Reneta, 2000).

Finalmente la relación entre “La Corriente de El Niño” y las alteraciones climáticas importantes, fue propuesta por Jacob Bjerknes, como conexiones entre el anormal calentamiento de las aguas del Pacífico Este (El Niño), el debilitamiento de los vientos del Este del Pacífico y las torrenciales precipitaciones de algunas zonas del planeta que coincidían con bajos IOS. Es a partir de los estudios realizados por Bjerknes entre 1966 y 1969 donde se establece la relación entre El Niño y la Oscilación del Sur, se acuñan las siglas ENOS (El Niño-Oscilación Sur) como el mejor término de referencia que involucra a la vez el componente oceánico y el componente atmosférico del fenómeno (Villalobos Flores & Reneta, 2000) La fase cálida del

ENOS corresponde a El Niño (calentamiento oceánico e IOS negativo), mientras que la fase fría del ENOS, corresponde a La Niña (enfriamiento oceánico e IOS positivo) (SENAMHI, Dirección General de Meteorología, & Dirección de Climatología, 2014)

Para entender el comportamiento del **ENOS (El Niño Oscilación del Sur)** primero se debe comprender la mecánica del funcionamiento de sus componentes principales. Este se manifiesta como un patrón climático recurrente que implica cambios en la temperatura de las aguas en la parte central y oriental del Pacífico tropical (Kurtz et al., 1998) un evento de naturaleza Oceánica- Atmosférica cuyos efectos se manifiestan en la interface de estos dos medios, aunque pareciera que su incidencia correspondiera a una extensa región continental de Sudamérica, su influencia presenta manifestaciones a escala global.

A nivel global el océano y la atmósfera interactúan entre sí por los sistemas de vientos que controlan la circulación superficial del mar, que al soplar en una dirección, arrastran masas de agua originando las diversas corrientes marinas como la de Humboldt, el Golfo, etc. Por efecto de la fuerza de Coriolis (fuerza generada por la rotación de la Tierra), los flujos son desviados hacia la derecha en el hemisferio norte y hacia la izquierda en el hemisferio Sur (Villalobos Flores & Reneta, 2000).

Frente a las costas de Ecuador, Perú y Chile se encuentran las regiones más importantes para la observación de ENOS, ya que el componente oceánico de esta zona resulta básico para caracterizar su comportamiento, y a nivel atmosférico es cerca del Ecuador terrestre donde las corrientes confluyen, los vientos toman dirección este-oeste, y sobre la línea ecuatorial se produce una contracorriente de dirección oeste-este. A su vez la corriente oceánica más importante del Océano Pacífico Sur la de Humboldt, corre de sur a norte paralela al continente suramericano (Villalobos Flores & Reneta, 2000). Mientras que a nivel oceánico, el fenómeno de El Niño genera unas condiciones oceanográficas anormales en el Pacífico Sudoriental, como un ciclo que puede ir de 3 a 7 años, en forma de una invasión masiva de aguas anormalmente más calientes frente a las costas de Perú y Ecuador, con una profundidad de hasta 30 km al sur del Ecuador y una persistencia de por lo menos 4 meses consecutivos, ocasionando un aumento de la temperatura superficial y el nivel del mar, un hundimiento de la termoclina y una profundización de la capa de mezcla. Con el fenómeno se eleva la temperatura superficial del agua hasta en 7°C, llegando a veces hasta 30,5 °C, mientras que la salinidad disminuye de 35 a 32,7 partes por mil. Lo que reduce considerablemente la productividad biológica y dando lugar a una emigración masiva de peces por falta de alimento (Montealegre Bocanegra & Zea Mazo, 1994).

Para estudiar el proceso de evolución de las temperaturas de la superficie del océano y los vientos, se ha sectorizado la franja ecuatorial del Océano Pacífico en cuatro cuadrantes. La mayoría de ellos concentran su atención en los cuadrantes Niño 3 y Niño 3.4 para tratar de comprender la formación de este fenómeno. A partir de esta división la Administración Nacional para el Océano y la Atmósfera de los Estados Unidos de Norteamérica (NOAA), utiliza el **Índice Oceánico El Niño** (ONI, por sus siglas en inglés) (Figura 9), para la identificación de El Niño y La Niña. Este índice consta de cálculo del promedio de 3 meses consecutivos en la serie mensual de las anomalías de la temperatura superficial del mar medidas en el Pacífico tropical en la Región El Niño 3.4 (5°N - 5°S, 120° - 170°W). Se consideran episodios cálidos y fríos cuando el índice ONI supera el umbral de +/- 0,5 °C. Cuando dicho umbral es cubierto durante un mínimo de 5 meses continuos, se definen como episodios El Niño (+) o de La Niña (-) (SENAMHI et al., 2014).

Categoría	Indice Oceánico El Niño (ONI)
Cálido fuerte	Mayor o igual a +1,4
Cálido moderado	Mayor que +0,9 y menor que +1,4
Cálido débil	Mayor que +0,5 y menor que +1,0
Neutro	Mayor que -0,5 y menor que +0,5
Frío débil	Menor que -0,5 y menor que -0,9
Frío moderado	Menor que -1,0 y menor que -1,5
Frío fuerte	Menor o igual que -1,5

Figura 8. Umbrales para identificar la magnitud de El niño/La Niña, según la NOAA en la región Pacífico tropical Niño 3.4. Fuente: SENAMHI et a (2014).

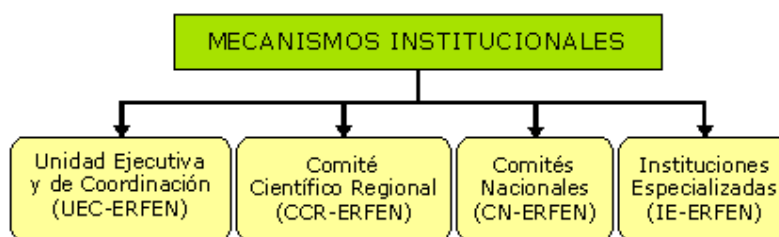
4.3.3.1. Estudio Regional Fenómeno El Niño ERFEN

Para el seguimiento, estudio y planeación de medidas que permitan afrontar los efectos que el **ENOS (El Niño Oscilación del Sur)** pueda tener sobre el territorio, se ha constituido un sistema marítimo regional denominado como la Comisión Permanente del Pacífico Sur (CPPS) creada por la Declaración de Santiago de 1952, como una alianza y opción estratégica, política y operativa en el Pacífico Sudeste, para consolidar la presencia de los países ribereños en dicha área geográfica y su proyección de manera efectiva y coordinada, tanto hacia las zonas aledañas cuanto a la vinculación con la Cuenca del Pacífico.

La Comisión Permanente del Pacífico Sur (CPPS) es un organismo intergubernamental, creado por acuerdo entre Chile, Ecuador y Perú, al que se adhirió posteriormente Colombia en 1979. La CPPS es la Secretaria Ejecutiva del Convenio de Lima 1981 y del Plan de Acción para la protección del Medio Marino y las Áreas costeras del Pacífico Sudeste.

Hoy en día la CPPS es un organismo que promueve y articula la cooperación entre sus miembros y con otros organismos de la comunidad internacional en asuntos marítimos. Entre sus principales actividades se contempla la realización de estudios científicos relacionados con el fenómeno del Niño a fin de pronosticar oportunamente este fenómeno y alertar a sus poblaciones de los riesgos climáticos asociados a este evento. Para esta tarea específica fue creado en 1974 el Programa para el Estudio Regional del Fenómeno El Niño en el Pacífico Sudeste (ERFEN) e inició sus actividades en 1976. Actividad que se desarrolla mediante cruceros de investigación efectuados con el apoyo de los buques de investigación de los cuatro países miembros, coordinados por la CPPS, constituyendo una acción conjunta, única a nivel internacional (Comisión Permanente del Pacífico Sur CPPS, 2015a). Los resultados que ha arrojado el programa ERFEN en especial tras los drásticos eventos ENOS El Niño 1972 - 1973 y de 1982 - 1983, han hecho que adquiera notable suficiencia en los aspectos físicos - oceanografía y meteorología - y en los biológico-pesqueros relativos a la incidencia del fenómeno en los recursos hidrobiológicos que sustentan importantes pesquerías de la región. Se ha avanzado significativamente en la evaluación del clima marino y sus variabilidades, especialmente en los episodios cálidos El Niño y los fríos de La Niña; en el aviso anticipado de su aparición, considerando la información regional y extraregional, y en la cuantificación de sus impactos en el ámbito pesquero principalmente.

El ERFEN realiza sus actividades sustentado en el esfuerzo integrado de más de 22 instituciones científicas de Chile, Colombia, Ecuador y Perú, la Figura 10 muestra los mecanismos institucionales para el desarrollo de sus actividades.



Tomado del Artículo XI "Mecanismos Institucionales" del "Protocolo sobre el Programa para el Estudio Regional del Fenómeno de El Niño en el Pacífico Sudeste (ERFEN).

Figura 9. Mecanismos institucionales Protocolo ERFEN. Fuente: CPPS (2015a)

El Protocolo ERFEN promueve el intercambio de investigadores para la cooperación científica y técnica, además de la ampliación de las capacidades nacionales para el manejo e interpretación de la información, a fin de mejorar la capacidad de pronóstico del fenómeno El Niño y contribuir a la mitigación de los efectos perjudiciales y aprovechar los efectos benéficos. (C. P. del P. sur CPPS, 2015).

De los mecanismos institucionales para Colombia está el **Comité Técnico Nacional para el Estudio del Fenómeno de El Niño (CTN ERFEN)** que de acuerdo con la "Declaración de Quito" firmada por los representantes de los países miembros, se establece la necesidad de fortalecer el Programa ERFEN en los campos meteorológico, oceanográfico (físico y químico), biológico-marino, biológico-pesquero, de capacitación, y socio-económico; además de dotar al programa de un adecuado marco jurídico, institucional, financiero y de cooperación a nivel regional.

El Niño en el Pacífico Sudeste (ERFEN) fue ratificado en Colombia mediante Ley 295 de 1996, con la meta básica del Protocolo ERFEN de poder predecir los cambios oceánico-atmosféricos, con anticipación suficiente para permitir políticas de adaptación o de emergencia frente a variaciones en el rendimiento pesquero, agrícola e industrial y decisiones de mercadeo, manejo de recursos hidrobiológicos y otras.

El principal objetivo del **CTN ERFEN** es integrar los esfuerzos y capacidades de las instituciones nacionales que lo conforman, no solo las del orden técnico y científico (DIMAR, IDEAM, Sub Gerencia de Pesca y Acuicultura) sino también aquellas que deben afrontar las situaciones de emergencia que este fenómeno pueda ocasionar. Con el fin de preparar a la comunidad en general, ante la ocurrencia de este evento natural, buscando atenuar los impactos de tipo ambiental y socioeconómico que genera un evento cálido (El Niño) o frío (La Niña) en el país (CCO, 2015).

Las labores de investigación que promueve el programa ERFEN en colaboración con los Comités Nacionales de cada Estado miembro incluye la realización de cruceros oceanográficos programados para tomar muestras de agua y mediciones de diferentes parámetros que permitan hacer un seguimiento de la variabilidad de las condiciones del mar en cada sección de la costa sudamericana.

Para mantener los estándares, el programa ERFEN establece una serie de estaciones de medición asignadas por coordenadas de latitud y longitud, con distribuciones diferentes para las áreas marítimas de cada país. Este conjunto de coordenadas se denomina MALLA ERFEN. En

el caso de Colombia inició su adopción a partir de 1992 para los cruceros oceanográficos en expedición científica en convenio con el programa ERFEN. La Malla ERFEN de la Cuenca Pacífica Colombiana (CPC) se puede apreciar en la Figura 11, en la cual las estaciones físicas están marcadas con color rojo y las físico-químicas-biológicas en color verde.

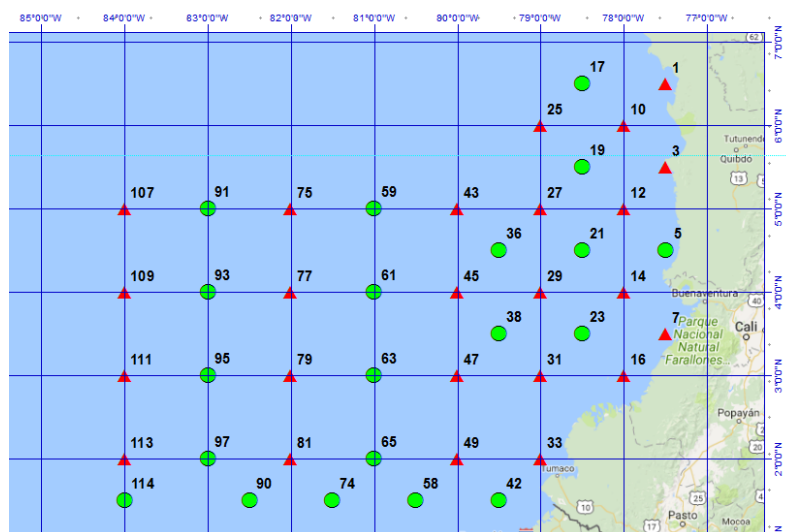


Figura 10. Malla ERFEN Cuenca Pacífica Colombiana (CPC). Fuente: Dimar.

En concordancia con las iniciativas del programa ERFEN para el seguimiento del fenómeno de El Niño, los países miembros desde su propia jurisdicción hacen monitoreo de las condiciones océano-atmosféricas de sus zonas marítimas. Aunque en un principio ERFEN operaba con énfasis en la oceanografía biológica, hoy día abarca los aspectos físicos – oceanográficos y meteorológicos, además de los biológico-pesqueros. Los cruceros oceanográficos ejecutan sus actividades con el propósito de desarrollar el programa sobre el Protocolo ERFEN, incluyendo acciones de monitoreo, cooperación y desarrollo de capacidades, con el propósito de contribuir a la vigilancia climática para la gestión del riesgo; así como en el respaldo a investigación oceanográfica para la Región con la finalidad de apoyar el monitoreo y evaluación de ENOS (CPPS, 2015).

A nivel nacional el monitoreo y seguimiento del fenómeno de El Niño y La Niña, es una labor conjunta entre la Dirección General Marítima (Dimar), el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) y el CTN ERFEN de la Comisión Colombiana del Océano (CCO), mediante la vigilancia continua de factores oceánicos y atmosféricos para detectar posibles anomalías en la Temperatura Superficial del Mar (TSM) en el Océano Pacífico Tropical Central para la emisión de comunicados y boletines periódicos acerca de las condiciones océano-atmosféricas a nivel nacional que permitan informar a la población sobre estos eventos climáticos y sus posibles efectos en el resto del país. Así también el Instituto de Hidrología, Meteorología y Estudios Ambientales emite periódicamente boletines sobre la predicción climática mensual para informar al público sobre las condiciones meteorológicas a nivel nacional (IDEAM, 2014b).

Debido al amplio espectro de efectos que el fenómeno de El Niño puede tener en varios aspectos del ámbito nacional, también existen otros actores que se involucran en el seguimiento de este evento climático. Como lo son la Federación nacional de Cafeteros de Colombia (CENICAFE) con publicaciones como “Variabilidad climática en la zona cafetera colombiana asociada al evento de El Niño y su efecto en la caficultura” (Jaramillo Robledo, Arcila Pulgarín, & CENICAFE, 2009); y El Centro de Investigación de la Caña de Azúcar de Colombia (CENICANÑA) con reportes sobre la evolución e impactos del fenómeno de El Niño en el valle del Río Cauca (2015) (CENICANÑA, 2017).

4.3.3.2. Impactos del Fenómeno El Niña- La Nina en Colombia

El impacto del fenómeno El Niño se manifiesta de diferentes maneras de acuerdo a las regiones en el territorio nacional. En la región Andina es altamente significativo debido a deficiencias de lluvia, el territorio es afectado por deficiencias de precipitación cuando ocurre el evento; en la región Caribe el impacto la situación es similar por déficit de lluvias, ya que para los eventos El Niño 1972 - 1973, 1976 - 1977, 1991-1992 y 1997 - 1998, el territorio afectado lo fue por más del 80% del tiempo de duración de cada uno de estos eventos; en la región Pacífica el número de eventos El Niño con deficiencias de lluvia es significativo, y la deficiencia de precipitación también es significativo (Hurtado Moreno, González, & IDEAM, 2015).

Para la Amazonia la presencia de un evento El Niño no presenta efectos significativos. Tan solo el evento 1976 -1977 mostró déficits significativos, mientras que el evento 2002 – 2003 mostró predominio de meses lluviosos; caso similar ocurre en la Orinoquia colombiana (Hurtado Moreno et al., 2015).

En el caso de La Niña la región Andina manifiesta un impacto significativo, por el exceso de lluvia en la mayoría de los casos. Sin embargo, los eventos 1984 - 1985 y 2001 – 2002 registraron un impacto totalmente contrario siendo afectados por sequía; la causa puede atribuirse a eventos muy débiles. Mientras que en la región Caribe el impacto de la Niña en general es variable. En dos eventos predominó el déficit (1984 – 1985 y 2001 – 2002) y en dos eventos predominó el exceso (1970 – 1972 y 1999 – 2000) (Hurtado Moreno et al., 2015).

En la región Pacífica es significativo el exceso de las lluvias, en la mayor parte de los eventos La Niña. Existen excepciones como fue el evento 1984 – 1985, el cual presentó una sequía. Para la Amazonia, durante un evento La Niña, mayormente se presentan sequías aunque sin afectar grandes porciones de territorio; y en la Orinoquia el impacto de la Niña en general es variable. Con frecuencia se presentan eventos mayormente secos, aunque para los eventos 1970 – 1972 y el 1999 – 2000 predominó el tiempo lluvioso (Hurtado Moreno et al., 2015).

Los impactos que tanto el fenómeno de El Niño como el de La Niña han tenido en el territorio colombiano en los últimos años, así como también las particularidades de su comportamiento y evolución en la región han sido estudiadas por Pabón J.D, Poveda G, Martínez R y Montealegre J.E muchos de cuyos aportes se encuentran recopilados publicaciones como “ Los Fenómenos de El Niño y de La Niña, su efecto climático e impactos socioeconómicos” (Pabón Caicedo & Montealegre Bocanegra, 2017). Otras investigaciones como “La Variabilidad Climática Interanual Asociada al Ciclo El Niño-La Niña-oscilación del Sur y su Efecto en el Patrón Pluviométrico de Colombia. En: Meteorología Colombiana” (Montealegre Bocanegra & Pabón Caicedo, 2000), y “Seasonality in ENSO-Related Precipitation, River Discharges, Soil Moisture, and Vegetation Index in Colombia. En: Water Resources Research” (Poveda, Jaramillo, Gil, Quiceno, & I.Mantilla, 2001).

4.3.4 Calidad de Datos

La UNESCO a través de la Comisión Oceanográfica Intergubernamental (COI) establece el control de calidad de los datos oceanográficos con el objetivo de asegurar la coherencia de los mismos dentro de un único conjunto de datos que a su vez hacen parte de una colección de conjuntos de datos, con el fin de garantizar que la calidad y los errores de los datos son evidentes para el usuario, y que este cuenta con suficiente información para evaluar su idoneidad para un uso específico (UNESCO, 2013b).

Ya que el concepto de calidad no tiene un significado explícito, se relaciona con el grado en que un conjunto de características inherentes cumple con los requisitos, o se refiere a la percepción

del grado en que cumple con las expectativas del cliente, también puede referirse a un conjunto específico de requisitos (UNESCO, 2013a) por lo que los procesos de control de calidad pueden ser tan variados, para diferentes casos.

El control de datos permite verificar si el valor de un dato notificado es representativo de la medición que se pretendía efectuar y no se ha visto afectado por factores no relacionados con el mismo (Organización Meteorológica Mundial OMM, 2011) es decir, que el dato en cuestión logre reflejar el comportamiento de la variable que se midió inicialmente, representando las lecturas de una determinada serie de datos. Los datos no deben ser considerados aptos para su archivo si no han aprobado los controles de calidad respectivos (Organización Meteorológica Mundial OMM, 2011).

De acuerdo a las mejores prácticas adoptadas por el Cecoldo, los datos primarios deben pasar por un proceso de normalización, para dejarlos en un formato estándar e interoperable (Centro Colombiano de Datos Oceanográficos, 2016). Durante el proceso de normalización se incluye la asignación de banderas de calidad, como indicadores de calidad de datos, proporcionando al usuario información clara acerca de las acciones tomadas para evaluar los datos originales (UNESCO, 2013b).

Ya que existe la misma preocupación en lo referente a la gestión, calidad, procesamiento, almacenamiento y transferencia de una gran cantidad de datos. Con ligeras modificaciones el presente trabajo de investigación explora las metodologías de procesamiento de datos aplicados en la climatología que podrían ser útiles a nivel estadístico y matemático para garantizar la calidad de datos de parámetros oceanográficos.

En capítulos anteriores se ha descrito las diferentes metodologías de muestreo de parámetros oceanográficos colombianos; a este proceso se le denomina aseguramiento de la calidad (QA, *Quality Assurance*), es decir, la parte de la gestión de la calidad que se centra en brindar confianza de que se cumplirán los requisitos de calidad, lo que incluye el monitoreo y evaluación sistemática de los procesos asociados con la generación de un producto o servicio; mientras que el control de calidad (QC, *Quality Control*) se refiere al proceso de monitoreo de la producción de las actividades de aseguramiento de la calidad para mejorar los productos y servicios para que se cumplan los requisitos y / o estándares de calidad (UNESCO, 2013a), y en el caso del presente proyecto hace referencia al procesamiento de los datos primarios.

Se hacen múltiples esfuerzos porque los procedimientos de aseguramiento de la calidad (QA) que consideren todos los tipos, fuentes y causas posibles de errores (Organización Meteorológica Mundial OMM, 2011) haciendo todo lo posible para reducirlos. La “Guía de Normalización de Conjunto de Datos Oceanográficos” (Centro Colombiano de Datos Oceanográficos, 2016) es un referente que para propósitos del desarrollo de este trabajo se tiene constantemente en cuenta sobre los pasos, acciones, regulaciones y especificaciones en los formatos, nomenclatura y almacenamiento de archivos para garantizar la calidad de los datos recolectados en los cruceros oceanográficos.

4.3.5. Procedimientos Estadísticos de Control de Calidad

En la literatura se encuentran múltiples procedimientos y metodologías sugeridas para el **proceso de control de calidad** de datos, de acuerdo a las necesidades del área en que se trabaja. Para entender las denominaciones de calidad que se le puede atribuir a un conjunto de datos se debe entender la secuencia estadística y matemática que permite llegar a calificar los datos con ciertas marcas de calidad. Las técnicas estadísticas son de gran valor para detectar errores y en algunos casos para indicar cuál debería ser el valor “correcto” (Organización Meteorológica Mundial OMM, 2011).

Existe un paralelo entre el manejo de datos climatológicos y oceanográficos, dos disciplinas transversalmente relacionadas que sin embargo operan de manera casi independiente. Pero en

general las observaciones marítimas pueden ser sometidas a procedimientos estadísticos similares a aquellos que se aplican a las estaciones terrestres de superficie (Organización Meteorológica Mundial OMM, 2011).

Bajo el supuesto de que los controles aseguramiento de la calidad (QA) a nivel operativo fueron aplicados en la medida de lo posible, y confiando en la información suministrada por el metadato, se procede a la aplicación de las técnicas estadísticas pertinentes correspondientes al proceso de control de calidad (QC).

4.3.5.1. Pruebas de Coherencia

Las pruebas de coherencia permiten identificar la lógica del valor de un dato con respecto al conjunto de datos al que pertenece. Las pruebas de coherencia de los datos que se tienen en cuenta para propósitos del presente trabajo son la interna, la temporal y la espacial. A nivel climatológico y oceanográfico se debe tener presente que los valores de los datos guardan una relación mutua tanto temporal como espacialmente y con el conjunto de mediciones al que pertenece.

La **coherencia interna** se basa en las relaciones físicas que existen entre los elementos climatológicos. Todos los elementos han de verificarse exhaustivamente haciendo una comparación con otros elementos afines dentro de cada observación, relacionado a los límites físicos del parámetro a medir (Organización Meteorológica Mundial OMM, 2011). Así como la temperatura superficial del mar no puede sobrepasar los 31,3°C (NOAA & Operations, n.d.) o llegar al punto de congelación en aguas tropicales; la radiación mundial no puede ser superior a la irradiancia en lo alto de la atmósfera, la dirección del viento debe oscilar entre 0° y 360° y la precipitación no puede ser negativa, entre otros.

La **coherencia temporal** examina la variación de un elemento en el tiempo. Muchos conjuntos de datos climatológicos muestran una importante correlación en serie. Los controles constan de una comparación del elemento con las observaciones anteriores y posteriores en el ordenamiento. (Organización Meteorológica Mundial OMM, 2011). Ha de identificarse un comportamiento lógico y muchas veces secuencial a lo largo del tiempo. Los análisis permiten establecer la magnitud del cambio que podría experimentar un determinado elemento en cualquier intervalo de tiempo, teniendo en cuenta factores como el comportamiento de la variable, la estación del año, la ubicación y el lapso de tiempo transcurrido entre dos observaciones sucesivas, y a su vez identificar valores sospechosos al compararse con el comportamiento global de los demás elementos en la serie de tiempo, y reconocer las posibles causas de error.

La **coherencia espacial** compara cada observación con observaciones efectuadas al mismo tiempo en diferentes ubicaciones geográficas dentro de la misma zona. Cada observación puede compararse con las que cabría esperar en su ubicación teniendo en cuenta las observaciones de las estaciones vecinas. Aquellos datos que difieren notablemente entre las observaciones esperadas y las que, de hecho, se obtienen deberán marcarse para ser sometidos a una revisión, corrección o eliminación, según proceda. (Organización Meteorológica Mundial OMM, 2011).

Para tratar el tema de la coherencia espacial y para propósitos del análisis de datos oceanográficos es importante recurrir a los elementos de la **geo estadística**, como una herramienta que ofrece una manera de describir la continuidad espacial, la cual es un rasgo distintivo de muchos fenómenos naturales, además de proporcionar adaptaciones de las técnicas clásica de regresión para tomar ventajas de esta continuidad (Giraldo Henao, n.d.). Como ya se ha explicado en capítulos anteriores los parámetros oceanográficos pueden ser representados gráficamente en el océano mediante cartografía (isotermas, termoclinas, halo clinas, etc.) que permiten apreciar la distribución e intensidad de las variables a través del espacio muestreado.

La **geo estadística** tiene en cuenta la posición, dirección, magnitud y distancia entre los valores de observación. Con esta información se pueden hacer análisis estructurales, en donde describe la correlación entre los puntos en el espacio y posteriormente incurrir en una fase de predicción en sitios de la región no muestreados por medios de la técnica *kriging*. Para propósitos del desarrollo de este trabajo se aborda la primera fase referente al análisis de correlación de las variables oceanográficas en diferentes puntos muestreados en el Pacífico Colombiano; dejando a un lado el tema de predicción y construcción de modelos que probablemente pueda ser abordado en investigaciones posteriores.

En el caso de datos oceanográficos se cuenta con muestreos realizados en un conjunto de coordenadas específicas de acuerdo al año de captura de los datos. Según I. Clark (1979), a cada valor muestral puede serle asignado un “peso” y existen una serie de métodos desarrollados para asignar dicho “peso”. Se parte del supuesto de que la relación existente entre los valores de un punto A y un punto B, cualquiera sea el valor de la muestra depende de la distancia, y en ciertas ocasiones de la dirección entre las dos posiciones, en nada más. Aunque de este supuesto surgen múltiples interrogantes de acuerdo a las características de cada caso.

Para el desarrollo del presente trabajo se hace énfasis en la correlación existente entre las estaciones de muestreo. Esta fase es conocida también como análisis estructural, y para llevarla a cabo con base en la información muestral, se usan funciones de correlación espacial, como son el semivariograma, el covariograma y el correlograma.

El **semivariograma** permite describir gráficamente la diferencia esperada en valores entre pares de muestras con una orientación relativa dada. Su función se define como:

$$\gamma(h) = \frac{\sum (Z(x+h) - Z(x))^2}{2n}$$

Donde $g(x)$ denota el valor de una muestra en una posición x ; $g(x+h)$ el valor de una muestra en una posición diferente de x ya que h representa la distancia entre muestras y su relativa orientación y n el número de pares de muestras que se tienen. La mitad del variograma $\gamma(h)$, se conoce como la función de semivarianza y caracteriza las propiedades de dependencia espacial del proceso.

La función de semivarianza se calcula para varias distancias h . En la práctica, debido a irregularidad en el muestreo y por ende en las distancias entre los sitios, se toman intervalos de distancia $\{[0, h], [h, 2h], [2h, 3h], \dots\}$ Se llama semivariograma experimental al semivariograma que utiliza una distancia promedio entre parejas de sitios dentro de cada intervalo y no a una distancia h específica. Se debe tener en cuenta que el número de parejas de puntos n dentro de los intervalos no es constante (Giraldo Henao, n.d.).

La interpretación del semivariograma experimental parte del criterio de que a menor distancia entre los sitios mayor similitud o correlación espacial entre las observaciones. Por ello en presencia de auto correlación se espera que para valores de h pequeños el semivariograma experimental tenga magnitudes menores a las que este toma cuando las distancias h se incrementan.

El eje x representa la distancia, mientras que el eje y la correlación. Desde el origen la correlación tiende a aumentar hasta que llega a un punto máximo a partir de cual se vuelve constante o empieza a decaer, dicho punto representa la distancia máxima a la que se identifica la mayor correlación entre los datos, cualquier h mayor o menor es inferior a la correlación máxima del conjunto de datos.

El **covariograma y correlograma** parte de la función de covarianza muestral entre parejas de observaciones que se encuentran a una distancia h , calculado por:

$$C(h) = COV(Z(x+h), Z(x)) = \frac{\sum_{i=1}^n (Z(x+h) - m)(Z(x) - m)}{n}$$

$$= \frac{\sum_{i=1}^n (Z(x+h) \cdot Z(x))}{n} - m^2 = C(h)$$

Donde m representa el valor promedio en todo punto de la región de estudio y n es el número de parejas de puntos que se encuentran a una distancia h .

En este caso se asume que el fenómeno es estacionario y estimando la varianza de la variable regionalizada a través de la varianza muestral, se tiene que el correlograma muestral está dado por:

$$r(h) = \frac{COV(Z(x+h), Z(x))}{S_{x+h} \cdot S_x} = \frac{C(h)}{S_x^2} = \frac{C(h)}{C(0)}$$

Bajo el supuesto de estacionariedad cualquiera de las tres funciones de dependencia espacial mencionadas, es decir semivariograma, covariograma o correlograma, puede ser usada en la determinación de la relación espacial entre los datos. Sin embargo como se puede observar en las fórmulas, la única que no requiere hacer estimación de parámetros es la función de semivarianza. Por esta razón, fundamentalmente, en la práctica se emplea el semivariograma y no las otras dos funciones (Giraldo Henao, n.d.).

Según I. Clark (1979) se puede esperar que la distribución este gobernada por la distancia entre pares de estaciones y su relativa orientación.

Llegado a este punto la construcción del correlograma, según sea el caso permite identificar la correlación de los valores entre puntos de muestreo. El trabajo requerido en la organización de los datos y su elaboración puede ser llevado a cabo a través de software de análisis geo estadístico como *PASSAGE* (Rosenberg & Corey D. Anderson, 2011), el cual al introducir los datos de las estaciones y sus respectivas coordenadas, arroja junto con la gráfica el índice de Moran, medida de auto correlación espacial que lleva el nombre de su desarrollador Patrick Alfred Pierce Moran (1950) que se calcula como:

$$I = n \frac{\sum_{ij} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_{i=1}^n (x_i - \bar{x})^2}$$

Donde x_i es el valor de la variable en la i -ésima posición; n es el número de puntos; w_{ij} es un peso que indica algo sobre la relación espacial de los puntos i y j ; \sum_g indica la suma doble sobre todo i y todo j donde $i \neq j$; e $W = \sum_g w_{ij}$ la suma de los valores en la matriz de pesos. El índice de Moran es similar a un coeficiente de correlación, ya que se basa en una formulación de momento del producto. Normalmente oscilará de 1 a -1. Para n grande los valores se aproximan a cero. Valores positivos del índice indican autocorrelación espacial positiva, valores negativos autocorrelación espacial negativa.

Es pertinente elaborar un procedimiento integrado para examinar la coherencia de los datos en la totalidad de sus dimensiones. Todas las pruebas de coherencia deben documentarse por completo mediante procedimientos, fórmulas y criterios de decisión.

4.3.5.2. Series de Tiempo

Una serie de tiempo es un conjunto de observaciones ordenadas según la secuencia en que se efectuaron. Para propósitos de la climatología, oceanografía y en general variables que representan fenómenos naturales, las series de tiempo permiten identificar características tendenciales, que constituyen un resumen del comportamiento histórico de las observaciones de un elemento. Las tendencias pueden ser lineales o no lineales, como una curva, o incluso un desplazamiento abrupto hacia arriba o hacia abajo, y por lo general se mantienen en las series climáticas durante un período finito, que puede ser bastante largo (Organización Meteorológica Mundial OMM, 2011).

El conjunto de datos suministrado por la Dimar se identifica básicamente como una serie de tiempo de variables oceanográficas. Aunque discontinua a nivel temporal se tiene información de parámetros oceanográficos de varios años y en diferentes meses a lo largo del tiempo. Para un estudio completo de las series temporales, determinar las tendencias es solo una parte del trabajo, también requiere de la identificación de las oscilaciones periódicas o cuasi periódicas, así como las variaciones irregulares o aparentemente fortuitas que se observan en los datos. En esta parte del proceso el objetivo del análisis de las series temporales es comprender la manera en la que la variabilidad de estas se distribuye como función de la escala temporal (Organización Meteorológica Mundial OMM, 2011).

Normalmente en el ámbito de la meteorología y la climatología, las observaciones sucesivas tienden más a asemejarse que a diferenciarse. La medida utilizada para resumir la relación entre cada observación y la que le precede constituye el coeficiente de autocorrelación. Esta medida se calcula de la misma manera que el coeficiente de correlación, con la excepción de que la segunda serie es la misma que la primera, pero con un desplazamiento de uno o más saltos de tiempo. Las medidas que simplifican las tendencias dependen del tipo de tendencia que se está determinando.

Las series temporales de referencia dan buen resultado cuando el conjunto de datos tiene una cantidad de valores lo suficientemente grande como para garantizar una buena relación climatológica entre cada serie propuesta y las zonas aledañas utilizadas para establecer la serie de referencia y cuando no haya faltas de homogeneidad que afecten a todas o a la mayoría de las estaciones o valores disponibles (Organización Meteorológica Mundial OMM, 2011).

4.3.5.3. Intervalos de Confianza

Las pruebas de tolerancia anteriormente descritas permiten conducir el proceso de evaluación de los datos hacia la construcción de límites superiores o inferiores sobre los posibles valores de un elemento climatológico. Se debe tener en cuenta que por lo general, los límites suelen depender del tiempo y la ubicación y deberían establecerse recurriendo a los valores históricos o mediante métodos de interpolación espacial. También es importante determinar y remediar rápidamente los sesgos sistemáticos que presenta la información proporcionada por los instrumentos. Para ello es necesario conservar documentación con información de las pruebas de tolerancia que se han efectuado, los límites climáticos establecidos respecto de cada elemento estudiado y las razones fundamentales para determinar dichos límites (Organización Meteorológica Mundial OMM, 2011).

En general, las pruebas de tolerancia comparan un valor en cuestión con algún patrón que utiliza un umbral estadístico. Algunas de las pruebas de tolerancia simples consisten en comparar un valor observado con el valor extremo o récord o con algún múltiplo de desviaciones típicas en torno al valor promedio correspondiente a esa fecha. En el segundo caso, se ha de tener en cuenta la posibilidad de que tal vez el elemento no tenga necesariamente una distribución simétrica o gaussiana, o que algunos valores extremos determinados a partir del multiplicador

de la desviación típica puedan ser incorrectos. (Organización Meteorológica Mundial OMM, 2011).

La guía de prácticas climatológicas (Organización Meteorológica Mundial OMM, 2011), recomienda utilizar datos históricos de largo período para el control de calidad, es decir, efectuar comparaciones con información de referencia previamente normalizada y evaluada. Se pueden aplicar algunas pruebas de tolerancia utilizando series de datos completamente distintos, tales como datos obtenidos por satélite o por radar. En este caso se recurre al Atlas de los Datos Oceanográficos de Colombia 1922-2013 el cual proporciona una visión tanto a nivel espacial como temporal de las condiciones climatológicas de la cuenca Pacífica Colombiana. Así con base en la información que suministra es posible construir un intervalo de confianza para cada punto de muestreo y para cada mes o década de los que haya información disponible.

Una ventaja que suministra el Atlas es que la información oceanográfica se adhiere a un procesamiento estadístico propuesto por la National Oceanic and Atmospheric Administration (NOAA, 2013), en el que las mediciones son evaluadas dentro de un rango de valores de acuerdo a la desviación estándar de la media del conjunto de datos al que pertenecen por niveles estándar de profundidad y cercanía a zonas costeras para descartarlos como datos atípicos fuera del intervalo de confianza. Así:

- Costero: El valor de los datos de nivel estándar no debe superar las 5 desviaciones estándar calculadas dentro de la cuadrícula 5x5 (grados de latitud y longitud) en los 50 m superiores.
- Cercano a la costa: El valor de datos de nivel estándar no debe superar 4 desviaciones estándar calculadas para 5x5 (grados de latitud y longitud) en la parte superior 50 m.
- Océano abierto: El valor de los datos de nivel estándar no debe superar tres desviaciones estándar calculado para la cuadrícula 5x5 (grados de latitud y longitud), excepto cuando un perfil estaba en o por debajo de la profundidad media. Nivel para sección de un grado en la que está contenida, o cualquiera de las cuadrículas adyacentes de un grado, se usan 4 desviaciones estándar.

Según Ospina Botero (2001) el intervalo de confianza recomendado para este tipo de fuente es el la **distribución t de Student**, definido por la expresión:

$$\bar{X} \pm t_{n-1} ee[\bar{X}]$$

Donde \bar{X} representa la media aritmética del conjunto de datos, $ee[\bar{X}]$ el error estándar que se puede calcular al dividir la desviación estándar en la raíz cuadrada del tamaño de muestra (s/\sqrt{n}) y t un valor asociado a los grados de libertad de acuerdo al nivel de confianza y al tamaño de la muestra n establecido en la tabla *t de Student*.

Lo que caracteriza este tipo de distribución de la metodología z , es el factor t_{n-1} y gráficamente:

- Tener forma de montículo y es simétrica alrededor de $t = 0$, igual que z como se observa en el Figura 12.
- Es más variable que z , con “colas más pesadas”; esto es, la curva t no aproxima al eje horizontal con la misma rapidez que z . Esto es porque el estadístico t abarca dos cantidades aleatorias, X y s , en tanto que el estadístico z tiene sólo la media muestral \bar{X} .
- La forma de la distribución t depende del tamaño muestral n . A medida que n aumenta, la variabilidad de t disminuye porque la estimación s de σ está basada en más y más

información. En última instancia, cuando n sea infinitamente grande, las distribuciones t y z son idénticas. (Mendenhall, Beaver, & Beaver, 2010)

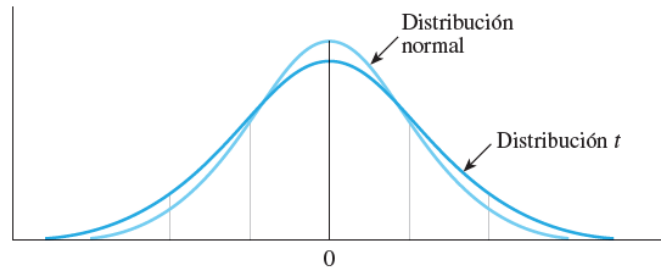


Figura 11. Gráfico de la distribución normal. Fuente: Mendenhall et al (2010)

La construcción de intervalos de confianza supone un reto en el que se tengan en cuenta la mayor cantidad de elementos posibles que puedan influir sobre el rango de los datos. En una aproximación entre el análisis estadístico y espacial de correlación de datos, a partir del principio del intervalo de confianza de Student se explora una metodología que va más allá; en donde el intervalo de confianza considere la correlación espacial entre las observaciones.

El tema de la correlación espacial como elemento influenciador en la determinación del intervalo, primero se debe considerar un conjunto de muestras independientes X_1, X_2, \dots, X_n con una distribución normal con media \bar{x} y varianza conocida σ^2 . Donde el estimador más imparcial es el promedio de la muestra $\bar{x} = \sum_i X_i/n$, que sigue una distribución normal con media \bar{x} y varianza σ^2/n , dando un intervalo de dos lados del 95% (ejemplo) de μ .

$$\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}}$$

Nótese la similitud con el intervalo propuesto por Student

$$\bar{x} \mp t_{n-1} s/\sqrt{n}$$

Ahora en lugar de considerar datos independientes, se asume que estos exhiben una correlación espacial en R^1 que decrece exponencialmente a medida que la separación entre los puntos de muestreo aumenta.

Así la varianza de la función se ve modificada para introducir la correlación espacial existente entre los datos, representada por ρ ; que para efectos prácticos se toma como el valor de coeficiente de Moran calculado a partir del conjunto de datos y sus coordenadas y que fue ilustrado anteriormente (Li, 2009).

$$var(\bar{x}) = \frac{\sigma^2}{n} \left[1 + 2 \left\{ \frac{\rho}{1-\rho} \right\} \left(1 - \frac{1}{n} \right) - 2 \left\{ \frac{\rho}{1-\rho} \right\}^2 \frac{(1-\rho^{n-1})}{n} \right]$$

$$ee(\bar{x}) = \sqrt{var(\bar{x})} = \frac{\sigma}{\sqrt{n}} \sqrt{\left[1 + 2 \left\{ \frac{\rho}{1-\rho} \right\} \left(1 - \frac{1}{n} \right) - 2 \left\{ \frac{\rho}{1-\rho} \right\}^2 \frac{(1-\rho^{n-1})}{n} \right]}$$

4.3.5.4. Homogeneidad de Datos

Los conjuntos de datos homogeneizados permiten analizar conjuntos de datos climatológicos para detectar cambios y tendencias de forma más fiable, ya que en un grupo de valores homogéneo todas las fluctuaciones contenidas en su serie temporal reflejan la variabilidad y el cambio reales del elemento representado. La conveniencia de contar con un registro homogéneo es fundamentalmente para satisfacer la necesidad de detectar y señalar los cambios en el comportamiento de los parámetros de interés a gran escala (Organización Meteorológica Mundial OMM, 2011).

La mayoría de los métodos estadísticos parten de la hipótesis de que los datos que se están examinando están lo más exentos posible de errores debidos a los instrumentos, la codificación, el proceso y otros errores no asociados al fenómeno natural en sí. Sin embargo, por lo general, los datos meteorológicos o climáticos ni son homogéneos ni están libres de errores (Organización Meteorológica Mundial OMM, 2011). Por lo que se recurre tanto a recomendaciones y acciones para mantener la homogeneidad del registro a lo largo del tiempo, como de técnicas de homogenización que corrijan tales errores.

Tras el análisis de los metadatos, las pruebas estadísticas pueden encontrar otras faltas de homogeneidad. Normalmente, las pruebas suelen depender de la escala temporal de los datos; las pruebas utilizadas para los datos diarios son diferentes de las utilizadas para datos mensuales y de otras escalas. Los resultados de esos procedimientos de homogeneización estadística se han de verificar una vez más con los metadatos existentes.

Cuando se evalúe la fiabilidad de los resultados de cualquier prueba, deberían tomarse en consideración los efectos de la autocorrelación serial, el número de posibles puntos de cambio en una serie (documentados con metadatos y sin documentar), las tendencias y oscilaciones y los períodos de registro breves que pueden ser anómalos. Muchos métodos se basan en la comparación de los datos que se han de homogeneizar (serie propuesta) con una serie de referencia. Idealmente, una serie temporal de referencia deberá haber experimentado todas las influencias climáticas generales de la serie propuesta, pero ninguno de sus sesgos posibles y artificiales. Si la serie propuesta es homogénea, cuando esta y la serie de referencia se comparan mediante el método de las diferencias (en el caso de elementos medidos en una escala de intervalos, como la temperatura) o mediante el cálculo de los cocientes o los cocientes logarítmicos (en el caso de los elementos medidos en una escala proporcional, como la precipitación), la serie temporal resultante no mostrará ni cambios repentinos ni tendencias, sino que oscilará en torno a un valor constante. Sin embargo, si hay una o más faltas de homogeneidad, estas resultarán evidentes en la serie temporal de diferencias o de cocientes (Organización Meteorológica Mundial OMM, 2011).

4.3.6. Banderas de Calidad (QF)

Las banderas de calidad son un código (numérico o alfabético) que es asignado a los datos oceanográficos y proveen al usuario de los mismos información clara sobre las acciones tomadas sobre los datos originales (UNESCO, 2013b) hasta llegar a los datos finales puestos a disposición del público. Las banderas revelan características de calidad o confiabilidad de los valores que los datos registran, así como verificaciones que se han llevado a cabo para alterar las cifras después del control, rellenar los vacíos de datos, etc., pueden variar de un proyecto a otro y entre diferentes laboratorios y centros de datos (UNESCO, 2013b).

El objetivo del control de la calidad de los datos oceanográficos es garantizar la coherencia de los datos dentro de un solo conjunto de datos y dentro de una colección de conjuntos de datos, y asegurar que la calidad y los errores de los datos sean evidentes para el usuario y evaluar su idoneidad para una tarea específica (UNESCO, 2013b).

Existe una variedad de diferentes códigos y esquemas para asignar banderas de calidad bajo diversos parámetros de control de los datos, pero en esencia procuran el mismo principio de calidad de los datos. Para propósitos del presente trabajo cada uno de los datos procesados tendrán asignada una bandera de calidad que corresponda al esquema recomendado por el programa internacional IODE de la Unesco-COI, las banderas IODE (como se denominaran de aquí en adelante), el cual propone dos niveles para la asignación de banderas de calidad los cuales se explican a continuación.

4.3.6.1. Primer Nivel

En los indicadores de nivel primario a medida que aumenta el valor del indicador señala una disminución de la calidad de los datos. El primer nivel consta de valores de calidad y sus correspondientes definiciones, como se observa en la Tabla II.

Tabla II. Banderas de calidad IODE Fuente: Centro Colombiano de Datos Oceanográficos (2016)

Valor de Nivel Primario	Nombre corto	Definición
1	Bueno	Pasó las pruebas de calidad documentadas.
2	No evaluada, no disponible o desconocida	Se utiliza cuando no se ha realizado una prueba de calidad o la calidad de los datos es desconocida.
3	Cuestionable o sospechoso	Falló una prueba documentada métrica no-crítica o una prueba subjetiva.
4	Malo	Falló una prueba documentada fundamental sobre la calidad del dato.
9	Dato ausente	Usada para indicar que hay datos ausentes.

En el marco del proyecto, la aplicación de este nivel primario facilita el filtrado y / o procesamiento de la calidad de los datos, incluyendo la herencia de los valores de indicador de calidad para las variables derivadas. La calidad de un valor calculado hereda el calificador de calidad más baja de las variables utilizadas en el cálculo. Un ejemplo simple es el cálculo de la densidad a partir de la temperatura (T) y la salinidad (S), entonces si T es de calidad "buena" (1) y S es de calidad "desconocida" (2), entonces la densidad debe heredar la "calidad desconocida"(2).

4.3.6.2. Segundo Nivel

El objetivo del nivel secundario es complementar las banderas de nivel primario al informar los resultados de las pruebas de control de calidad, específicas realizadas y el historial de procesamiento de datos.

El contenido del nivel secundario puede variar en número y descripción y es elegido por aquellos que implementan el esquema, representando información sobre las pruebas de calidad aplicadas e historial de procesamiento de datos.

Tabla III. Pruebas de control de calidad IODE de nivel secundario. Fuente: UNESCO (2013b)

Ejemplos de pruebas de control de calidad / Procesamiento histórico de datos (descripción)
Valor globalmente imposible
Prueba de desviación estándar climática mensual
Comprobación de punta excesiva
Compensación / sesgo excesivos en comparación con un conjunto de datos de referencia
Excesiva incertidumbre en los datos
Relación X / Y inesperada (por ejemplo, estequiometría química o propiedad-propiedad X a T, S, densidad, entre otras)
Exceso de gradiente espacial o comprobación de patrones
Por debajo del límite de detección del método
Valor interpolado (no medido)
Valor de corrección de compensación de datos relativo a un dato de referencia
Opinión de expertos

Para poder llegar a abordar las banderas IODE, es preciso ir de lo particular a lo general, es decir, recurrir a otros sistemas de clasificación previos que permitan proporcionar la mayor cantidad de información sobre los datos respecto a las pruebas de control que se aplican, ya que estos en su descripción son más puntuales a la hora de establecer los códigos, también es conveniente en este punto evaluar cada conjunto de datos de acuerdo al parámetro de interés, ya que tanto para variables físicas, químicas como biológicas los procesos de control pueden variar.

En el marco del proyecto, la aplicación del nivel secundario de calidad se abordó desde dos referencias, las cuales se describen a continuación:

a) Banderas de Calidad del Programa GTSPP

El Programa Global de Perfil de Salinidad de Temperatura (Global Temperature-Salinity Profile Programme GTSPP) propone una serie de procedimientos y pruebas de control para la asignación de banderas de calidad de datos de temperatura y salinidad. La especificidad de este sistema de clasificación permite contemplar ensayos particulares al comportamiento de estos dos parámetros oceanográficos.

El GTSPP maneja todos los datos del perfil de temperatura y salinidad. Esto incluye las observaciones recolectadas usando muestras de agua, instrumentos de perfil continuo tales como CTDs, datos de cadena de termistor y observaciones adquiridas usando termosalinógrafos. Se pretende que los datos lleguen a los centros de procesamiento de datos del programa a través de los canales en tiempo real del Sistema Mundial de Telecomunicaciones (GTS) de la Organización Meteorológica Mundial (OMM) o en modo diferido a través del sistema IODE (UNESCO, 2010).

Los códigos de banderas de calidad utilizados por GTSPP se aprecian en la Tabla IV.

Tabla IV. Banderas de calidad GTSP.

Fuente: UNESCO (2010)

Valor	Definición
0	No se ha asignado ningún control de calidad a este elemento
1	El elemento que aparece es correcto
2	El elemento que aparece es probablemente bueno
3	El elemento que aparece es dudoso
4	El elemento que aparece es erróneo
5	El elemento ha sido cambiado
6 a 8	Reservado para uso futuro
9	Elemento faltante

El Manual de la GTSP recomienda no cambiar los datos. Puesto que estos sólo deben hacerse cuando está claro cuál debería ser el cambio y que si no se hiciera un cambio los datos serían inutilizables. Las descripciones de las pruebas permiten inferir valores para aquellos que no han superado los procedimientos de prueba. La inferencia de un valor correcto se hace a discreción de la persona que hace el control de calidad. Debe basarse en información que no está disponible para el procedimiento de prueba pero que el operador tiene a mano y ayuda a saber cuál debería ser el valor correcto. Los valores deben ser cambiados sólo cuando hay certeza de cuál es el valor correcto. En el caso en que se cambien los valores de los datos, el valor original también se conserva y está disponible para los usuarios o para otras pruebas si es necesario.

Por último, dado que la evaluación de la calidad se comparte entre centros de procesamiento, es posible que los datos marcados como dudosos por un centro sean considerados aceptables por otro o viceversa. Las banderas pueden ser cambiadas por cualquier centro de procesamiento siempre que se mantenga un registro de los cambios.

GTSP propone una serie de pruebas a aplicar para los datos de temperatura, salinidad y densidad a todas las profundidades, es a criterio del analista de los datos discernir entre las pruebas de control cuales aplicar y cuales omitir según las características del conjunto de datos a evaluar. Algunas pruebas pueden ser de no aplicación como la de *punto de congelación* (2.6) si se trata de datos provenientes del trópico, o la de batimetría (2.11) si se trabaja con mediciones a nivel superficial; mientras que otros ensayos son de carácter casi obligatorio como las pruebas de climatología, que aplica a todas los parámetros sin importar los métodos de muestreo o alcance de las mediciones.

PRUEBAS DE CONTROL DE CALIDAD

1. Pruebas de Localización e Identificación
 - 1.1. Identificación de plataforma
 - 1.2. Tiempo imposible
 - 1.3. Localización imposible
 - 1.4. Posición en tierra
 - 1.5. Velocidad imposible
 - 1.6. Sondeo imposible
2. Pruebas de Perfil
 - 2.1. Valores de parámetros globales imposibles
 - 2.2. Valores de parámetros regionales imposibles
 - 2.3. Aumento de la profundidad
 - 2.4. Sobre perfil
 - 2.5. Perfil constante
 - 2.6. Punto de congelación
 - 2.7. Spike
 - 2.8. Picos superiores e inferiores
 - 2.9. Gradiente

- 2.10. Inversión de densidad
- 2.11. Batimetría
- 2.12. Inversión de temperatura
- 3. Pruebas de Climatología
 - 3.1. Estadística estacional de *Levitus*
 - 3.2. Climatología *Emery & Dewar*
 - 3.3. Climatología *Asheville*
 - 3.4. Climatología mensual
- 4. Pruebas de Consistencia de Perfil
 - 4.1. Prueba en cascada
- 5. Inspección Visual
 - 5.1. Travesía de crucero
 - 5.2. Perfil

GTSP Real-Time Quality Control Manual explica en detalle cada una de las pruebas de control de calidad de datos así las reglas y condiciones que deben cumplir los datos para ser evaluados correctamente. (UNESCO, 2010).

b) Banderas de Calidad OceanSites

El Manual de la GTSP es un referente importante en cuanto a asignación de banderas de calidad se refiere, sin embargo no es el único. A modo de adaptación se hizo uso de otro tipo de nomenclatura en banderas de calidad para variables adicionales a la temperatura y la salinidad establecida por el programa OceanSites, el cual es una red mundial de sitios de series de tiempo sostenido de océano abierto, llamadas estaciones de referencia oceánica, que está siendo implementado por una asociación internacional de investigadores. OceanSites proporciona series de tiempo fijo de varias variables físicas, biogeoquímicas y atmosféricas en diferentes lugares alrededor del globo, desde la atmósfera y la superficie del mar hasta el fondo marino. El objetivo del programa es construir y mantener una red global multidisciplinaria para una amplia gama de aplicaciones operacionales y de investigación, incluyendo la variabilidad y predicción del clima, el carbono y los ecosistemas y la validación del estado del océano (OceanSites, 2010). Los códigos de banderas de calidad utilizados por OceanSites se aprecian en la Tabla IV.

Tabla V. Banderas de calidad de la OceanSites.
Fuente: OceanSites (2010)

Bandera de Calidad	Nombre corto	Definición
0	Control de calidad no realizado	-
1	Dato bueno	Todas las pruebas de control de calidad aprobadas
2	Dato probablemente bueno	-
3	Dato malo que es potencialmente corregible	El dato no debe utilizarse sin corrección científica o recalibración.
4	Dato malo	El dato ha fallado en una o más pruebas
5	Valor cambiado	Data may be recovered after transmission error.
6	-	No usado
7	Valor nominal	No se observaron datos pero se informó. Ejemplo: profundidad del objetivo del instrumento.
8	Valor interpolado	Los datos que faltan pueden ser interpolados a partir de datos vecinos en espacio o tiempo
9	Valor ausente	-

5. METODOLOGÍA

Para el cumplimiento de los objetivos propuestos en la presente investigación se desarrollaron una serie de actividades y/o etapas que permitieron gracias a la implementación de una serie de instrumentos, llegar a los resultados de la presente investigación. El desarrollo metodológico se presenta en la Tabla VI.

Tabla VI. Síntesis de la metodología.

SÍNTESIS DE LA METODOLOGÍA			
<i>Objetivos específicos</i>	<i>Etapas / Actividades</i>	<i>Instrumentos y Métodos</i>	<i>Resultados</i>
1. Realizar un análisis exploratorio y descriptivo de las variables oceanográficas tanto físicas como químicas.	1. Inventario de datos de crucero de tipo espacial, temporal y temático para los parámetros oceanográficos TSM, OD y Nutrientes. 2. Preselección de puntos de muestreo y definición de criterios de selección, conformación de bases de datos y asignación de código ONI para las mediciones. 3. Gráficos de caja para la identificación datos atípicos.	1. Hojas de cálculo de Excel. 2. Software ARCGIS 10.1 Herramientas: Georeferencing. 3. Software R versión 3.3.1 paquete Rcmdr. 4. Oceanic Niño Index (ONI)	1. Inventario de datos. 2. Selección de 21 puntos de muestreo para aplicación de controles de calidad. 3. Datos en formato.
2. Analizar los casos de datos faltantes de series de tiempo de variables oceanográficas.	1. Inventario de datos de crucero de tipo espacial, temporal y temático para los parámetros oceanográficos TSM, OD y Nutrientes. 2. Preselección de puntos de muestreo y definición de criterios de selección, conformación de bases de datos y asignación de código ONI para las mediciones. 3. Acorde a la “Guía para la Normalización de Conjuntos de Datos oceanográficos” (Centro Colombiano de Datos Oceanográficos, 2016) se identifican con el código -99999 los datos ausentes en las mediciones inventariadas.	1. Hojas de cálculo de Excel. 2. Software ARCGIS 10.1 Herramientas: Georeferencing. 3. “Guía para la Normalización de Conjuntos de Datos oceanográficos” (Centro Colombiano de Datos Oceanográficos, 2016) 4. Oceanic Niño Index (ONI) (NOAA, 2015a)	1. Inventario de datos de crucero. 2. Selección de 21 puntos de muestreo para aplicación de controles de calidad. 3. Base de datos en formatos normalizados.

SÍNTESIS DE LA METODOLOGÍA

<i>Objetivos específicos</i>	<i>Etapas / Actividades</i>	<i>Instrumentos y Métodos</i>	<i>Resultados</i>
<p>3. Evaluar la calidad de los datos oceanográficos.</p>	<p>1. Búsqueda bibliográfica de guías metodológicas de procesos de control de calidad de datos oceanográficos.</p> <p>2. Etapa 1: Estandarización de formatos, conformación de bases de datos y asignación de código ONI para las mediciones.</p> <p>3. Etapa 2: Identificación de Datos Atípicos mediante gráficos de caja y casos particulares para OD con media recortada.</p> <p>4. Etapa 3: Prueba de Coherencia Temporal para TSM mediante comparación con Atlas de los Datos Oceanográficos de Colombia 1922-2013 (Andrade Amaya et al., 2015) (evaluación a disposición de información).</p> <p>5. Etapa 4: Prueba de Coherencia Espacial por intercomparación con puntos de muestreo vecinos (evaluación a disposición de información por parámetro) (Bernal et al., 2012; Bernal S, Correa, & Rangel M, 2011; Barrios Moreno, 2013)</p> <p>6. Etapa 5: Resultados síntesis de asignación de banderas de calidad de acuerdo a estándares internacionales de la Global Temperature Salinity Profile Programme (GTSP), OceanSites y IODE</p>	<p>1. Hojas de cálculo de Excel.</p> <p>2. Software ARCGIS 10.1 Herramientas: Georeferencing.</p> <p>3. Software R versión 3.3.1 paquete Rcmdr. Gráficos de Caja / Script programado para la ejecución de cada proceso estadístico.</p> <p>4. Atlas de los Datos Oceanográficos de Colombia 1922-2013 (Andrade Amaya et al., 2015)</p> <p>5. Software <i>PASSAGE</i> versión 2.0.11.6 / Coeficiente de Moran / Correlograma Espacial.</p> <p>6. Construcción de Intervalos de Confianza: Clásico, con ajuste de Correlación Espacial.</p> <p>7. Asignación de Banderas de Calidad: TSM (GTSP), OD y Nitratos (Ocean Sites): IODE Dimar</p>	<p>1. Script ejecutable en R para la Etapa 2.</p> <p>2. Script ejecutable en R para la Etapa 3.</p> <p>3. Script ejecutable en R para la Etapa 4.</p> <p>4. Correlogramas para TMS, OD y NO3 en los respectivos puntos de muestreo.</p> <p>5. Propuesta metodológica para asignación de banderas de calidad mediante pruebas de control estadístico.</p> <p>6. Series de tiempo homogeneizadas para TSM, OD y NO3.</p> <p>7. Base de datos con banderas de calidad asignadas.</p>

SÍNTESIS DE LA METODOLOGÍA

Objetivos específicos	Etapas / Actividades	Instrumentos y Métodos	Resultados
<p>4. Explorar la relación entre variables físicas y químicas, como un aporte al tema de monitoreo ambiental.</p>	<p>1. Etapa 2: Comparación entre períodos: Normal, Niño y Niña de parámetros de TSM, OD y Nitratos mediante gráficos de caja.</p> <p>2. Etapa 6: Detección de cambios entre períodos Normal vs. El Niño y Normal vs. La Niña mediante la aplicación de la prueba de Mann-Whitney (Prueba de suma de rangos de Wilcoxon)</p>	<p>1. Hojas de Cálculo de Excel.</p> <p>2. Software ARCGIS 10.1 Herramientas: Georeferencing.</p> <p>3. Software R versión 3.3.1 paquete Rcmdr / Gráficos de caja / Script programado para la ejecución de cada proceso estadístico.</p> <p>4. Software IBM Statistics SPSS 19</p>	<p>1. Script ejecutable en R para la Etapa 6.</p> <p>2. Resultados de la prueba de Mann-Whitney por cada punto de muestreo para TSM, OD y NO3.</p> <p>3. Resultados de cálculo de anomalías por estación y construcción de series de tiempo.</p>
<p>5. Realizar un artículo en colaboración Universidad Distrital y La Dimar.</p>	<p>1. Propuesta a evento académico Seminario Nacional de Ciencias y Tecnologías del Mar 2017, titulado: “Propuesta Metodológica Para La Asignación De Banderas De Calidad De Datos Oceanográficos: Estudio De Caso Temperatura Superficial Del Mar De La Cuenca Pacífica Colombiana” aceptado en modalidad de Póster</p>	<p>1. Síntesis de Resultados de TSM</p> <p>2. Word</p>	<p>Artículo “Propuesta Metodológica Para La Asignación De Banderas De Calidad De Datos Oceanográficos: Estudio De Caso Temperatura Superficial Del Mar De La Cuenca Pacífica Colombiana” primera versión en proceso de revisión. Anexo14.</p>

A continuación se escribe la metodología propuesta.

5.1. Búsqueda bibliográfica

La búsqueda bibliográfica permite contemplar una visión general de las metodologías existentes en cuanto a procesos de control de calidad de datos oceanográficos y similares, tanto a nivel nacional o internacional. Lo cual se traduce en siete diferentes guías metodológicas mostradas en la Tabla VII. De esta forma es posible evaluar y seleccionar elementos propios de cada guía que sean de utilidad para propósitos de la investigación.

Tabla VII. Guías metodológicas revisadas.

Guía	Sección aplicada	Fecha	Autor/ Entidad
Guía para la Normalización de Conjuntos de Datos oceanográficos	Metodología /Registros	2016	Dirección General Marítima Centro Colombiano de Datos Oceanográficos
Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos	Metodología /Registros	2016	Dirección General Marítima Centro Colombiano de Datos Oceanográficos
GTSP Real- Time Quality Control Manual	Quality Flagging /Test Monitoring/ Pre and Post Processing /Quality Control tests / Implementation Details	2010	UNESCO- COI
Guía de Prácticas Climatológicas	Gestión de datos climáticos/ Caracterización del clima a partir de conjuntos de datos /Métodos estadísticos para analizar conjuntos de datos	2011	Organización Meteorológica Mundial (OMM)
Guía para el Procesamiento de Series de Tiempo de Precipitación y temperatura: Estimación de datos faltantes, Detección de Cambios y Homogenización	Homogenización de series de tiempo de precipitación y temperatura	En proceso de publicación	Universidad Distrital Francisco José de Caldas Lorena Lombana González, Néstor Ricardo Bernal Suárez, Juan Sebastián Barrios Moreno Apoyo Técnico del Instituto de Hidrología, Meteorología y Estudios Ambientales - IDEAM
OceanSITES User's Manual	Data_type_and_data code/_Variable quality_control_flag scale/_Overall quality_control procedure_indicator	2011	OceanSites

Ocean Data Standards: Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data/ Volumen 3	Data quality flagging / The quality flag scheme	2013	UNESCO
--	---	------	--------

Adicionalmente se consultaron bases de datos de acceso abierto, de la Universidad Distrital Francisco José de Caldas, de Dimar, así como buscadores académicos como herramienta para la recopilación de información, utilizando como palabras claves: calidad de datos, control de calidad, variables oceanográficas, monitoreo ambiental y homogenización. Donde el enfoque principal fueron artículos científicos y boletines científicos.

5.2. Inventario de Datos

Para alcanzar el primero de los objetivos del presente proyecto de investigación: “Realizar un análisis exploratorio y descriptivo de las variables oceanográficas tanto físicas como químicas” es necesario realizar un inventario de datos para esquematizar los datos disponibles producto de los cruceros oceanográficos información suministrada por Dimar correspondiente al período 1970-2016, para así obtener una visión general de la conformación de los datos y las posibles metodologías de control de calidad que podrían ser aplicados a los mismos de manera eficiente.

Los archivos recibidos están organizados según las especificaciones de la “Guía para la normalización de conjunto de datos oceanográficos”(Centro Colombiano de Datos Oceanográficos, 2016) Debido a que la información de cada año varía en número de cruceros realizados, cantidad de variables oceanográficas medidas y distribución espacial y temporal de la misma, el inventario aborda tres aspectos principales: temático, espacial y temporal.

En el aspecto temático, se tiene en cuenta el tipo de parámetro oceanográfico de interés y la presencia de mediciones en cada crucero. Para objetivos de la investigación se priorizan las mediciones de temperatura superficial del mar (TSM), oxígeno disuelto (OD) y nutrientes, por lo que se revisa una a una cada hoja de registro para saber el conjunto de variables medidas en cada expedición oceanográficas.

A nivel espacial, se considera la carta de ruta de cada crucero y de acuerdo al conjunto de coordenadas de las mediciones es posible mediante el software ArcGis 10.1 usando la herramienta *Georeferencing* ubicar los puntos de muestreo donde se efectuó la captura de datos. Aparte de los puntos de muestreo recorrido por cada crucero en sus respectivas expediciones, se toma como referencia espacial la MALLA ERFEN, como conjunto de coordenadas que propone tanto por ubicación como por distribución unos puntos de muestreo estándar o estaciones como se les puede denominar en ocasiones.

En el aspecto temporal, se utiliza el registro de fecha de las mediciones permite saber el año y mes en que fueron capturados los datos en campo, incluso si se quisiera está disponible la información de la hora de cada medición.

En la Figura 13 se puede observar que al lado de cada variable do de cada variable existe una columna con el encabezado [QF_IODE] la cual corresponde a la bandera de calidad asignada para los datos primarios y su respectivo parámetro oceanográfico. El código “2” se refiere a datos no evaluados y el “9” a datos ausentes.

Fecha de cada medición durante la realización del crucero				Coordenadas geográficas correspondientes a cada medición			Parámetros oceanográficos medidos durante la realización del crucero oceanográfico					
Fecha [aaaa-mm-dd UT-5]	Hora [hh:mm:ss UT-5]	Latitud [deg]	Longitud [deg]	Estacion [#]	Profundidad [dbar]	Temperatura [DEGC]	QF [IODE]	Salinidad [psu]	QF [IODE]	OD [mL O2/L]	QF [IODE]	
ADATAA01	AHMSAA01	ALATGP01	ALONGP01	ACYCAA01	ADEPZZ01	TEMPET01	FLAGIODE	SSALSG01	FLAGIODE	DOXYWITX	FLAGIODE	
1982-11-05	11:30:00	6.50000	-79.66667	1	0	27.90	2	25.518	2	4.86	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	8	28.31	2	26.419	2	4.77	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	15	28.32	2	26.727	2	4.52	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	31	28.12	2	29.253	2	4.43	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	46	26.68	2	33.791	2	4.46	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	61	24.06	2	34.444	2	4.44	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	77	20.47	2	34.569	2	3.13	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	115	18.68	2	34.857	2	2.88	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	153	17.57	2	34.970	2	2.36	2	
1982-11-05	11:30:00	6.50000	-79.66667	1	230	13.18	2	-99999	9	-99999	9	
1982-11-05	11:30:00	6.50000	-79.66667	1	306	9.85	2	-99999	9	-99999	9	
1982-11-08	06:24:00	6.00000	-77.66667	2	0	28.70	2	27.698	2	4.32	2	
1982-11-08	06:24:00	6.00000	-77.66667	2	8	28.40	2	27.819	2	4.76	2	
1982-11-08	06:24:00	6.00000	-77.66667	2	16	28.20	2	29.996	2	4.58	2	

Figura 12. Ejemplo de formato de datos suministrados por Dimar.

Este procedimiento permite identificar la información necesaria para la posterior construcción del Metadato que acompaña el conjunto de datos producto del procesamiento estadístico para la asignación de banderas de calidad. El metadato sigue los estándares establecidos por el *Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos* (CECOLDO & DIMAR, 2016).

5.3. Preselección

El inventario de datos esquematiza los datos presentes en los cruceros oceanográficos, sin embargo para propósitos del desarrollo del presente proyecto y para dar cumplimiento al tercer objetivo de “*Evaluar la calidad de los datos oceanográficos*” se prioriza la información para contar con una cantidad de datos que permita ensayar las metodologías estadísticas para el control de calidad **pero que no exceda las capacidades de procesamiento del investigador**.

A partir de los datos primarios suministrados por la Dimar la preselección plantea dos criterios, de tipo espacial y temático:

- Criterio espacial: A partir de las 44 estaciones de la Malla ERFEN, se seleccionan los puntos de muestreo cuyas coordenadas coincidan con las ubicaciones estándar propuestas por la Malla ERFEN. No se tiene en cuenta la información de estaciones vecinas.
- Criterio temático: Presencia de medición de la variable de interés (TSM, OD, nutrientes) en los registros, en la mayor cantidad de años y de cruceros.

La Guía para la Normalización de Conjuntos de Datos Oceanográficos de Cecoldo nombra una serie de errores frecuentes que pueden ocurrir a la hora de normalizar datos; la GTSP Real-Time Quality Control Manual por su arte recomienda pruebas de control para verificar la información base asociada con la captura de los datos, tales pruebas son:

- Pruebas de localización e identificación
 - ❖ Identificación de plataforma
 - ❖ Dato/tiempo imposible
 - ❖ Localización imposible
 - ❖ Posición en tierra
 - ❖ Velocidad imposible
 - ❖ Sondeo imposible

Cabe anotar que los datos suministrados por la Dimar debidamente formalizados han pasado por un proceso de revisión equivalente a las pruebas anteriormente mencionadas, por lo que se tiene seguridad de las coordenadas, fechas y datos registrados en cada uno de los archivos debido a que siguen los parámetros establecidos por la Guía para la Normalización de Conjuntos de Datos Oceanográficos (Centro Colombiano de Datos Oceanográficos, 2016).

5.4. Etapa 1: Estandarización de Formatos

Una vez hecha la preselección se sabe los años, cruceros y puntos geográficos de los que se obtienen datos de los parámetros de interés (temperatura, OD, nutrientes). La “Guía para el Procesamiento de Series de Tiempo de Precipitación y Temperatura” (Lombana González et al., n.d.) sugiere que los datos estén organizados de forma matricial y por estaciones de muestreo de acuerdo a la ubicación geográfica en la que fueron capturados. Para lo cual la información debe ser extraída de los archivos primarios y organizada por estaciones en documentos nuevos, este proceso supone la construcción de una base de datos a partir de la cual se aplican las metodologías estadísticas de control de calidad. Para construir la base de datos, se siguen los siguientes pasos:

1. Extraer la información de los datos primarios, de cada estación correspondiente a los parámetros fisicoquímicos oceanográficos de cada registro, de cada año y cada crucero a todas las profundidades disponibles. Cada archivo creado corresponde a una estación según las coordenadas de los puntos de muestreo propuestos por la Malla ERFEN.
2. Seleccionar los datos correspondientes al parámetro de interés (temperatura, OD, nutrientes) en el conjunto de datos y a la profundidad deseada, para hacer el correspondiente tratamiento estadístico. En este paso se sugiere seleccionar trabajar todas las variables a una sola profundidad (0m) para reducir la cantidad de datos a procesar.

Asignación de código ONI

Una vez los datos se encuentran en un formato estándar, se crea una nueva columna en las hojas de cálculo en la cual se asigna el código ONI (Oceanic Niño Index) de acuerdo a la fecha de cada medición (Figura 14).

	A	B	C	D	E	F	G	H	I	J
	ONI	Fecha [aaaa-mm-dd UT-5]	Mes	Hora [hh:mm:ss UT-5]	Latitud [deg]	Longitud [deg]	Estacion [#]	Profundidad [dbar]	Temperatura [DEGC]	QF [IODE]
1										
2	NINO	1976-10-07	Octubre	05:30:00	4.50000	-78.50000	39	0	27.40	2
3	NINO	1977-10-30	Octubre	11:12:00	4.50000	-78.50000	23	0	27.3	2
4	NINO	1982-11-10	Noviembre	17:00:00	4.50000	-78.50000	21	0	28.50	2
5	NORMAL	1986-06-24	Junio	12:30:00	4.50000	-78.50000	30	0	26.64	2
6	NINO	1987-10	Octubre	-99999	4.50000	-78.50000	21	0	26.73	2
7	NINA	1988-11-13	Noviembre	22:00:00	4.50000	-78.50000	21	0	26.73	2
8	NORMAL	1993-04-18	Abril	16:20:00	4.50000	-78.50000	21	0	27.4	2
9	NORMAL	1993-09-30	Septiembre	01:35:00	4.50000	-78.50000	21	0	26.57	2
10	NINO	1994-11-09	Noviembre	15:00:00	4.50000	-78.50000	21	0	26.34	2

Figura 13. Ejemplo de documento con asignación de código ONI Etapa 1.

La clasificación puede ser alfabética o numérica. Para evitar introducir caracteres especiales en las celdas se usa el término de forma simplificada con un código numérico, así: NORMAL (0); NINO (1) y NINA (-1) (Figura 15).

Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
1980	0.6	0.5	0.3	0.4	0.5	0.5	0.3	0.2	0	0.1	0.1	0
1981	-0.2	-0.4	-0.4	-0.3	-0.2	-0.3	-0.3	-0.3	-0.2	-0.1	-0.1	0
1982	0	0.1	0.2	0.5	0.6	0.7	0.8	1.0	1.5	1.9	2.1	2.1
1983	2.1	1.8	1.5	1.2	1.0	0.7	0.3	0	-0.3	-0.6	-0.8	-0.8
1984	-0.5	-0.3	-0.3	-0.4	-0.4	-0.4	-0.3	-0.2	-0.3	-0.6	-0.9	-1.1
1985	-0.9	-0.7	-0.7	-0.7	-0.7	-0.6	-0.4	-0.4	-0.4	-0.3	-0.2	-0.3
1986	-0.4	-0.4	-0.3	-0.2	-0.1	0	0.2	0.4	0.7	0.9	1.0	1.1
1987	1.1	1.2	1.1	1.0	0.9	1.1	1.4	1.6	1.6	1.4	1.2	1.1
1988	0.8	0.5	0.1	-0.3	-0.8	-1.2	-1.2	-1.1	-1.2	-1.4	-1.7	-1.8
1989	-1.6	-1.4	-1.1	-0.9	-0.6	-0.4	-0.3	-0.3	-0.3	-0.3	-0.2	-0.1

Figura 14. Valores del Oceanic Niño Index por meses y años.
Fuente: NOAA (2015a)

Para definir el valor del índice que va de acuerdo a la fecha de cada medición, de las tres celdas en las que aparece el mes de interés se selecciona el de la mitad. Los valores en negro corresponden a periodos normales, los rojos a épocas de El Niño y los azules a periodos La Niña.

5.5. Etapa 2: Identificación de Datos Atípicos

La identificación de datos atípicos o valores extremos que forman parte del grueso de las observaciones, pueden provocar que el rango de variabilidad de los datos resulte engañoso (Organización Meteorológica Mundial OMM, 2011) En este caso la *GTSP Real-Time Quality Control Manual* recomienda una serie de pruebas que permiten identificar fácilmente valores extremos de acuerdo a estándares fijos para cada parámetro oceanográfico. En investigaciones previas como *Diseño e implementación de una herramienta computacional para el control de calidad y validación de datos oceanográficos* (Sánchez Escobar, 2006) se recurre a pruebas similares para la identificación de datos por fuera de los rangos globales imposibles para cada parámetro. Así:

- Pruebas de perfil
 - ❖ Valores de parámetros globales imposibles
 - ❖ Valores de parámetros regionales imposibles

Paralelamente a la identificación de datos atípicos también se hace una exploración de la distribución de los datos. La forma de presentar los datos para obtener una apreciación cualitativa de lo que los datos revelan; así las observaciones ordenadas pueden entonces presentarse gráficamente o como un cuadro, en el que se pondrán de relieve algunas características, tales como los valores extremos y el rango. Para este caso se hace mediante diagramas de caja. El software R a través del paquete Rcmdr permite construir gráficos de caja de un conjunto de datos organizado por grupos, ya sea por mes, años, o periodos Normal, Niño o Niña.

Para este paso es necesario construir un gráfico de caja, el cual es una representación visual que describe características importantes de un conjunto de datos, tales como el centro, la dispersión, la simetría o asimetría y la identificación de observaciones atípicas. El diagrama de caja representa los tres cuartiles, y los valores mínimo y máximo de los datos sobre un rectángulo (caja), alineado horizontal o verticalmente (Rangel, n.d.)

El gráfico de caja se construye a partir de los cuartiles del conjunto de datos de interés. Los cuartiles son los valores que dividen a una colección de datos que ordenados en forma creciente,

en cuatro partes iguales. De esta manera, existen tres cuartiles que se denotan Q1, Q2 y Q3. El gráfico de caja representa los cuartiles y las observaciones que están entre 1.5 y 3 veces el rango intercuartílico (RQ) a partir de los extremos del rectángulo reciben el nombre de valores atípicos. Las observaciones que están más allá de tres veces el rango intercuartílico (RQ) a partir de los lados del rectángulo se conocen como valores atípicos extremos, identificados con algún símbolo.

El rectángulo o caja representa el 50% de los datos que particularmente están ubicados en la zona central de la distribución. La caja representa el cuerpo de la distribución y los bigotes sus colas (Figura 16).

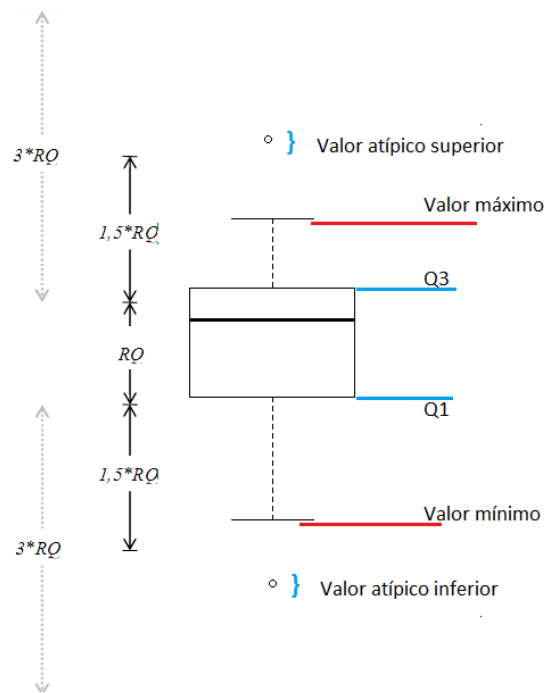


Figura 15. Explicación partes gráfico de caja.
Modificado de Rangel, n.d.

La representación gráfica proporciona una idea general de la distribución de los datos y los valores atípicos que se pueden encontrar. Matemáticamente se puede identificar mediante el cálculo de los percentiles del conjunto de datos. Los percentiles son valores que dividen a los datos ordenados de forma creciente, en cien partes iguales. Existen noventa y nueve percentiles que se denotan por P1, P2, ... , P99. Entre dos percentiles consecutivos se encuentra el 1% de los datos (Rangel, n.d.).

5.6. Etapa 3: Prueba de Coherencia Temporal

La Organización Mundial Meteorológica recomienda pruebas de coherencia temporal para examinar la variación de un elemento en el tiempo. Ya que los conjuntos de datos climatológicos muestran una importante correlación en serie. Se hace un control comparando la observación en cuestión con la anterior y la posterior en el ordenamiento (Organización Meteorológica Mundial OMM, 2011). La GTSP Real- Time Quality Control Manual establece pruebas de control las cuales son:

- Pruebas de climatología
 - ❖ Estadística estacional de *Levitus*

- ❖ Climatología *Emery and Deward*
- ❖ Climatología *Ashville*
- ❖ Climatología mensual *Levitus*

Las pruebas incluyen en hacer una comparación con Atlas que contienen información estadística de los comportamientos estacionales del océano en diferentes partes del mundo. Para propósitos de esta investigación se hace uso del Atlas de los Datos Oceanográficos de Colombia 1922-2013 (Andrade Amaya et al., 2015), como referente que permita hacer dicha comparación.

La comparación se efectúa a través de pruebas de tolerancia que establecen límites superiores e inferiores sobre los posibles valores de una variable. En otras palabras, se establecen intervalos de confianza, para este caso se aplica el intervalo clásico t-Student (Ospina Botero, 2001). Y una segunda alternativa es la construcción del intervalo es posible por la determinación del coeficiente de correlación espacial y rango de vecindad en Software *PASSAGE* realizado por Bernal S, Correa, & Rangel M, (2011) en trabajo sobre datos climatológicos en “Homogenización de series de tiempo mensuales de precipitación”.

5.7. Etapa 4: Prueba de Coherencia Espacial

La Organización Mundial Meteorológica recomienda pruebas de coherencia espacial que consta de comparar cada observación con observaciones efectuadas al mismo tiempo en otras estaciones de la zona. Cada observación puede compararse con las que cabría esperar en el emplazamiento teniendo en cuenta las observaciones de las denominadas estaciones vecinas (Organización Meteorológica Mundial OMM, 2011). Paralelamente la “Guía para el Procesamiento de Series de Tiempo de Precipitación y Temperatura” (Lombana González et al., n.d.) utiliza una metodología para la homogenización de series de tiempo de precipitación y temperatura desde el punto de vista espacial en donde escoge pares de estaciones vecinas para la elaboración de una curva de dobles masas teniendo en cuenta los criterios de selección específicos para este tipo de variables climatológicas, dicha metodología fue propuesta y aplicada para una investigación previa denominada Aporte al Proceso de Toma de Decisiones en la Planificación Ambiental de la Región Climática del Bajo Magdalena con Base en Análisis Estadístico Empleando una Metodología Para la Homogenización de Series Mensuales de Precipitación (Barrios Moreno, 2013). Así, para el procesamiento de los datos oceanográficos se aplica el principio de las estaciones vecinas como forma de comparar los datos; nuevamente se recurre a una prueba de tolerancia mediante la construcción de intervalos de confianza, pero en lugar del intervalo clásico se aplica el intervalo propuesto por Cressie (1993) (Li, 2009), que tiene en cuenta la correlación espacial entre puntos de muestreo en diferentes ubicaciones geográficas.

Con el software R a través del paquete Rcmdr es posible construir los intervalos de confianza y ejecutar la prueba de tolerancia respectiva.

Correlograma espacial en PASSAGE

El software *PASSAGE* en su versión 2.0.11.6 permite calcular la correlación espacial de un conjunto de datos a través de la construcción de correlogramas espaciales y calculando el Índice de Moran, como la propuesta que hace Bernal S et al.,(2011) en la “Homogenización de series de tiempo mensuales de precipitación” ; de esta forma se utiliza para establecer un rango de vecindad entre los puntos de muestreo en función del parámetro oceanográfico de interés; así igual que para las estaciones meteorológicas en tierra como lo propone la “Guía para el Procesamiento de Series de Tiempo de Precipitación y Temperatura” (Lombana González et al., n.d.), es posible asociar un grupo de puntos de muestreo y hacer comparaciones y

verificaciones de datos a nivel espacial, como propone y aplica para tres regiones climatológicas en dicha investigación.

5.8. Etapa 5: Resultados Síntesis de Asignación de Banderas de Calidad

Para el cumplimiento de tercer objetivo “ Evaluar la calidad de los datos oceanográficos.” desde la etapa 2 hasta la etapa 4, se aplican pruebas estadísticas de control que asignan banderas de calidad a cada dato al final de cada proceso. En esta etapa se evalúa el resultado de las etapas por las que las mediciones fueron procesadas y se asigna una bandera de calidad final; primero utilizando nomenclaturas de banderas de calidad intermedias como la GTSP Real-Time Quality Control Manual (UNESCO, 2010) y la OceanSITES User’s Manual de acuerdo al tipo de parámetro oceanográfico a evaluar y finalmente haciendo al equivalencia a la banderas establecida por la IODE según la “Guía para la normalización de conjunto de datos oceanográficos” (Centro Colombiano de Datos Oceanográficos, 2016)

5.9. Etapa 6: Detección de Cambios Entre Períodos Normal vs. El Niño y Normal vs. La Niña.

A diferencia de las etapas anteriores la etapa 6 no designa banderas de calidad a los datos sino que sirve a propósitos de análisis. De cada punto de muestreo se pretende someter los datos que han aprobado exitosamente las pruebas estadísticas de las etapas anteriores. La prueba permite demostrar que existen diferencias entre grupos independientes con variables cuantitativas que tienen libre distribución, en este caso de cada estación los grupos corresponden a los períodos Normal Vs. Niña y Normal Vs. Niño, grupos de datos de la misma ubicación geográfica, correspondiente a las variables de TSM, oxígeno disuelto y nitratos. Para lo cual se utiliza la prueba de Mann-Whitney. Esta prueba tiene su base en la diferencia de rango y es la contraparte de la t de Student que se emplea en las variables cuantitativas con distribución normal (Rivas Ruiz, Moreno Palacios, & O. Talavera, 2013) que permite confirmar o descartar una hipótesis previamente establecida.

Paralelamente a la detección de cambios se estiman las anomalías entre las épocas El Niño y La Niña comparado con el período Normal. Desde el concepto climatológico de la anomalía como la diferencia entre el valor registrado de la una variable durante un periodo de tiempo determinado y su promedio; así, las fluctuaciones observadas en el clima durante períodos de tiempo relativamente cortos pueden registrar valores por encima (excesos) o por debajo de lo normal (déficits) (IDEAM, 2014a).

Para calcular la anomalía se toman las mediciones (de acuerdo al parámetro oceanográfico) durante época EL Niño/La Niña y se resta a cada una el promedio de las mediciones durante época Normal, por estación. Se debe tener en cuenta que los promedios son calculados a partir del conjunto de datos aprobados por el control de calidad de las anteriores etapas.

$$\text{Anomalía} = x_i \text{ El Niño} - \bar{x} \text{ Normal}$$

$$\text{Anomalía} = x_i \text{ La Niña} - \bar{x} \text{ Normal}$$

Las anomalías negativas representan una alteración “negativa” (comportamiento de la variable) en comparación con la media en condiciones normales; mientras que una anomalía positiva representa una alteración “positiva” superávit (comportamiento de la variable) en comparación con la media en condiciones normales, que es el valor de referencia para la comparación. Los valores de la anomalía se expresan en las unidades en las que se miden los diferentes parámetros (IDEAM, 2014a). Posteriormente, se seleccionan los datos aprobados (con bandera de calidad

la misma, esta sección hace una descripción y análisis de la organización y estructura de los datos.

Estructura de los datos

La información correspondiente está organizada por carpetas de acuerdo a cada año de muestreo. En cada carpeta existe un metadato, un informe de crucero y una matriz de datos presentada en una hoja de registro. Para cada año la información sobre los cruceros que se efectuaron y los muestreos registrados para ese período pueden presentarse de alguna de las siguientes maneras:

1. Un único crucero con una sola hoja de registro.
2. Un único crucero con dos hojas de registro.
3. Dos cruceros, cada uno con una hoja de registro diferente.
4. Dos cruceros, cada uno con una o más hojas de registro diferentes.

Cada matriz de datos comprende un conjunto de mediciones realizadas en campo, de acuerdo al tipo de variables oceanográficas medidas, las hojas de registro que las contienen pueden ser de tipo: físico, químico o fisicoquímico.

Organización de los datos

Después de efectuar el primer sondeo de los datos, la información existente presenta:

- 36 años no consecutivos de datos (1970-2016).
- 57 cruceros realizados.
- 74 hojas de registro.

Se debe tener en cuenta que el número de registros y de cruceros realizados por año varía según la necesidad y disponibilidad de los buques para ese momento. Los cruceros efectuados en un mismo año no necesariamente midieron las mismas variables oceanográficas y no comprenden el mismo periodo de tiempo ni las mismas estaciones de muestreo, así como los tipos de hojas de registro varían.

La Tabla VIII muestra la disposición de la información por año y número de crucero, así como los tipos de hojas de registro que hay registrados para cada periodo, encontrándose un total de 74 registros distribuidos en 57 cruceros, efectuados en 36 años de mediciones.

Tabla VIII. Tipos de Datos suministrados por la Dimar para el período 1970-2016.

AÑO	CRUCERO	TIPO DE REGISTRO	AÑO	CRUCERO	TIPO DE REGISTRO
1970	Pacífico I	Fisicoquímico	2001	Pacífico XXXIV_ERFENXXX II	Fisicoquímico
1972	Pacífico II	Fisicoquímico		Pacífico XXXV_ERFENXXXI II	Fisicoquímico
	Pacífico III	Fisicoquímico			Químico
1975	Pacífico IV_ ERFEN I	Fisicoquímico	2002	Pacífico XXXVI_ERFENXXX IV	Fisicoquímico
1976	Pacífico V_ ERFEN II	Fisicoquímico	2003	Pacífico XXXVII_ERFENXX XV	Fisicoquímico
	Pacífico VI_ ERFEN III	Fisicoquímico		Pacífico XXXVIII	Fisicoquímico

1977	Pacífico_VII	Fisicoquímico	2004	Pacífico XXXIX	Pacífico_ Fisicoquímico
1978	ERFEN_IV	Fisicoquímico			Malpelo_ Fisicoquímico
1982	Pacífico VIII_ERFEN V	Fisicoquímico	2005	Pacífico XL_ERFENXXXVIII	Fisicoquímico
1986	Pacífico	Fisicoquímico		Pacífico XLI_ERFENXXXIX	Fisicoquímico
1987	Pacífico X_ERFEN_VII	Fisicoquímico	2006	Pacífico XLII	Químico
1988	Pacífico	Fisicoquímico		Pacífico XLIII	Fisicoquímico
	Pacífico_XII_ERFEN IX	Fisicoquímico	2007	CPC XLV	Físico
1989	Pacífico_XIII_ERFEN X	Fisicoquímico		Pacífico XLIV	Fisicoquímico
	Pacífico XIV_ERFEN_XI	Fisicoquímico	2008	Pacífico XLVI	Físico
1990	Pacífico XV_ERFEN_ XII	Fisicoquímico		Pacífico XLVII	Fisicoquímico
	Pacífico_ XVI_ERFEN_ XIII	Fisicoquímico	2009	CPC XLVII	Fisicoquímico
1991	Pacífico_ XVII_ERFENXIV	Fisicoquímico		Pacífico XLVIII	Fisicoquímico
	Pacífico XVIII_ERFEN_ XV	Fisicoquímico	2010	CPC 48	Químico
1992	Pacífico XIX_ERFEN_ XVI	Fisicoquímico		CPC49	Físico
	ERFEN_ XVII	Fisicoquímico	2011	CPC50_CPCSXIV	Químico
1993	ERFEN_ XVIII	Físico		2012	CPC51_CPCSXV
	ERFEN_ XX	Fisicoquímico	CPC52_CPCSXVI		Químico
1994	ERFEN_ XXI	Fisicoquímico	2013	CPC53_CPCSXVII	Físico
	Pacífico_XXIV_ERFE N_XXII	Fisicoquímico			Pacífico LIV
1996	Pacífico_XXV_ERFEN_ XXIII	Fisicoquímico	2014	2016	Físico
	Pacífico_XXVI_ERFE N_XXIV	Fisicoquímico			Físico
1997	Pacífico XVII_ERFEN_ XXV	Fisicoquímico	2015	2016	Químico
	Pacífico XVIII_ERFEN_ XXVI	Fisicoquímico			Físico
1998	Pacífico XXIX_ERFENXXVII	Fisicoquímico	2016	2016	Químico
	Pacífico XXX_ERFENXXVIII	Fisicoquímico			Físico
1999	Pacífico XXXI	Fisicoquímico	2016	2016	Químico
		Fisicoquímico			Físico
2000	Pacífico XXXII_ERFENXXX	Fisicoquímico	2016	2016	Físico
	Pacífico XXXIII_ERFENXXXI	Fisicoquímico			Físico

Los 36 años de datos recolectados presentan en total 18 variables o parámetros oceanográficos medidos. La toma de las muestras no es homogénea a lo largo de los años, cada crucero cuenta con un número y tipo de variable diferente en función de las necesidades de medición para el momento de la expedición.

Existen variables cuyo muestreo es constante a lo largo de los años. Las variables de tipo físico como la temperatura y la salinidad se presentan de manera continua en los años de muestreo, mientras que otras variables, como las de tipo químico las hay con menor frecuencia en los registros como en el caso de los nutrientes.

Para propósitos de este proyecto la TSM, el oxígeno disuelto y los nutrientes son las variables objeto de estudio. La TSM se halló en el 85% de los registros inventariados, al ser una variable física no se incluye en los registros de tipo químico, aún así en varios cruceros los dos tipos de hojas están simultáneamente.

El oxígeno disuelto es una variable química con un menor porcentaje de aparición en los muestreos (37%) es un parámetro del que se tiene registro desde 1982, más no presenta periodos continuos de medición. Se tienen 28 registros diferentes distribuidos en 19 años de información, siendo la última década (2006-2016) los que presentan la mejor continuidad a lo largo del tiempo.

Los nutrientes son el caso más particular de información, el primer registro data de 1988, y no vuelven a haber mediciones hasta 2005; los nutrientes comprenden el conjunto de variables de interés del presente estudio con los mayores vacíos en la continuidad de los registros. Existen ocho años de mediciones para el grupo de nutrientes distribuidos en su mayoría en la década 2006-2016. La disponibilidad de registros por variable se muestra en la Tabla IX.

Tabla IX. Variables oceanográficas presentes en los datos 1970-2016 suministrados por la Dimar.

Variables oceanográficas presentes en los datos 1970-2016			
TIPO	VARIABLE	Nº de Registros presentes	Porcentaje de aparición en los registros
FÍSICO	Profundidad	74	100,00
	Temperatura	63	85,14
	Salinidad	58	78,38
	Densidad	27	36,49
	Velocidad del sonido	6	8,11
	Presión	20	27,03
	Velocidad de la corriente	4	5,41
	Dirección de la corriente	3	4,05
	Conductividad	23	31,08
QUÍMICO	Oxígeno Disuelto	28	37,84
	pH	4	5,41
	NH4	8	10,81
	NO2	8	10,81
	NO3	8	10,81
	PO4	8	10,81
	SiO3	8	10,81
BIOLÓGICO	Clorofila a	5	6,76
	Fluorescencia	3	4,05

Distribución temporal de los datos

Los datos anuales representan muestreos tomados en periodos de 15 a 25 días aproximadamente, es decir, que se trabaja con datos diarios, sin embargo dependiendo del año hay muestreos continuos mientras que otros presentan registros con intervalos de varios días. Cada muestreo fue tomado en diferentes épocas del año según disponibilidad de las embarcaciones.

Para cada periodo de información, las épocas en de muestreo varían entre los diferentes meses. En los primeros años de registro (1970-1982) no se evidencia un patrón que haya sido tenido en cuenta en el muestreo a nivel temporal. Para el periodo 1986-2000 se puede observar dos temporadas diferentes en el año que suelen repetirse en la mayoría de los casos, las cuales son febrero-abril y septiembre- octubre aunque, para esta franja de tiempo rara vez años consecutivos poseen datos en los mismos meses, mientras que los últimos quince años de registros 2001-2016 presentan mediciones más frecuentes efectuadas durante los meses de septiembre-octubre y en años consecutivos.

Las fechas de inicio y finalización de las mediciones para cada matriz de datos, crucero y año se observan en la Tabla X.

Tabla X. Distribución temporal de los datos 1970-2016 suministrados por la Dimar.

AÑO	CRUCERO	REGISTROS	MES
1970	Pacífico I	Fisicoquímico	Mar 13- Mar 29
1972	Pacífico II	Fisicoquímico	Feb 20- Mar 8
	Pacífico III	Fisicoquímico	Jun17- Jun17
1975	Pacífico IV_ ERFEN I	Fisicoquímico	Ene 31- Feb 17
		Fisicoquímico	Feb 25- Mar 11
1976	Pacífico V_ ERFEN II	Fisicoquímico	Abr 14- May 17
	Pacífico VI_ ERFEN III	Fisicoquímico	Sept 27- Oct 29
1977	Pacífico_ VII	Fisicoquímico	Oct 22-Nov 2
1978	ERFEN_ IV	Fisicoquímico	Nov 29- Dic 12
1982	Pacífico VIII_ ERFEN V	Fisicoquímico	Nov 5 - Dic 12
1986	Pacífico_ IX	Fisicoquímico	May 1- Jun 25
1987	Pacífico X_ ERFEN_ VII	Fisicoquímico	Mar 31- Abr 14
	Pacífico_ XI	Fisicoquímico	Octubre
1988	Pacífico_ XII_ ERFEN IX	Fisicoquímico	Mar 7- Mar 31
		Químico	Mar 7- Mar 31
	Pacífico_ XIII_ ERFEN X	Fisicoquímico	No 7- Nov 23
1989	Pacífico XIV_ ERFEN_ XI	Fisicoquímico	Mar 27- Abr 11
	Pacífico XV_ ERFEN_ XII	Fisicoquímico	Ago 7 - Ago 21
1990	Pacífico_ XVI_ ERFEN_ XIII	Fisicoquímico	Mar 27- Abr 8
	Pacífico_ XVII_ ERFENXIV	Fisicoquímico	Sept 24 - Oct 2
1991	Pacífico XVIII_ ERFEN_ XV	Fisicoquímico	Sept 20- Oct 8
	Pacífico XIX_ ERFEN_ XVI	Fisicoquímico	Mar 23- Abr 6
1992	ERFEN_ XVII	Fisicoquímico	Mar 30- Abr 3**
	ERFEN_ XVIII	Fisicoquímico	Oct **
1993	ERFEN_ XX	Físico	Abr 15- Abr 28
	ERFEN_ XXI	Fisicoquímico	Sept 29 –Oct 20

1994	Pacífico_XXIV_ERFEN_XXII	Fisicoquímico	Nov 1 - Nov 11
1996	Pacífico_XXV_ERFEN_XXIII	Fisicoquímico	Jun 5- Jun 20
	Pacífico_XXVI_ERFEN_XXIV	Fisicoquímico	Oct 14- Oct 26
1997	Pacífico XXVII_ERFEN_XXV	Fisicoquímico	May 25- Abr 4
	Pacífico XXVIII_ERFEN_XXVI	Fisicoquímico	Nov 27- Dic 9
1998	Pacífico XXIX_ERFENXXVII	Fisicoquímico	May 9- Jun 2
	Pacífico XXX_ERFENXXXVIII	Fisicoquímico	Oct 16- Oct 29
		Fisicoquímico	Oct 16- Oct 29
1999	Pacífico XXXI	Fisicoquímico	May 6- May 25
		Fisicoquímico	May 6- May 25
2000	Pacífico XXXII_ERFENXXX	Fisicoquímico	May 1- May 30
	Pacífico XXXIII_ERFENXXXI	Fisicoquímico	Nov 19- Dic 6
2001	Pacífico XXXIV_ERFENXXXII	Fisicoquímico	Jun 23-Jul 12
		Fisicoquímico	Jun 23-Jul 12
	Pacífico XXXV_ERFENXXXIII	Fisicoquímico	Ago 27- Sept 15
		Químico	Ago 28- Sept 8
2002	Pacífico XXXVI_ERFENXXXIV	Fisicoquímico	Feb 24- Mar 14
	Pacífico XXXVII_ERFENXXXV	Fisicoquímico	Sept 3- Sept 22
2003	Pacífico XXXVIII	Fisicoquímico	Sept 1- Sept 21
2004	Pacífico XXXIX	Pacífico_ Fisicoquímico	Sept 10- Oct 8
		Malpelo_ Fisicoquímico	Sept 26 **
2005	Pacífico XL_ERFENXXXVIII	Fisicoquímico	Jul 9 - Jul 28
	Pacífico XLI_ERFENXXXIX	Fisicoquímico	Sept 1- Sept 24
		Químico	Sept 1- Sept 24
2006	Pacífico XLII	Fisicoquímico	Mar 3- Mar 26
	Pacífico XLIII	Fisicoquímico	Sept 15- Sept 17
		Físico	Sept 15- Sept 21
2007	CPC XLV	Físico	Ago30- Sept 28
		Fisicoquímico	Sept 4- Sept 28
	Pacífico XLIV	Fisicoquímico	Ene24 -Feb 17
		Físico	Ene 28- Feb 18
2008	Pacífico XLVI	Fisicoquímico	Mar 9- Mar 28
	Pacífico XLVII	Fisicoquímico	Sept 6- Sept 26
2009	CPC XLVII	Fisicoquímico	Sept 16- Oct 7
	Pacífico XLVIII	Fisicoquímico	Mar 2- Mar 26
		Químico	Mar 2- Mar 26
2010	CPC 48	Físico	Mar 7- Mar 9
	CPC49	Físico	Sept 17- Oct 6
		Químico	Sept 17- Oct 6
2011	CPC50_CPCSXIV	Físico	Oct 20- Nov 7
		Químico	Oct 20- Nov 8
2012	CPC51_CPCSXV	Físico	Sept 5 -Sept 18
		Químico	Sept 5 -Sept 18
2013	CPC52_CPCSXVI	Físico	Oct 8- Oct 28
		Químico	Oct 8- Oct 28

2014	CPC53_CPCSXVII	Físico	Oct 4- Oct 28
		Químico	Oct 4- Oct 28
2016	Pacífico LIV	Fisicoquímico	Ene 2- Ene 20

**Los datos primarios no especifica las fechas en las que se hicieron las mediciones, los registros solo muestran el mes en el que se efectuó.

Distribución espacial de los datos

La distribución espacial de las estaciones de muestreo para cada año de mediciones se evaluó mediante las coordenadas de los registros de cada crucero con la ayuda del software ArcGis 10.1 *Geographic Coordinate System GCS_WGS_1984*. Para ello, se utilizó como marco de referencia las estaciones establecidas por el Programa para el Estudio Regional del Fenómeno El Niño en el Pacífico Sudeste (ERFEN) del cual Colombia hace parte, y cuyas coordenadas fueron suministradas por la Dimar, y se ubicaron geográficamente en el Pacífico Colombiano. Como se observa en la Figura 19.

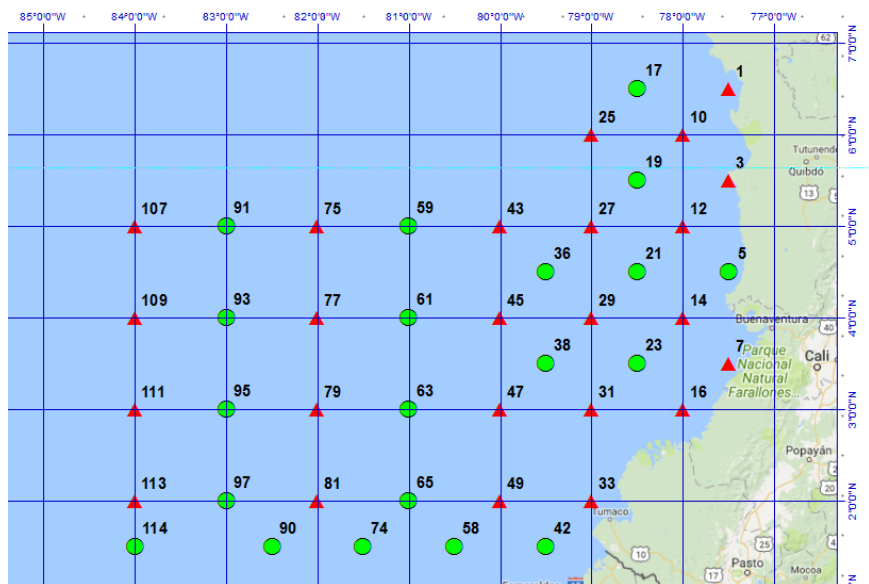


Figura 18. Malla ERFEN Cuenca Pacífica Colombiana. Fuente: Dimar.

La Malla ERFEN de la Cuenca Pacífica Colombiana cuenta con 44 estaciones (20 de tipo Físico y 24 de tipo Físico-Químico-Biológico) numeradas de 1 a 114. Comprende un área desde las zonas costeras hasta mar abierto a 84° longitud oeste (W) y en un rango de latitud entre 1° a 7° Norte (N). La mayoría de las estaciones conservan una distancia establecida entre ellas. La distancia mínima es de 58,51 Km y la máxima es de 111 Km a la redonda, aproximadamente. Se tiene en cuenta las posiciones de las estaciones de la Malla ERFEN como puntos de referencia oficiales, para con base en ellos evaluar la distribución de los muestreos a lo largo de los años.

Se elaboró una comparación de cada registro y la ubicación de sus estaciones con la Malla ERFEN, buscando identificar la coincidencia entre las estaciones de los cruceros y las oficiales. Se identificaron tres escenarios diferentes:

1. Las estaciones coinciden geográficamente con la malla de referencia.

2. Las estaciones no se ajustan exactamente con la malla de referencia, sin embargo existe una proximidad espacial que podría ser de utilidad para identificar estaciones vecinas que sirvieran a propósitos del análisis de los datos.
3. No hay estaciones, ni que coincidan espacialmente con la malla de referencia, ni en un rango de cercanía para identificar estaciones vecinas.

Hay que tener en cuenta que no en todos los cruceros tomaron muestras en todas las estaciones, ni en el mismo grupo de estaciones cada año. Para cada período se observaron estaciones adicionales o faltantes al compararlo con malla ERFEN. La distribución espacial de las estaciones muestreadas con los puntos de referencia de la malla ERFEN puede apreciarse en el banco de imágenes del Anexo 11.

El Protocolo ERFEN fue firmado y adoptado por Colombia en 1992; los datos recolectados recopilados antes de esa fecha presentan un gran número de estaciones extras, cuya distribución geográfica es muy diferente a la sugerida en la Malla ERFEN.

Para 1970, 1972, 1986, 1987 los puntos de muestreo se encuentran más cercanos a la costa. Los cruceros realizados en 1975, 1976, 1977, 1978, 1982 y desde 1988 hasta 2000 describen unas estaciones más mar abierto en forma de cuadrícula que año a año van aumentando su ajuste con la Malla ERFEN; en este último año todavía se observan un gran número de estaciones adicionales ubicadas más cerca entre ellas y próximas a la costa, ya que en la mayoría de los cruceros no pasan de los 82° de longitud oeste (W). Finalmente desde 2001 hasta 2016 la ubicación geográfica de las estaciones presentan una coincidencia casi total con los puntos de la malla ERFEN, y el número de estaciones adicionales o faltantes disminuye en algunos años.

NOTA: Para ver la ubicación geográfica de los puntos de muestreo por año y crucero ver **Anexo 11_Banco Imágenes 1970_2016**. Allí se encuentra la distribución de las estaciones de acuerdo al año de realización del crucero en contraste con los puntos de muestreo estándar establecido por la MALLA ERFEN:

6.3. Priorización de Estaciones de Muestreo

A partir del análisis exploratorio y descriptivo de los datos de crucero suministrados por la Dimar, se realizó una priorización de estaciones de muestreo a trabajar, con el objetivo de contar con una base homogénea de datos de características similares que permitieran probar adecuadamente cada una de las etapas de la metodología propuesta. De 44 estaciones pertenecientes a la Malla ERFEN, se seleccionaron los puntos que cumplieran con los siguientes criterios:

- a. Coincidencia de las coordenadas de la estación con las propuestas en la Malla ERFEN. No se tiene en cuenta la información de estaciones vecinas o coordenadas cercanas.
- b. Presencia de la variable de interés (temperatura, oxígeno disuelto, nutrientes) en las hojas de registro, en la mayor cantidad de años y de cruceros.

Se seleccionaron las estaciones que cumplieran con los anteriores criterios con un percentil por encima del 55%, es decir, de 74 hojas de registro, aquellas estaciones que tuvieran más de 40 coincidencias tanto espacial como temáticamente. Finalmente, para las variables de temperatura y oxígeno disuelto se seleccionaron 21 estaciones (Figura 20); mientras que para nutrientes 16 (Figura 21). Las coordenadas de estas estaciones prioritarias se muestran a continuación en la Tabla XI.

Tabla XI. Coordenadas de las 21 estaciones prioritarias.

Coordenadas de estaciones prioritarias		
Código de la Estación	Latitud	Longitud
10	6.00 N	-78.00 W
12	5.00 N	-78.00 W
14	4.00 N	-78.00 W
16	3.00 N	-78.00 W
17	6.50 N	-78.50 W
19	5.50 N	-78.50 W
21	4.50 N	-78.50 W
25	6.00 N	-79.00 W
27	5.00 N	-79.00W
29	4.00 N	-79.00 W
31	3.00 N	-79.00 W
33	2.00 N	-79.00 W
43	5.00 N	-80.00 W
45	4.00 N	-80.00 W
47	3.00 N	-80.00 W
61	4.00 N	-81.00 W
63	3.00 N	-81.00 W
65	2.00 N	-81.00 W
77	4.00 N	-82.00 W
79	3.00 N	-82.00 W
81	2.00 N	-82.00 W

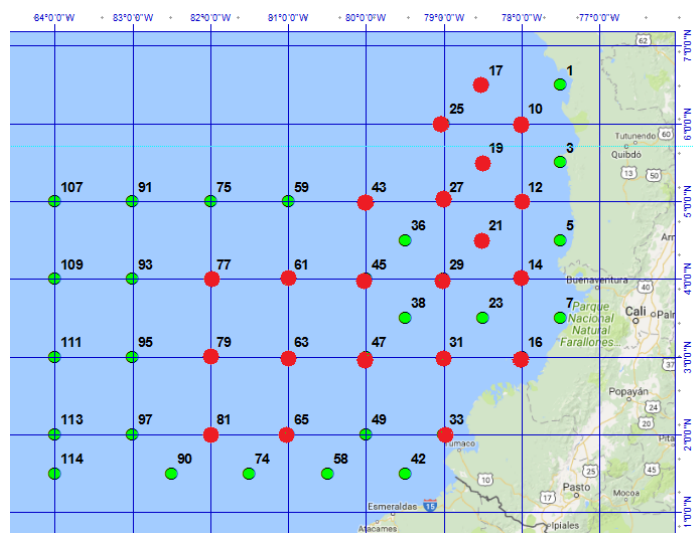


Figura 19. Estaciones prioritarias (señaladas en rojo) para las variables de temperatura y oxígeno disuelto.

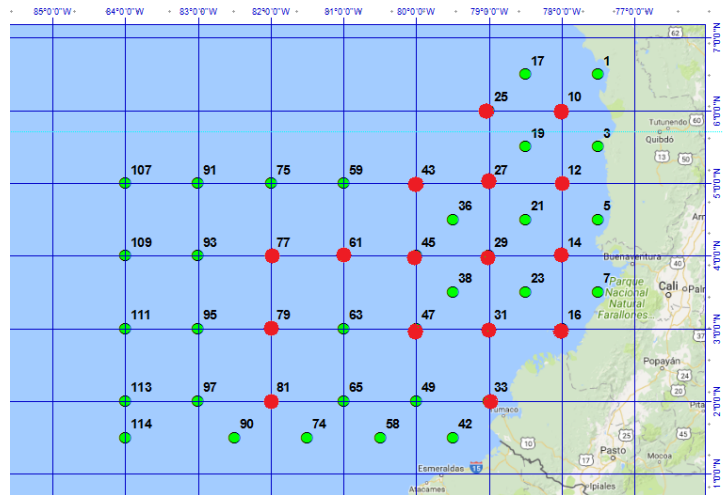


Figura 20. Estaciones prioritarias (señaladas en rojo) para la variable de nutrientes.

6.4. Etapa 1- Estandarización de Formatos

Una vez seleccionadas las estaciones a trabajar, es importante establecer un formato de fácil interpretación para el procesamiento de los datos en R. Este debe preservar la información adjunta a cada dato, como lo es fecha, hora, coordenadas, profundidad. Además de información adicional importante para el análisis posterior. Para pasar de los datos primarios al formato apto para R, se realizan los siguientes pasos:

1. Ya que los datos primarios están organizados por años y para el procesamiento estadístico se requiere un formato determinado por ubicación geográfica, se procede a hacer un traslado de los datos de años a estaciones de muestreo. Consta de extraer la información de los datos primarios, de cada estación distribuida en los diferentes años y cruceros, correspondiente a los parámetros fisicoquímicos oceanográficos de cada hoja de datos a todas las profundidades disponibles. Cada archivo creado corresponde a una estación según las coordenadas de los puntos de muestreo propuestos por la Malla ERFEN (producto de la preselección). La organización de los archivos nombrados está de acuerdo a la Guía de Normalización.
2. Se crea una nueva columna denominada ONI por Oceanic Niño Index, en la cual de acuerdo a la fecha de medición se clasificaron los datos identificándolos como pertenecientes a periodos de Niño, Niña o Normal. (se denominan periodos NORMAL (0), NINO (1) y NINA(-1) para evitar introducir caracteres especiales se puede utilizar tanto de forma alfabética como numérica)
3. Seleccionar por columnas los datos correspondientes al parámetro de interés (temperatura, oxígeno disuelto, nutrientes) en el conjunto de datos y a la profundidad deseada. Para propósitos de esta investigación se seleccionaron datos superficiales, es decir aquellos medido a 0 m de profundidad.
4. Se debe añadir una columna denominada “Mes” y de acuerdo a la fecha de medición de cada dato asignar el mes correspondiente. Esta asignación puede ser por letras o números, según como se prefiera.
5. Simplificar el encabezado de los archivos en Excel. La Guía de Normalización establece unas características que debe cumplir el encabezado de cada columna, sin embargo para propósitos de procesamiento de los datos en R y hacer más sencilla la visualización de los mismos, es preferible trabajar con encabezados cortos y simples en cada columna en el archivo de Excel que posteriormente será exportado a R, también es

requerido modificar el formato de las celdas que contienen la información de Fecha y Hora a tipo general (reemplazando “/” por “_” y “:” por “.”) de esta forma la información asociada a cada dato no se ve alterada por el cambio de formato. Es importante que todos los encabezados de todos los archivos sean iguales así como el formato de celdas de los documentos, esto para evitar problemas en R más adelante.

6.5. Etapa 2- Identificación De Datos Atípicos

La Etapa 2 consta de la detección de los datos atípicos en los conjuntos de datos, y así poder identificar los valores extremos según cada variable de interés. Esta etapa tiene en cuenta la coherencia temática de los valores a tratar, sobre los cuales se procura mantener rangos acordes al comportamiento de los parámetros oceanográficos según los estándares globales y regionales.

Para este paso se construye un gráfico de caja, como representación visual que describe características importantes de un conjunto de datos. (Rangel, n.d.) Luego de observar el gráfico de caja es necesario calcular los percentiles 25, 50 y 75. Por debajo del P_{25} se encuentra el 25% de los datos de menor valor y por encima del P_{75} están los datos de mayor valor. El P_{50} indica la mediana de los datos. Con estos valores se puede establecer el rango intercuartílico (RQ) para fijar los límites máximos y mínimos de cada conjunto de datos. Así:

$$\text{Limite de detección atípico inferior} = P_{25} - 1,5 (P_{75} - P_{25})$$

$$\text{Limite de detección atípico superior} = P_{75} + 1,5 (P_{75} - P_{25})$$

Al comparar cada (X_i) dato con el respectivo límite superior e inferior del mismo conjunto de datos al que pertenece se puede clasificar así:

$$\text{Si } X_i > P_{75} + 1,5 (P_{75} - P_{25}) \text{ entonces } X_i = \text{atípico superior}$$

$$\text{Si } X_i < P_{25} - 1,5 (P_{75} - P_{25}) \text{ entonces } X_i = \text{atípico inferior}$$

Si se estima que los valores pueden alejarse más de 1,5 del rango intercuartílico (RQ), se puede usar el atípico extremo, descrito como:

$$\text{Si } X_i < P_{25} - 3 (P_{75} - P_{25}) \text{ entonces } X_i = \text{atípico inferior extremo}$$

$$\text{Si } X_i > P_{75} + 3 (P_{75} - P_{25}) \text{ entonces } X_i = \text{atípico superior extremo}$$

Ya que se pretende evaluar la coherencia temática de los datos de una misma estación, se hace una clasificación de los mismos por períodos Normal, Niño y Niña, ya que hay evidencia que dichos eventos influyen en el rango normal de los valores de ciertos parámetros oceanográficos. Para la primera asignación de banderas de calidad se siguen los siguientes pasos:

Parte 1

1. Dividir el conjunto de datos de cada estación en tres subgrupos (Normal, Niño, Niña) según el código ONI adicionado en la anterior etapa.
2. Realizar un primer gráfico de caja a modo exploratorio para evaluar la presencia de datos extremos en el conjunto de datos (Figura 22).

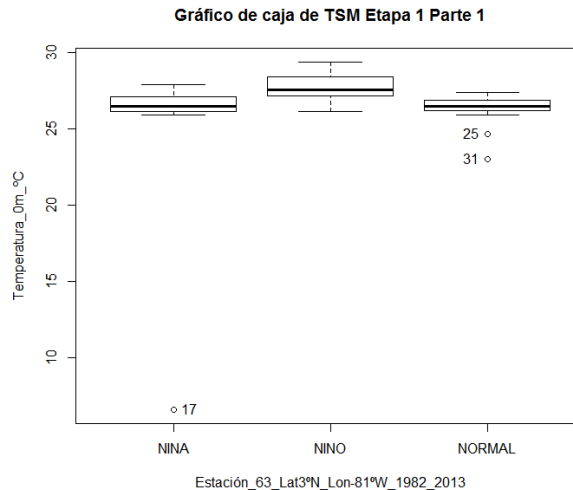


Figura 21. Ejemplo gráfico de caja para una estación con datos de TSM. Etapa 2 parte 1

3. Calcular el rango intercuartílico para cada subconjunto mediante la estimación de percentiles.
4. Comparar cada dato de cada grupo con los valores establecidos en el intervalo.
5. Asignar las banderas de calidad correspondientes a cada datos según la clasificación establecida por la GTSP (Figura 23), así:
 - Los datos ausentes identificados con -9999,0000 en los documentos primarios se clasifican inmediatamente con “9”
 - Los valores que se encuentre por fuera del intervalo establecido para cada subconjunto son asignados con “4”
 - Los datos que cumplen con la prueba estadística aplicada se les asigna “1”.

Flag Description	ODY	GTSP	ARGO	SEADATANET	ESEAS	WOD	WODSTATION	WOCEBOTTLE	WOCECTD	WOCEAMPLE	QARTOD	BODC	PANGAEA	SMHI	OceanSITES	IODE
no quality control (QC) was performed	1	0	0	0	0	0	0	2	2	2	0	Q	*	blank	0	2
QC was performed; good data	0	1	1	1	1	0	0	2	2	2	3	blank	blank	blank	1	1
QC was performed; probably good data	0	2	2	2	1	0	0	2	2	2	3	blank	blank	blank	2	1
QC was performed; probably bad data	4	3	3	3	3	4	3	3	3	7	2	K	?	?	3	3
QC was performed; bad data	8	4	4	4	4	4	3	4	4	7	1	K	/	B	4	4
the value was changed as a result of QC	1	5	5	5	2	0	0	2	2	2	3	R	*		5	2
the value is missing	1	9	9	9	9	0	0	5	5	5	9	N	*	B	9	9

Figura 22. Equivalencias de banderas de calidad GTSP. Fuente: Schlitzer, Wegener, & Instituto for Poland and Marine Research (2013)

** Para el caso de nitratos se utiliza una nomenclatura de banderas de calidad diferente debido a las características del parámetro oceanográfico. La asignación de banderas que se emplea es establecida por OceanSites (Figura 24), así:

- Los datos ausentes identificados con -9999,0000 en los documentos primarios se clasifican inmediatamente con “9”
- Los valores que se encuentre por fuera del intervalo establecido para cada subconjunto son asignados con “3”
- Los datos que cumplen con la prueba estadística aplicada se les asigna “1”.

Flag Description	<u>ODV</u>	<u>GTSP</u>	<u>ARGO</u>	<u>SEADATANET</u>	<u>ESEAS</u>	<u>WOD</u>	<u>WODSTATION</u>	<u>WOCEBOTTLE</u>	<u>WOCECTD</u>	<u>WOCEAMPLE</u>	<u>QARTOD</u>	<u>BODC</u>	<u>PANGAEA</u>	<u>SMHI</u>	<u>OceanSITES</u>	<u>IODE</u>
no QC was performed	1	0	0	0	0	0	0	2	2	2	0	Q	*	blank	0	2
good data	0	1	1	1	1	0	0	2	2	2	3	blank	blank	blank	1	1
probably good data	1	2	2	2	1	0	0	2	2	2	3	blank	blank	blank	2	1
bad data that are potentially correctable	4	3	3	3	3	4	3	3	3	7	2	K	?	?	3	3
bad data	8	4	4	4	3	4	3	4	4	7	1	K	/	B	4	4
value changed	1	5	5	5	0	0	0	2	2	2	3	R	*	blank	5	2
nominal value	1	0	0	0	0	0	0	2	2	2	3	R	*	blank	7	2
interpolated value	1	0	0	8	2	0	0	2	2	2	3	R	*		8	2
missing value	1	9	9	9	9	0	0	5	5	5	9	N	*	B	9	9

Figura 23. Equivalencias de banderas de calidad OceanSites Etapa 1. Fuente: Schlitzer, Wegener, & Instituto for Poland and Marine Research (2013)

- Con los datos ya clasificados los tres grupos se unifican en un solo conjunto de datos que pasará a la siguiente etapa para aplicar otras pruebas estadísticas.

Parte 2

Esta etapa consta de verificar la efectividad del método para detectar los datos atípicos. Para ello:

- El conjunto de datos con banderas de calidad ya asignadas se filtra para seleccionar únicamente los valores marcados con “1”, es decir, aquellos dentro del intervalo de confianza.
- Se elabora nuevamente un nuevo gráfico de caja, que permite inspeccionar la distribución de los valores sin presencia de datos atípicos (Figura 25).

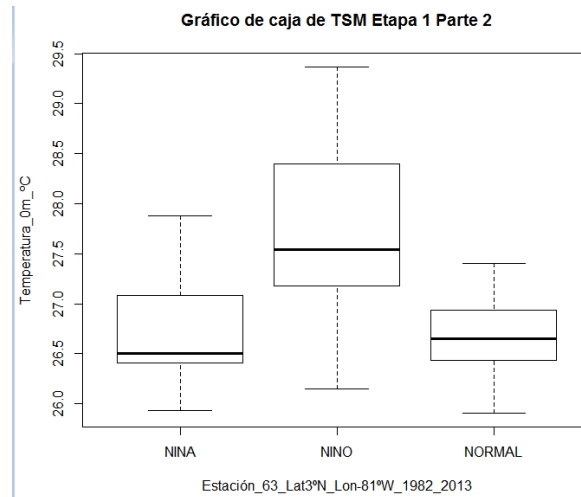


Figura 24. Ejemplo gráfico de caja para una estación sin presencia de datos atípicos de TSM. Etapa 2 parte 2

El archivo filtrado, aquel que solo contienen los datos marcados con bandera de calidad “1” libre de valores atípicos se exporta como un documento que va a ser utilizado en la Etapa 3.

NOTA: Para esta etapa se escribió un SCRIPT en el software R que ejecuta las operaciones estadísticas que permiten hacer la clasificación de cada conjunto de datos con la primera asignación de banderas de calidad. Al final de esta fase se obtiene un documento que contiene

una columna denominada QF_ETAPA1 la cual clasifica cada dato con alguno de los códigos de banderas de calidad correspondientes a las reglas establecidas según las pruebas estadísticas anteriormente descritas.

****Para ver el SCRIPT de ejecución y su explicación ver**

ANEXO 1_SCRIPT_co_tesisdimarUD_R_temperatura_0m_etapa2

ANEXO 8_SCRIPT_co_tesisdimarUD_R_NO3_0m_etapa2

6.6. Etapa 2- Identificación de Datos Atípicos (Caso de Oxígeno Disuelto)

En ocasiones de acuerdo a la distribución de los datos, puede que el método por percentiles no permita identificar datos que temáticamente salen de los rangos de variabilidad del parámetro oceanográfico de interés. Mediante una inspección visual y consultando la bibliografía es posible observar datos atípicos que mediante gráficos de caja no es posible detectar; como es el caso de los datos para oxígeno disuelto. La metodología aplicada para TSM fue ensayada para OD, sin embargo valores que eran claramente identificados como atípicos no fueron detectados por el método de percentiles, por ello para este tipo de casos, ya sea por la conformación del conjunto de datos, su distribución o incluso el tamaño de la muestra se plantea esta metodología alternativa para la identificación de datos atípicos (A.Colubi, A.Lubiano, & P.Terán, n.d.).

La metodología alternativa hace uso de la media recortada para identificar datos atípicos. La media recortada, plantea el cálculo de la media eliminando la posibilidad de que esté afectada por los valores extremos excluyendo del estudio sistemáticamente los valores altos y los bajos. Así, la media recortada al 10 % es la media aritmética del 90 % central de los valores de la muestra. Es decir, se eliminan el 10 % de los valores extremos (el 5 % de los valores más altos y el 5 % de los más bajos) y se calcula la media de los que quedan (A.Colubi et al., n.d.).

La mayor ventaja es que al eliminar el 10 % los valores extremos, en caso de que haya valores atípicos, estos no influyen en su cálculo y en este sentido es mejor que la media. Sin embargo, eliminar individuos también supone un inconveniente, porque sabemos que las muestras son más informativas cuanto más grandes sean y omitir datos supone una pérdida de conocimiento.

Se debe tener en cuenta el tamaño de la muestra. Si el tamaño de muestra es suficientemente grande, comparando los valores de la media aritmética y de la media recortada se puede saber si hay valores muy extremos que hacen que la media aritmética no sea una medida robusta a la presencia de dichos valores, lo anterior sin necesidad de hacer análisis exploratorios adicionales. Si se diferencian poco, significa que el 10 % de los valores que se excluyeron no aportaban mucho en el cálculo de la media, pero si se diferencian mucho significa que tenían mucho peso y que la media puede estar distorsionada por ellos (A.Colubi et al., n.d.). Así en los ensayos hechos para los datos de oxígeno disuelto se identificó que la media recortada al 10% no presentaba diferencia a la media que se calcula normalmente y los datos atípicos seguían sin ser detectados. Por ello se elevó el porcentaje de exclusión de datos hasta encontrar aquel que revelara una diferencia real a la media normal y permitiera identificar los datos atípicos que afectaban los estadísticos del conjunto de datos para esta variable. Finalmente se encontró que al 25% los datos atípicos eran detectados y la media se volvía más central.

Se debe tener en cuenta que utilizar este estadístico como medida de tendencia central cuando la muestra es suficientemente grande y hay valores atípicos que hacen que la media este muy desvirtuada.

La asignación de banderas de calidad en este caso también depende de la construcción de rangos con los cuales comparar los datos. Ya que se utiliza la media recortada, para mantener los cálculos estadísticos centrales sin que se vean influenciados por los datos atípicos presentes en el conjunto de datos, se usa la desviación media, que es el cálculo de la desviación estándar a partir de la media recortada de los datos en lugar de la media normal, lo cual permite hacer el cálculo sin que se vea influenciado por valores extremos en el conjunto de datos.

Es necesario establecer rangos para hacer la comparación de los datos, así al igual que en la metodología anterior los límites quedan establecidos por:

$$\begin{aligned} \text{Limite de detección atípico inferior} \\ = \text{media recortada} - 1,5 (\text{desviación estandar con el 75\% de los datos}) \end{aligned}$$

$$\begin{aligned} \text{Limite de detección atípico superior} \\ = \text{media recortada} + 1,5 (\text{desviación estandar con el 75\% de los datos}) \end{aligned}$$

Al comparar cada (X_i) dato con el respectivo límite superior e inferior del mismo conjunto de datos al que pertenece se puede clasificar así:

$$\begin{aligned} \text{Si } X_i > \text{media recortada} + 1,5 (\text{desviación estandar con el 75\% de los datos}) \text{ entonces } X_i \\ = \text{atípico superior} \end{aligned}$$

$$\begin{aligned} \text{Si } X_i < \text{media recortada} - 1,5 (\text{desviación estandar con el 75\% de los datos}) \text{ entonces } X_i \\ = \text{atípico inferior} \end{aligned}$$

Ya que se pretende evaluar la coherencia temática de los datos de una misma estación, se hace una clasificación de los mismos por períodos Normal, Niño y Niña, ya que se sabe que tales eventos influyen en el rango normal de los valores de ciertos parámetros oceanográficos. Para la primera asignación de banderas de calidad con la metodología alternativa ensayada en datos de oxígeno disuelto se siguen los siguientes pasos:

Parte 1

1. Dividir el conjunto de datos de cada estación en tres subgrupos (Normal, Niño, Niña) según el código ONI adicionado en la anterior etapa.
2. Realizar un primer gráfico de caja a modo exploratorio para evaluar la presencia de datos extremos en el conjunto de datos (Figura 26).

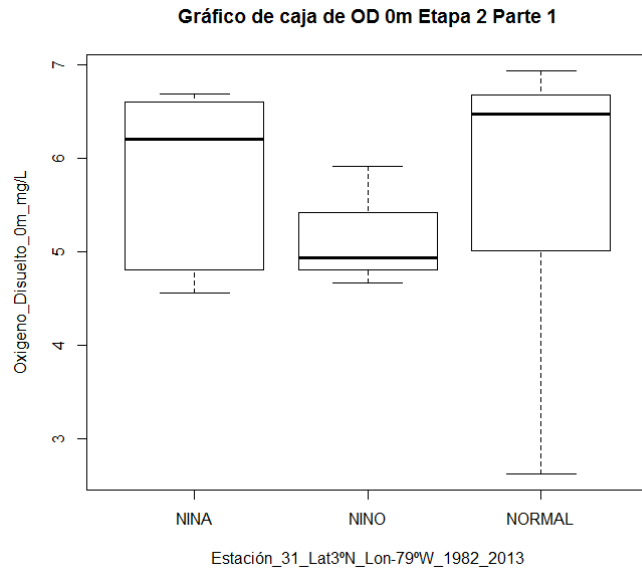


Figura 25. Ejemplo gráfico de caja para una estación con datos de oxígeno disuelto. Etapa 2 parte 1

3. Calcular la media recortada al 25% para cada subconjunto.
4. Calcular la desviación estándar con los datos que se emplearon para obtener la media recortada para cada subconjunto de datos.
5. Determinación de intervalos a partir de la desviación media y la media recortada.
6. Comparar cada dato de cada grupo con los valores establecidos en el intervalo.
7. Asignar las banderas de calidad correspondientes a cada dato según la clasificación establecida por la OceanSites (Figura 27), para el caso de oxígeno disuelto, así:
 - Los datos ausentes identificados con -9999,0000 en los documentos primarios se clasifican inmediatamente con “9”
 - Los valores que se encuentren por fuera del intervalo establecido para cada subconjunto son asignados con “4”
 - Los datos que cumplen con la prueba estadística aplicada se les asigna “1”.

Flag Description	<u>ODV</u>	<u>GTSP</u>	<u>ARGO</u>	<u>SEADATANET</u>	<u>ESFAS</u>	<u>WOD</u>	<u>WODSTATION</u>	<u>WOCEROTILE</u>	<u>WOCFCID</u>	<u>WOCESAMPLE</u>	<u>QARTOD</u>	<u>BODC</u>	<u>PANGAEA</u>	<u>SMHI</u>	<u>OceanSITES</u>	<u>LODE</u>
no QC was performed	1	0	0	0	0	0	0	2	2	2	0	Q	*	blank	0	2
good data	0	1	1	1	1	0	0	2	2	2	3	blank	blank	blank	1	1
probably good data	1	2	2	2	1	0	0	2	2	2	3	blank	blank	blank	2	1
bad data that are potentially correctable	4	3	3	3	3	4	3	3	3	7	2	K	?	?	3	3
bad data	8	4	4	4	3	4	3	4	4	7	1	K	/	B	4	4
value changed	1	5	5	5	0	0	0	2	2	2	3	R	*	blank	5	2
nominal value	1	0	0	0	0	0	0	2	2	2	3	R	*	blank	7	2
interpolated value	1	0	0	8	2	0	0	2	2	2	3	R	*		8	2
missing value	1	9	9	9	9	0	0	5	5	5	9	N	*	B	9	9

Figura 26. Equivalencias de banderas de calidad OceanSites para la Etapa 1. Fuente: Schlitzer et al. (2013)

8. Con los datos ya clasificados los tres grupos se unifican en un sólo conjunto de datos que pasará a la siguiente etapa para aplicar otras pruebas estadísticas.

Parte 2

Esta etapa consta de verificar la efectividad del método para detectar los datos atípicos. Para ello:

1. El conjunto de datos con banderas de calidad ya asignadas se filtra para seleccionar únicamente los valores marcados con “1”, es decir, aquellos dentro del intervalo de confianza.
2. Se elabora nuevamente un nuevo gráfico de caja, que permite inspeccionar la distribución de los valores sin presencia de datos atípicos (Figura 28).

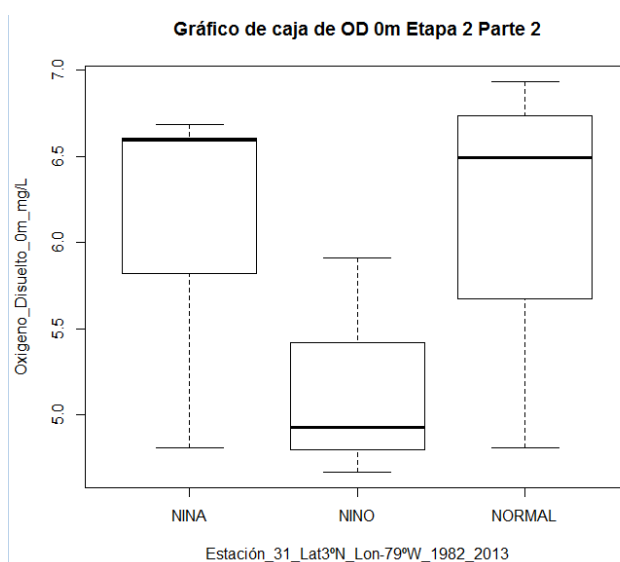


Figura 27. Ejemplo gráfico de caja para una estación sin presencia de datos atípicos para oxígeno disuelto. Etapa 2 parte 2

El archivo filtrado, aquel que solo contienen los datos marcados con bandera de calidad “1” libre de valores atípicos se exporta como un documento que va a ser utilizado en la Etapa 3.

NOTA: Para esta etapa se escribió un SCRIPT en el software R que ejecuta las operaciones estadísticas que permiten hacer la clasificación de cada conjunto de datos con la primera asignación de banderas de calidad. Al final de esta fase se obtiene un documento que contiene una columna denominada QF_ETAPA2 la cual clasifica cada dato con alguno de los códigos de banderas de calidad correspondientes a las reglas establecidas según las pruebas estadísticas anteriormente descritas.

**Para ver el SCRIPT de ejecución y su explicación ver

ANEXO 5_SCRIPT_co_tesisdimarUD_R_OD_0m_etapa2

6.7. Etapa 3- Prueba de Coherencia Temporal

La etapa 3 consta de evaluar la coherencia temporal de los datos de cruceo a partir de datos de referencia de otras fuentes. Las condiciones temporales y espaciales deben coincidir con las de los datos a los que se pretende hacer la asignación de banderas de calidad. Esta etapa aborda principalmente la temporalidad en la que los diferentes parámetros son evaluados y procurar la mayor coherencia posible entre estos y otra mediciones. Lo ideal es contar con una base de datos que ya haya sido evaluada previamente, que comprenda la misma zona geográfica y el mismo período de tiempo. Para el caso de TSM se recurre a la información proporcionada por el

Atlas de los Datos Oceanográficos de Colombia 1922-2013. Para las otras variables esta etapa no fue aplicada debido a la usencia de información de referencia para hacer la comparación.

Para la aplicación de esta etapa se extrajo la información del Atlas coincidente con el grupo de coordenadas de cada estación de muestreo, y se utilizó el promedio mensual de cada mes desde 1922 hasta 2013. Esta información ha se ser almacenada en un archivo debidamente nombrado. En él se observa la información correspondiente al promedio mensual de cada mes para temperatura, salinidad y densidad, a profundidades estándar. En este caso se hizo una selección de la información de interés, temperatura a 0 m de profundidad con la desviación estándar, error estándar y número de observaciones como parte de la información que proporciona el Atlas. Una vez organizada la información se procede a construir el intervalo de confianza con el método t-Student que tiene en cuenta el tamaño de la muestra para construir dicho intervalo.

Se construye un intervalo para cada mes, para cada estación, es decir que en total cada ubicación de muestreo cuenta con doce intervalos de confianza diferentes. El intervalo de confianza recomendado para esta etapa es el la **distribución t de Student**, que se define como:

$$\bar{x} \pm t_{n-1} ee[\bar{x}]$$

Donde \bar{x} = media aritmética del conjunto de datos,

$ee[\bar{x}]$ = el error estándar

n = tamaño de muestra

t = valor asociado a los grados de libertad de acuerdo al nivel de confianza y al tamaño de la muestra n establecido en la tabla *t de Student*. (Mendenhall et al., 2010).

Tabla XII. Información de Atlas de los Datos Oceanográficos de Colombia correspondiente a la estación 25, coordenadas 6°N, -79°W. Fuente: Andrade Amaya et al (2015).

Mes	Temperatura Media Mensual °C	Desviación Estándar	Error Estándar	Nº de Observaciones	Intervalo de confianza	Límite Superior	Límite Inferior
Enero	27.1306	0.9719	0.2696	13	0.8234	27.9540	26.3072
Febrero	26.9512	0.8085	0.1452	31	0.3993	27.3505	26.5519
Marzo	26.5298	1.1616	0.2371	24	0.6656	27.1954	25.8642
Abril	27.8394	0.9703	0.2353	17	0.6874	28.5268	27.1520
Mayo	27.9989	0.8719	0.1859	22	0.5263	28.5252	27.4726
Junio	27.7957	0.5995	0.2998	4	1.7508	29.5465	26.0449
Julio	27.2417	2.1337	1.0668	4	6.2314	33.4731	21.0103
Agosto	26.7975	2.2904	0.8657	7	3.2095	30.0070	23.5880
Septiembre	27.6577	0.3516	0.0677	27	0.1880	27.8457	27.4697
Octubre	27.6077	0.3821	0.1444	7	0.5354	28.1431	27.0723
Noviembre	27.2348	0.5567	0.1968	8	0.6888	27.9236	26.5460
Diciembre	27.0330	NaN	0.0000	1	NaN	NaN	NaN

Se debe tener en cuenta que no siempre para todos los puntos de muestreo hay información disponible de todos los meses del año. En este caso los datos de crucero a evaluar correspondientes a ese mes, quedan sin comparar y depende de las demás etapas de calidad asignar las banderas correspondientes.

La representación gráfica del intervalo calculado para una estación se observa en la Figura 29.

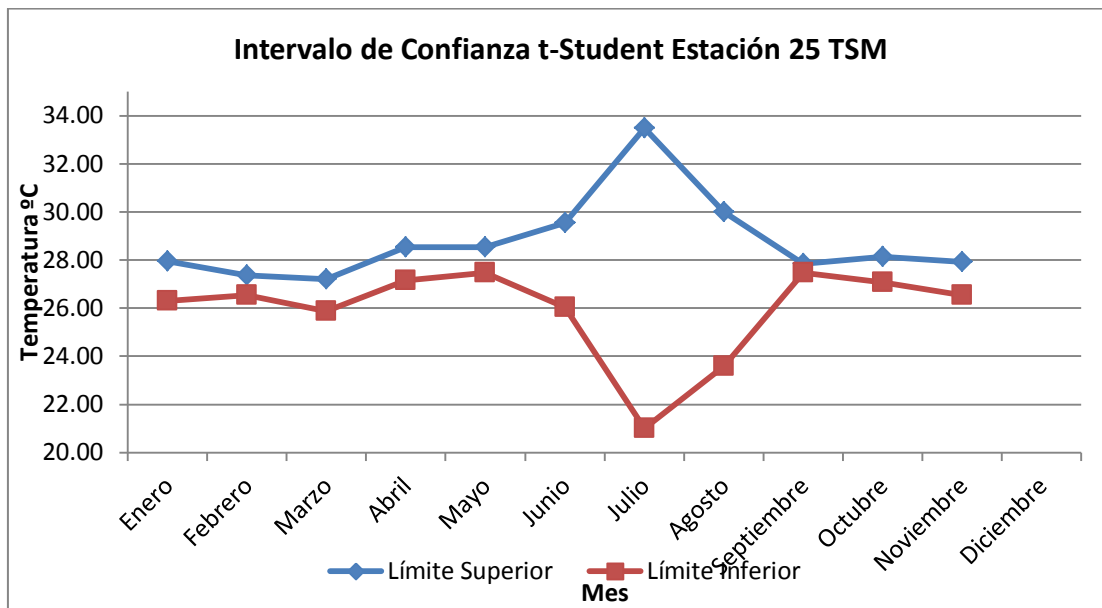


Figura 28. Ejemplo Intervalos de confianza t-Student para una estación.

Debido a que en ocasiones el método puede generar intervalos muy cerrados y se corre riesgo de descartar una gran cantidad de datos, se sugiere tomar un nivel de confianza del 96%. Los intervalos de confianza estrechos suponen una alta exigencia en la precisión de los datos a evaluar, y a la vez se ignora la influencia que el efecto de eventos como El Niño y La Niña puede tener sobre los mismos, igualmente la información suministradas por el Atlas no tienen en cuenta tales variaciones. Por lo tanto las banderas de calidad asignadas (Figura 30), para esta etapa se diferencian de la anterior y se definen como:

- Los datos ausentes identificados con -9999,0000 en los documentos primarios se clasifican inmediatamente con “9”.
- Valores que no pudieron ser evaluados por ausencia de información de referencia del Atlas son marcados con “0”.
- Los datos que cumplen con la prueba estadística aplicada se les asigna “2”.
- Los datos fuera del intervalo de confianza establecido se clasifican con “3”.

Flag Description	ODY	GTSP	ARGO	SEADATANET	ESEAS	WOD	WODSTATION	WOCEBOTTLE	WOCECTD	WOCE SAMPLE	QARTOD	BODC	PANGAEA	SMII	OceanSITES	LODE
no quality control (QC) was performed	1	0	0	0	0	0	0	2	2	2	0	Q	*	blank	0	2
QC was performed; good data	0	1	1	1	1	0	0	2	2	2	3	blank	blank	blank	1	1
QC was performed; probably good data	0	2	2	2	1	0	0	2	2	2	3	blank	blank	blank	2	1
QC was performed; probably bad data	4	3	3	3	3	4	3	3	3	7	2	K	?	?	3	3
QC was performed; bad data	8	4	4	4	4	4	3	4	4	7	1	K	/	B	4	4
the value was changed as a result of QC	1	5	5	5	2	0	0	2	2	2	3	R	*		5	2
the value is missing	1	9	9	9	9	0	0	5	5	5	9	N	*	B	9	9

Figura 29. Equivalencias de banderas de calidad GTSP para Etapa 3. Fuente: Schlitzer et al (2013)

El Atlas de los Datos Oceanográficos de Colombia 1922-2013 también proporciona información de tipo espacial sobre la distribución y el comportamiento de ciertas variables en diferentes periodos. Así fue posible extraer del atlas el mapa de isotermas para TSM para los diferentes

meses del año y contrastarlo con la posición de las estaciones de muestreo (color verde), tal como se muestra en las figuras de la 31 a la 42.

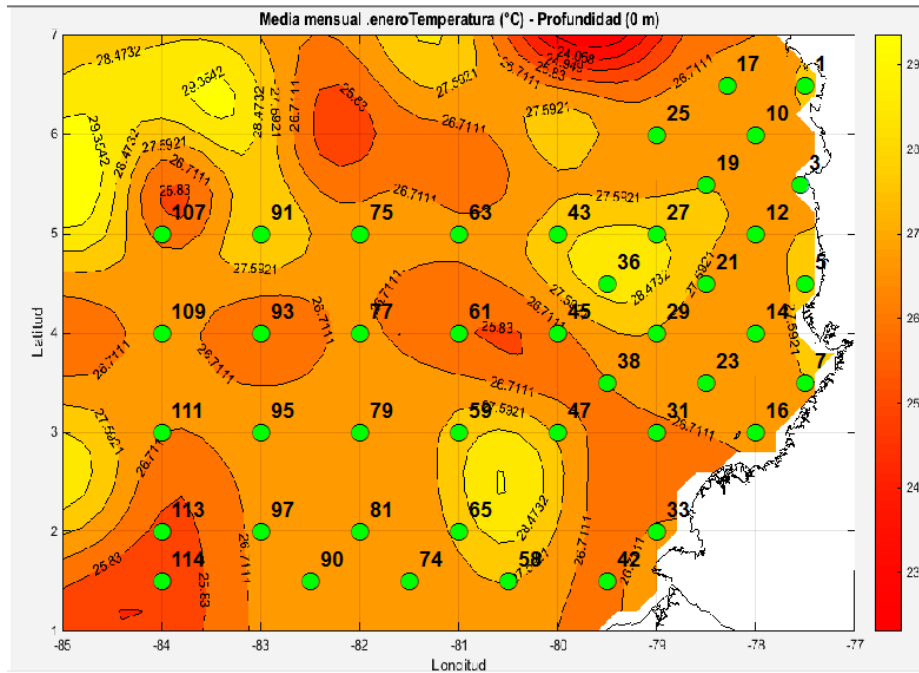


Figura 30. Distribución de las isotermas para media mensual de TSM en el mes de enero para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

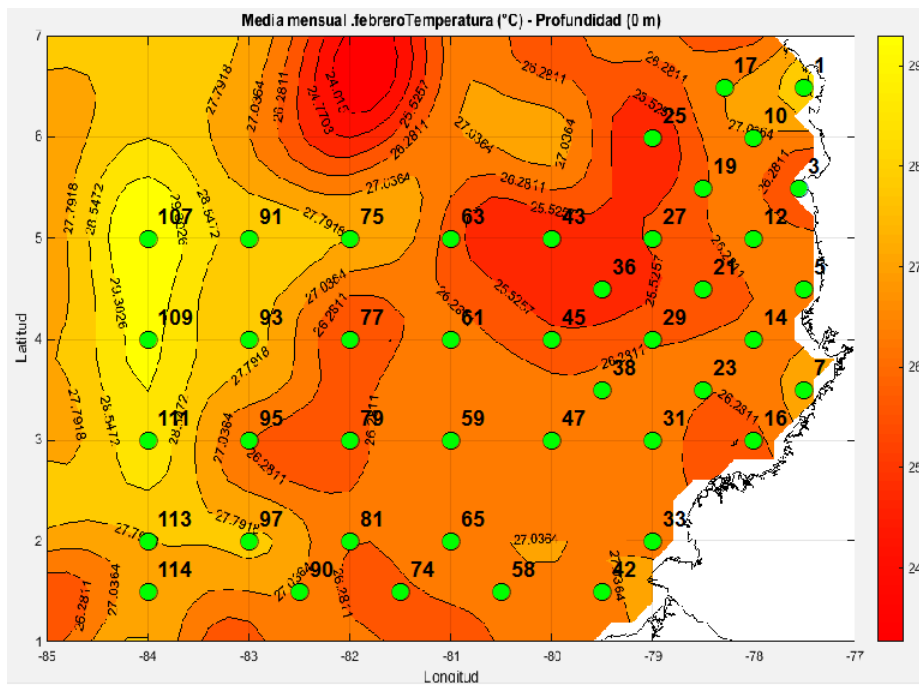


Figura 31. Distribución de las isotermas para media mensual de TSM en el mes de febrero para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

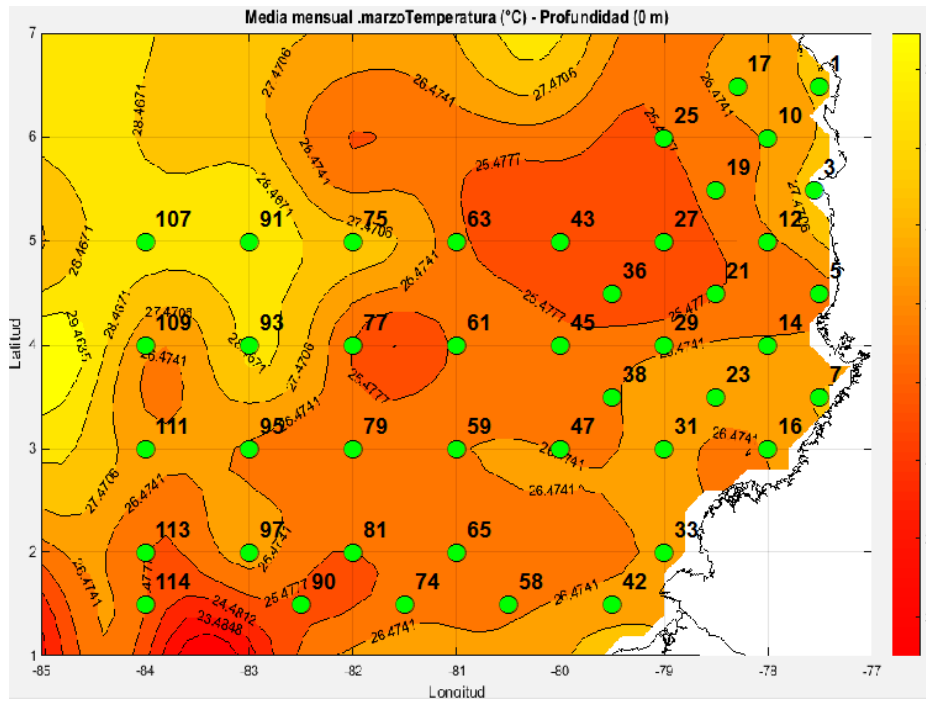


Figura 32. Distribución de las isotermas para media mensual de TSM en el mes de marzo para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

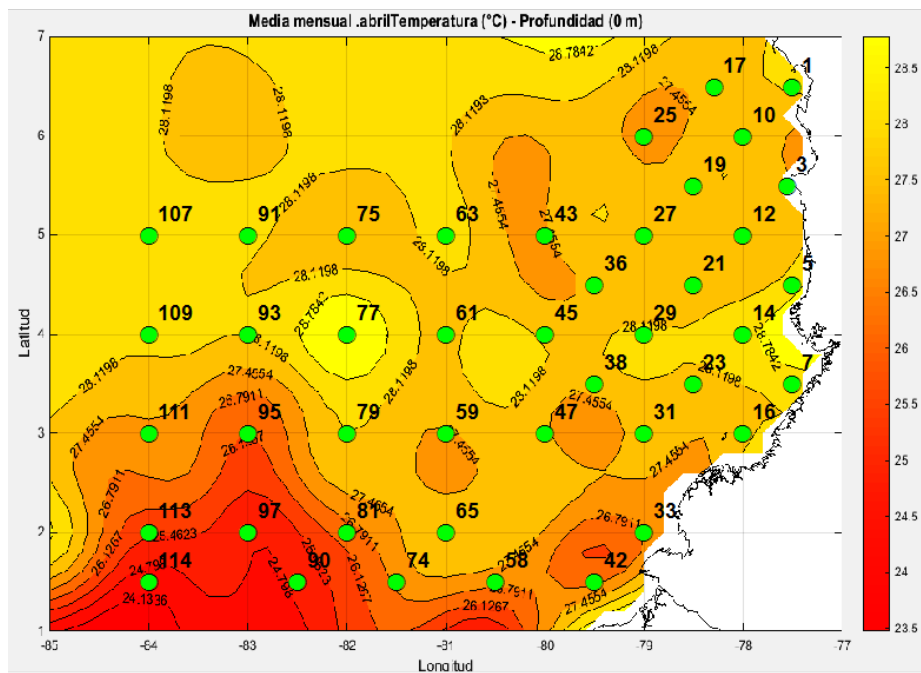


Figura 33. Distribución de las isotermas para media mensual de TSM en el mes de abril para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

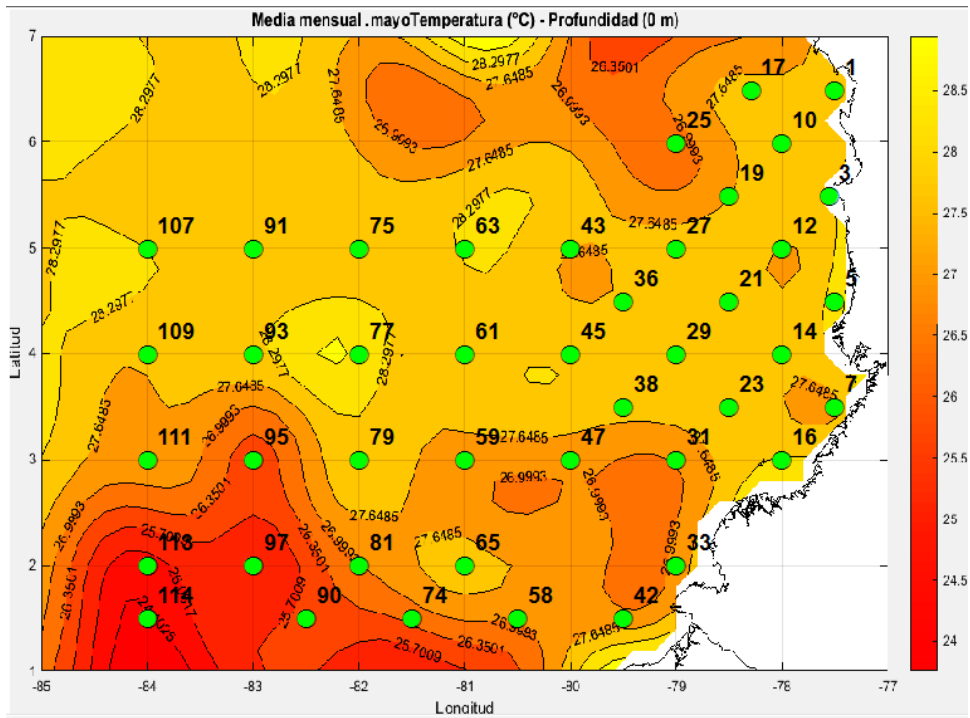


Figura 34. Distribución de las isotermas para media mensual de TSM en el mes de mayo para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

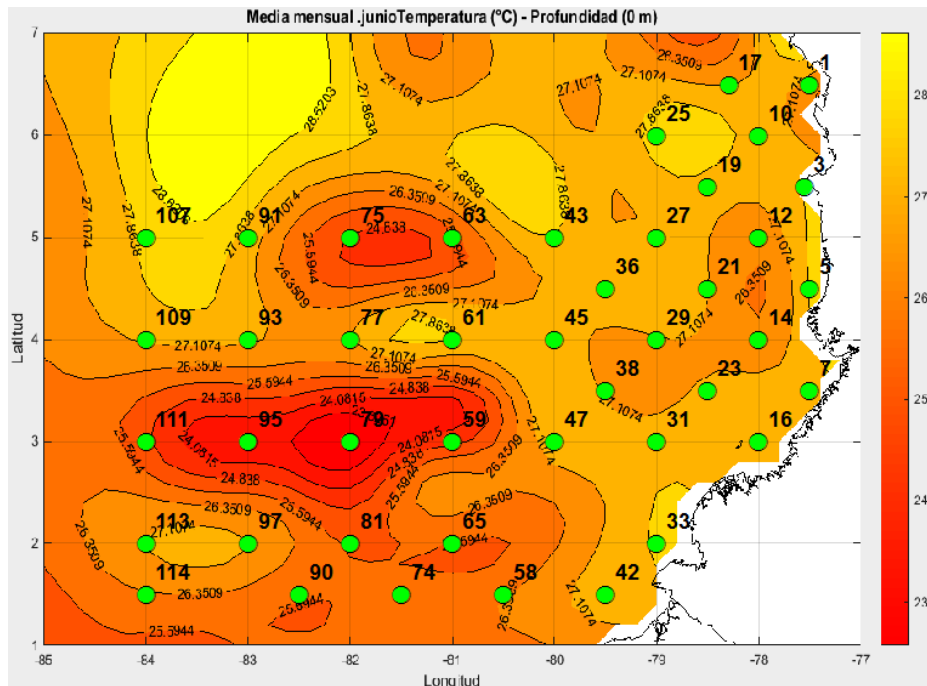


Figura 35. Distribución de las isotermas para media mensual de TSM en el mes de junio para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

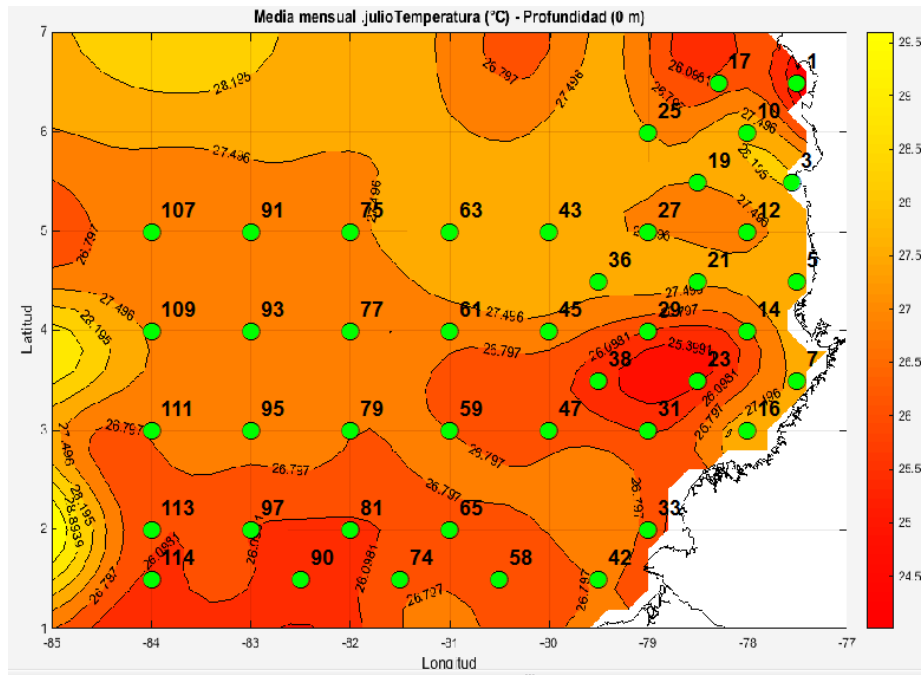


Figura 36. Distribución de las isotermas para media mensual de TSM en el mes de julio para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

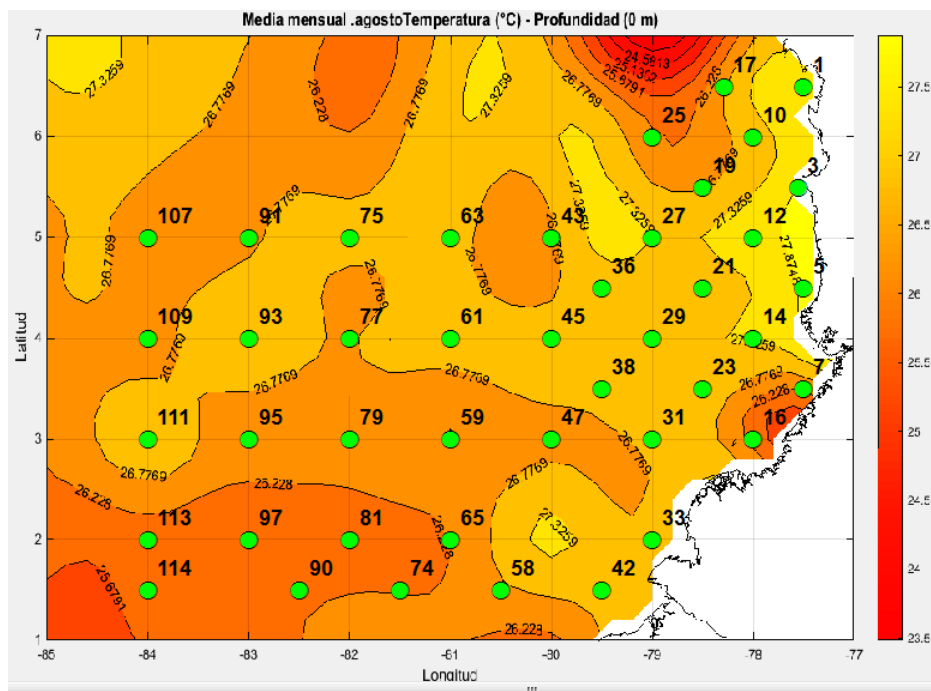


Figura 37. Distribución de las isotermas para media mensual de TSM en el mes de agosto para la CPC.
Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

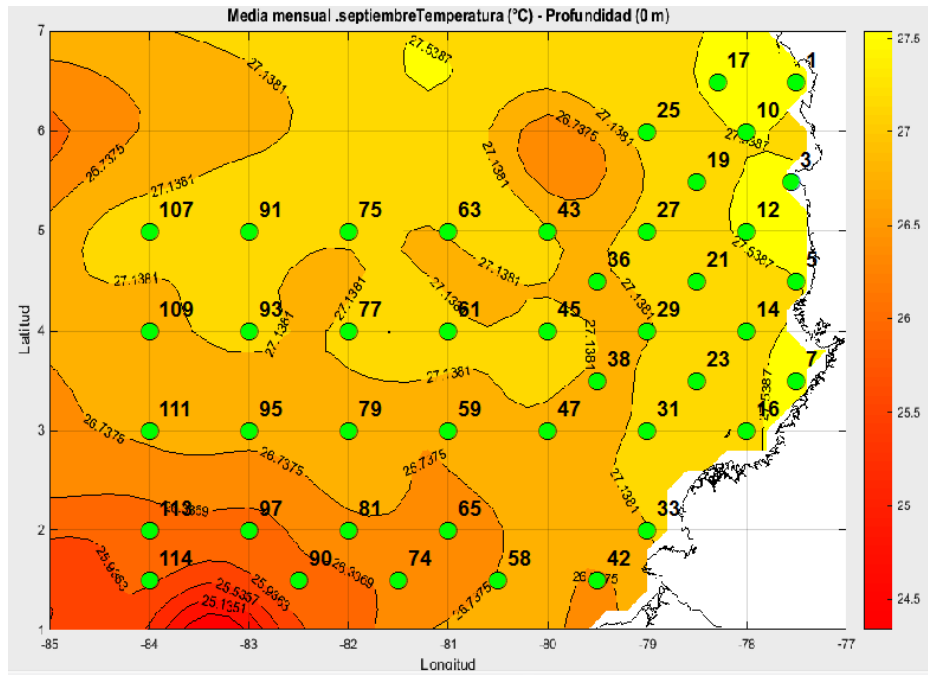


Figura 38. Distribución de las isotermas para media mensual de TSM en el mes de septiembre para la CPC. Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

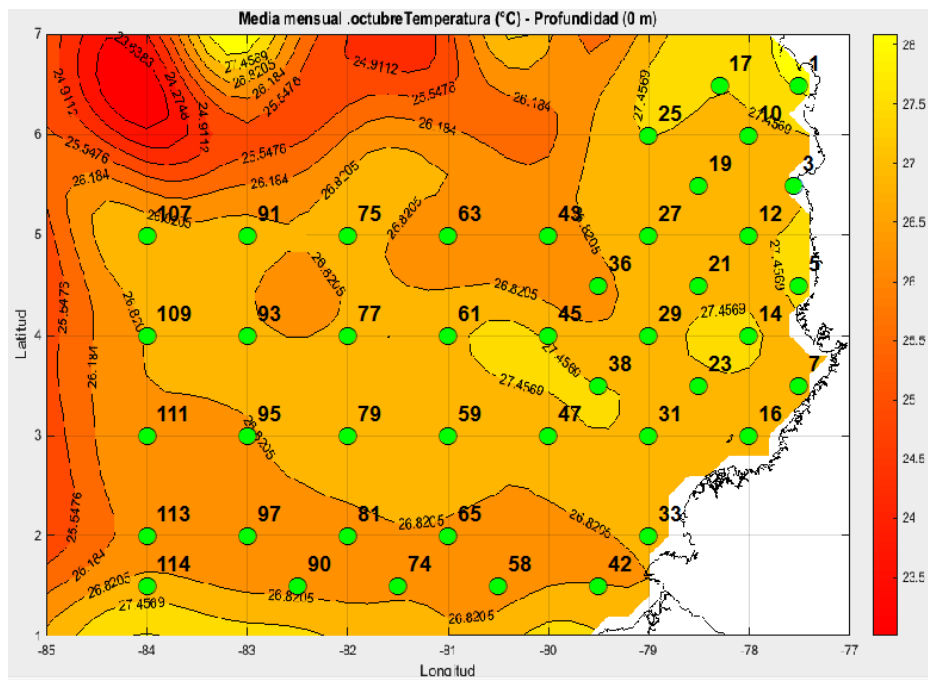


Figura 39. Distribución de las isotermas para media mensual de TSM en el mes de octubre para la CPC. Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

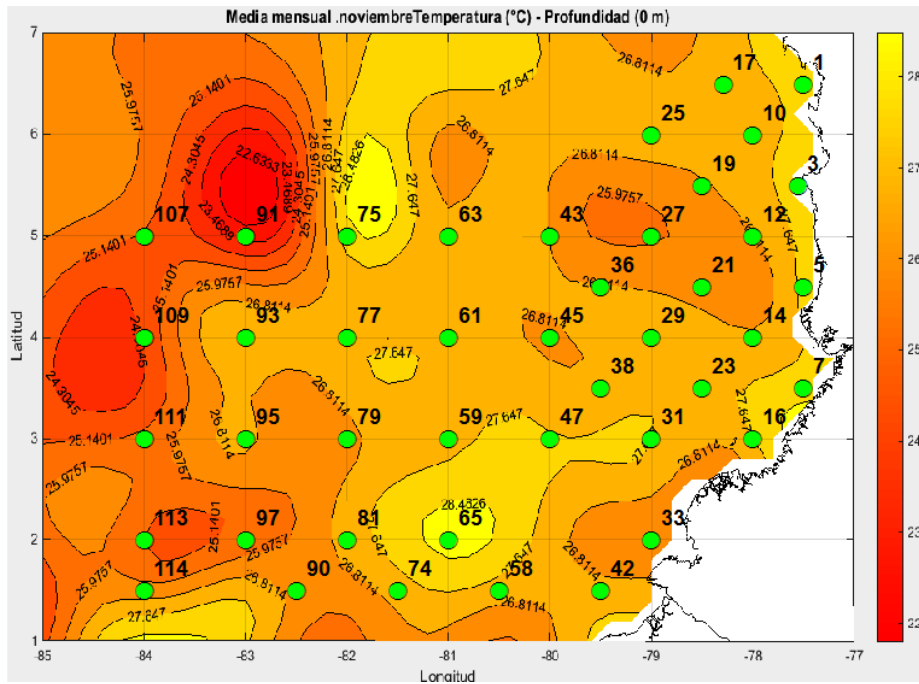


Figura 40. Distribución de las isotermas para media mensual de TSM en el mes de noviembre para la CPC. Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

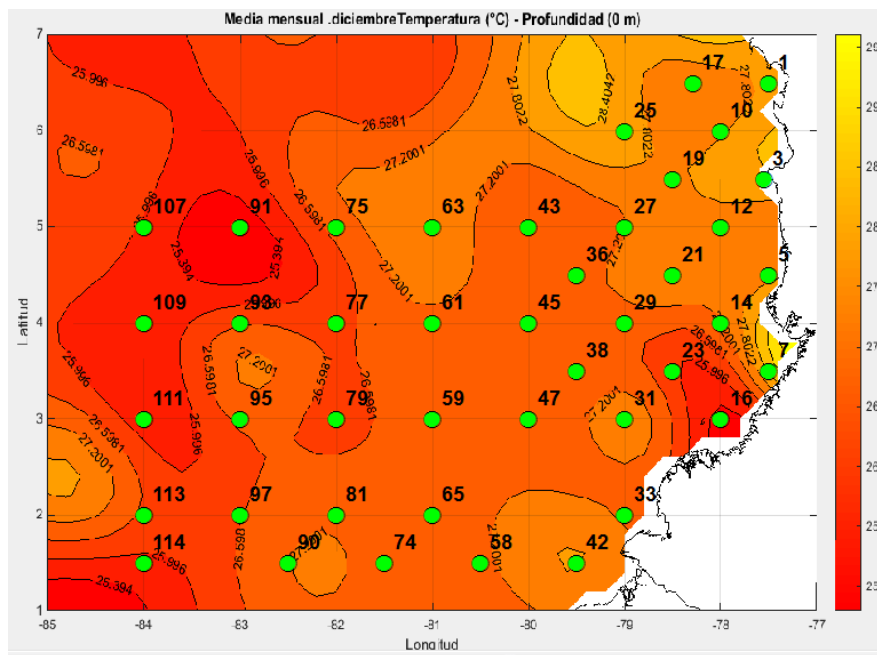


Figura 41. Distribución de las isotermas para media mensual de TSM en el mes de diciembre para la CPC. Fuente: Atlas de los Datos Oceanográficos de Colombia 1922-2013.

NOTA: Para esta etapa se escribió un SCRIPT en el software R que ejecuta las operaciones estadísticas que permiten hacer la clasificación de cada conjunto de datos con la segunda asignación de banderas de calidad. Al final de esta fase se obtiene un documento que contiene una columna denominada QF_ETAPA3 la cual clasifica cada dato con alguno de los códigos de banderas de calidad correspondientes a las reglas establecidas según las pruebas estadísticas anteriormente descritas.

Para ver el SCRIPT de ejecución y su explicación detallada ver ANEXO 2_SCRIPT_co_tesisdimarUD_R_temperatura_0m_etapa3

6.8. Etapa 4- Prueba de Coherencia Espacial

El objetivo de la Etapa 4 es aplicar una prueba de coherencia espacial haciendo una inter-comparación de los datos de cruceo con las mediciones hechas en estaciones vecinas; esto permite construir intervalos de confianza que fijan el promedio regional espacial para cada estación.

La base del intervalo de confianza se basa en la correlación espacial de los datos, expresado mediante la fórmula:

$$(\bar{x}-3.921 * ee[\bar{x}], \bar{x}+3.921 * ee[\bar{x}])$$

\bar{x} = Media del conjunto de datos

ee= error estándar

n= tamaño de muestra

3.921 es el valor del cuantil de una normal estándar asociado a un nivel de confianza del 99.9%

Aquí el error estándar se expresa en función de la varianza (muestral) que se ve modificada para introducir un factor asociado a la correlación espacial existente entre los datos, representada por ρ , así:

$$var(\bar{x}) = \frac{\sigma^2}{n} \left[1 + 2 \left\{ \frac{\rho}{1-\rho} \right\} \left(1 - \frac{1}{n} \right) - 2 \left\{ \frac{\rho}{1-\rho} \right\}^2 \frac{(1-\rho^{n-1})}{n} \right]$$

$$ee(\bar{x}) = \sqrt{var(\bar{x})} = \frac{\sigma}{\sqrt{n}} \sqrt{\left[1 + 2 \left\{ \frac{\rho}{1-\rho} \right\} \left(1 - \frac{1}{n} \right) - 2 \left\{ \frac{\rho}{1-\rho} \right\}^2 \frac{(1-\rho^{n-1})}{n} \right]}$$

(Li, 2009)

Los datos utilizados para calcular la varianza corresponden a las mediciones de las estaciones vecinas seleccionadas; mientras que para obtener el valor de ρ , según los avances sobre correlación espacial de variables climatológicas efectuados por Bernal S et al., (2011) se recurre al Software *PASSAGE* (Rosenberg & Corey D. Anderson, 2011), el cual permite no solo identificar la correlación espacial de un conjunto de datos si no también establecer un rango de distancia de estaciones vecinas a través de un correlograma. Para ello se recurre al conjunto de datos filtrados producto de la Etapa 2-Parte 2, es importante que los datos no contengan valores atípicos o extremos que puedan alterar el valor de correlación entre puntos geográficos.

Para el caso de TSM *PASSAGE* arroja el correlograma observado en la Figura 43.

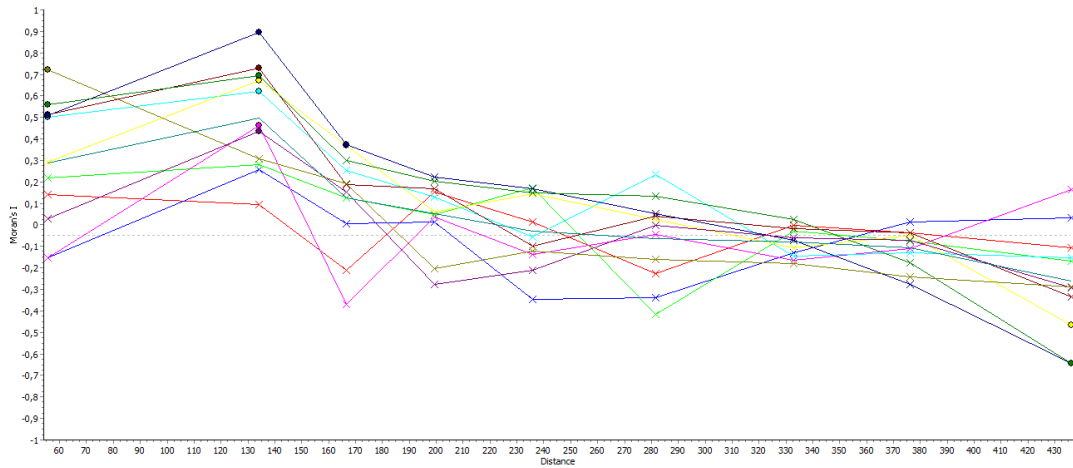


Figura 42. Correlograma de datos de mensuales TSM en software *PASSAGE*

De la forma como la información es introducida en *PASSAGE* el posible construir un correlograma casi para cada mes, representado por cada línea. En el eje x esta la distancia en Km y en el eje y el coeficiente de Moran que expresa la correlación de los datos de un mismo mes para el grupo de 21 estaciones. A medida que la distancia aumenta la correlación disminuye. Se escoge el punto de mayor correlación y la distancia que expresa antes que empiece a decaer. Así donde ocurre la mayor coincidencia de correlación y distancia es a 133 Km (Figura 44). Lo cual se fija como el rango de distancia a la cual se toma la información de estaciones vecinas para hacer la intercomparación.

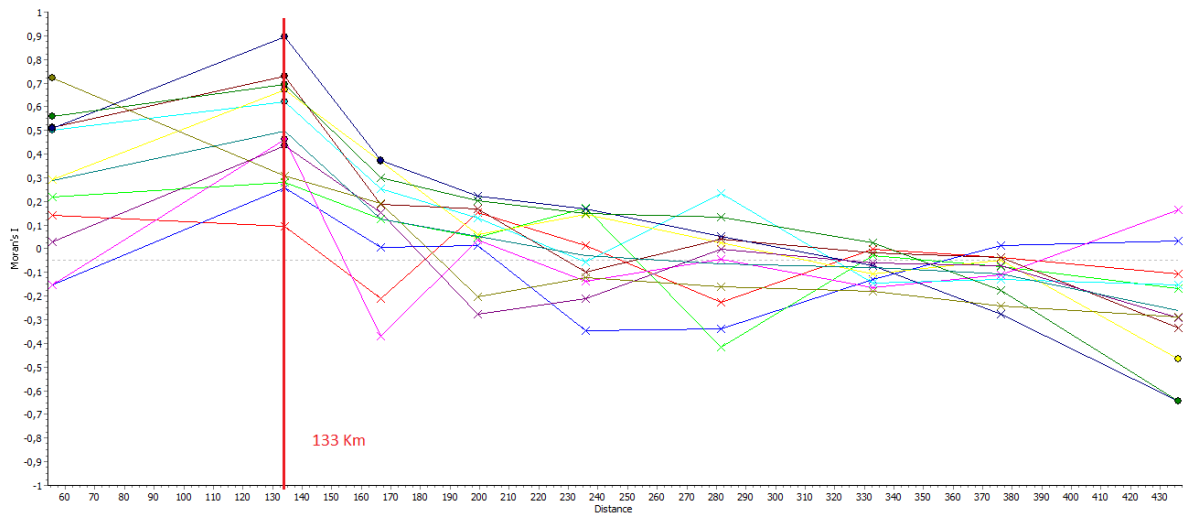


Figura 43. Estimación de rango de distancia de datos mensuales de TSM por Correlograma en software *PASSAGE*

La Tabla XIII se muestra el Índice de Moran para cada mes, para los meses de los que no se tenía suficiente información se utilizó el promedio (0.48) calculado por *PASSAGE*.

Tabla XIII. Índice de Moran mensual para datos de TSM.

Mes	Índice de Moran
Enero	0.48
Febrero	0.14
Marzo	0.48
Abril	0.26
Mayo	0.44

Junio	0.6
Julio	0.67
Agosto	0.72
Septiembre	0.9
Octubre	0.48
Noviembre	0.42
Diciembre	0.28

Así para la estación 63 con la ayuda del software ArcGis se identifican las estaciones vecinas como aquellas dentro de un rango de 133 Km, es decir, las estaciones 47, 61, 65 y 79 como lo muestra la Figura 45.

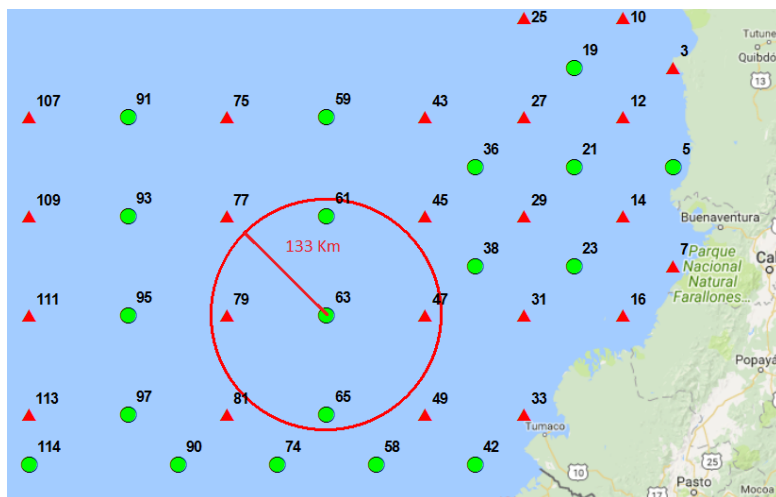


Figura 44. Rango de distancia de estaciones vecinas para TSM mensual de una estación, mediante el software ArcGis.

Para otros parámetros oceanográficos los resultados de correlograma y los rangos de distancia de estaciones vecinas pueden variar según las características de los datos a disposición. Así el mismo proceso realizado en *PASSAGE* para TSM se aplicó a oxígeno disuelto y nitratos, arrojando los resultados observados en las figuras de la 46 a la 50.



Figura 45. Estimación de rango de distancia de datos mensuales de oxígeno disuelto por Correlograma en software *PASSAGE*

Debido a la disposición de los datos de **oxígeno disuelto**. A diferencia que con TSM no fue posible establece un coeficiente de correlación para cada mes, en su lugar se utilizó el promedio mensual de todas las estaciones, paralelamente el rango de distancia es significativamente más amplio que con TSM, llegando a 280 Km (Figura 47), para tomar el índice de Moran más alto (0.4) para evitar el cálculo de intervalos demasiado cerrados. En ArcGis se puede apreciar el rango de distancia de vecindad para este parámetro, donde debido a una mayor distancia de vecindad también aumenta el número de puntos de muestreo considerados vecinos.

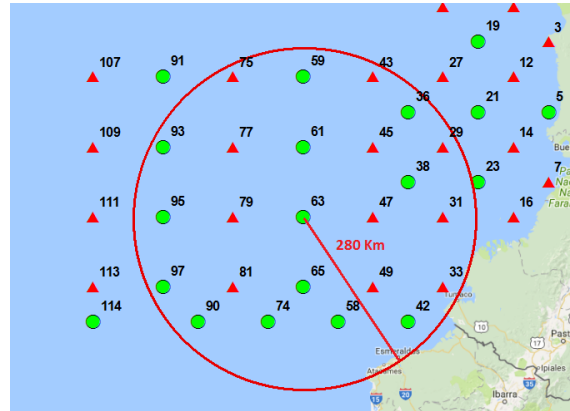


Figura 46. Rango de distancia de estaciones vecinas para oxígeno disuelto mensual de una estación, mediante el software ArcGis.

En el caso de **nitratos** la información disponible solo estaba completa para los meses de septiembre y octubre. Debido a que *PASSAGE* opera dentro del área designada por las coordenadas de cada punto de medición y a la escasez de datos para hacer comparaciones, se decidió designar un área más homogénea para que el software designara las correlaciones, así de las 16 estaciones seleccionadas para nitratos se introdujo solo la información de 11, cuyas ubicaciones se muestran en la Figura 48.

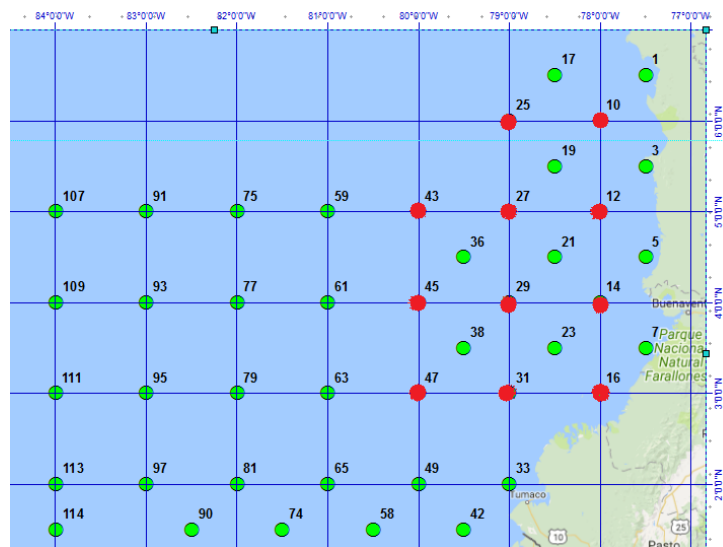


Figura 47. Estaciones seleccionadas para distribución homogénea de datos de nitratos para ensayo en software *PASSAGE*.

Con esta nueva configuración el correlograma construido por *PASSAGE* es:

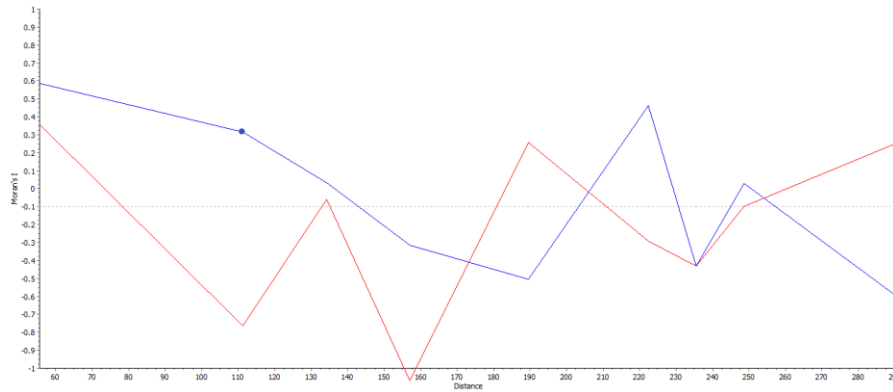


Figura 48. Estimación de rango de distancia de datos mensuales de nitratos con coordenadas homogéneas por Correlograma en software *PASSAGE*.

Estimación de rango de distancia observado en la Figura 49, de datos para dos meses: septiembre (rojo) y octubre (azul) de nitratos por Correlograma en software *PASSAGE*.

Debido al comportamiento irregular de la correlación arrojada por *PASSAGE* para el mes de septiembre (rojo) se decide tomar los valores de correlación espacial que presentan los datos del mes de octubre (azul). Así se designa un rango de correlación de estaciones vecinas de 112 Km (Figura 50) para nitratos con un índice de Moran de 0.32. De tal forma en ArcGis se puede apreciar el rango de distancia de vecindad para este parámetro.

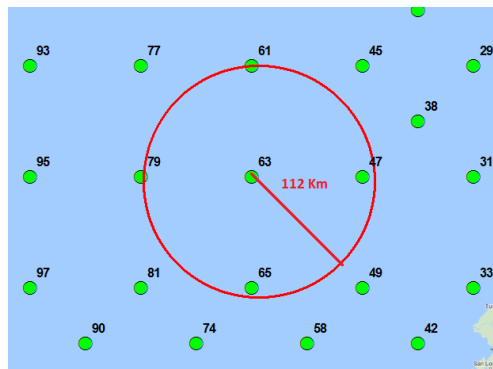


Figura 49. Rango de distancia de estaciones vecinas para nitratos de una estación, mediante el software ArcGis.

Además del proceso de determinación de los rangos de vecindad para cada parámetro oceanográfico y los índices de correlación (índice de Moran) calculados para cada caso, la metodología se aplica de la misma manera para todas las variables que se analizaron: TSM, OD y Nutrientes. Una vez identificadas las estaciones vecinas la evaluación consta de los siguientes pasos:

1. Se toma cada conjunto de datos de las estaciones vecinas y se combinan en uno solo, este nuevo conjunto de datos es el referente local espacial para la estación a evaluar.
2. La cantidad de datos disponibles de cada parámetro oceanográfico permite asignar un valor de correlación espacial (ρ), ya sea mensual o un promedio calculado, es necesario dividir el conjunto de datos en 12 grupos correspondientes a cada mes. Es posible que en ocasiones no haya información disponible de todos los meses, al igual que en el caso de la Etapa 3 los datos que no poseen un intervalo para ser comparados quedan a disposición de las pruebas aplicadas en las demás etapas para asignarles una bandera de calidad concreta.

- Cada grupo de datos con su propio ρ y n permite calcular el intervalo mensual de cada mes para esa región del océano, reemplazando los valores correspondientes en la fórmula.

$$var(\bar{x}) = \frac{\sigma^2}{n} \left[1 + 2 \left\{ \frac{\rho}{1-\rho} \right\} \left(1 - \frac{1}{n} \right) - 2 \left\{ \frac{\rho}{1-\rho} \right\}^2 \frac{(1-\rho^{n-1})}{n} \right]$$

$$ee(\bar{x}) = \sqrt{var(\bar{x})} = \frac{\sigma}{\sqrt{n}} \sqrt{\left[1 + 2 \left\{ \frac{\rho}{1-\rho} \right\} \left(1 - \frac{1}{n} \right) - 2 \left\{ \frac{\rho}{1-\rho} \right\}^2 \frac{(1-\rho^{n-1})}{n} \right]}$$

(Cressie, 1993)(Li, 2009)

Con los ensayos pertinentes, se identifica que debido que al combinar conjunto de datos de varias estaciones se pueden obtener tamaños de muestra (n) grandes que al estar en el denominador generan intervalos de confianza muy cerrados, por lo que para aumentar el rango de los valores y así evitar excluir una gran cantidad de datos se decide aumentar el nivel de confianza a un 99,9%.

- Una vez comparado cada dato con el intervalo mensual regional al que corresponde, la clasificación de las banderas de calidad de acuerdo al parámetro que se desea clasificar.

Para TSM según la GTSP (Figura 51) queda:

- Los datos ausentes identificados con -9999,0000 en los documentos primarios se clasifican inmediatamente con “9”.
- Valores que no pudieron ser evaluados por ausencia de información de referencia del Atlas son marcados con “0”.
- Los datos que cumplen con la prueba estadística aplicada se les asigna “1”.
- Los datos fuera del intervalo de confianza establecido se clasifican con “4”.

Flag Description	ODY	GTSP	ARGO	SEADATANET	ESEAS	WOD	WODSTATION	WOCEBOTTLE	WOCECTD	WOCEAMPLE	QARTOD	BODC	PANGAEA	SMHI	OceanSITES	LODE
no quality control (QC) was performed	1	0	0	0	0	0	0	2	2	2	0	Q	*	blank	0	2
QC was performed: good data	0	1	1	1	1	0	0	2	2	2	3	blank	blank	blank	1	1
QC was performed: probably good data	0	2	2	2	1	0	0	2	2	2	3	blank	blank	blank	2	1
QC was performed: probably bad data	4	3	3	3	3	4	3	3	3	7	2	K	?	?	3	3
QC was performed: bad data	8	4	4	4	4	4	3	4	4	7	1	K	/	B	4	4
the value was changed as a result of QC	1	5	5	5	2	0	0	2	2	2	3	R	*		5	2
the value is missing	1	9	9	9	9	0	0	5	5	5	9	N	*	B	9	9

Figura 50. Equivalencias de banderas de calidad GTSP para Etapa 5. Fuente: Schlitzer et al (2013)

Para oxígeno disuelto según la OceanSites queda:

- Los datos ausentes identificados con -9999,0000 en los documentos primarios se clasifican inmediatamente con “9”.
- Valores que no pudieron ser evaluados por ausencia de información de referencia del Atlas son marcados con “0”.

- Los datos que cumplen con la prueba estadística aplicada se les asigna “1”.
- Los datos fuera del intervalo de confianza establecido se clasifican con “3”.

Para nitratos según la OceanSites (Figura 52) queda:

- Los datos ausentes identificados con -9999,0000 en los documentos primarios se clasifican inmediatamente con “9”.
- Valores que no pudieron ser evaluados por ausencia de información de referencia del Atlas son marcados con “0”.
- Los datos que cumplen con la prueba estadística aplicada se les asigna “1”.
- Los datos fuera del intervalo de confianza establecido se clasifican con “2”.

Flag Description	ODY	GTSPP	ARGO	SEADATANET	ESFAS	WOD	WODSTATION	WOCEROTILE	WOCESID	WOCESAMPLE	QARTOD	BODC	PANGAEA	SMHI	OceanSITES	IODE
no QC was performed	1	0	0	0	0	0	0	2	2	2	0	Q	*	blank	0	2
good data	0	1	1	1	1	0	0	2	2	2	3	blank	blank	blank	1	1
probably good data	1	2	2	2	1	0	0	2	2	2	3	blank	blank	blank	2	1
bad data that are potentially correctable	4	3	3	3	3	4	3	3	3	7	2	K	?	?	3	3
bad data	8	4	4	4	3	4	3	4	4	7	1	K	/	B	4	4
value changed	1	5	5	5	0	0	0	2	2	2	3	R	*	blank	5	2
nominal value	1	0	0	0	0	0	0	2	2	2	3	R	*	blank	7	2
interpolated value	1	0	0	8	2	0	0	2	2	2	3	R	*		8	2
missing value	1	9	9	9	9	0	0	5	5	5	9	N	*	B	9	9

Figura 51. Equivalencias de banderas de calidad OceanSites para Etapa 5. Fuente: Schlitzer et al (2013)

NOTA: Para esta etapa se escribió un SCRIPT en el software R que ejecuta las operaciones estadísticas que permiten hacer la clasificación de cada conjunto de datos con la tercera asignación de banderas de calidad. Al final de esta fase se obtiene un documento que contiene una columna denominada QF_ETAPA4 la cual clasifica cada dato con alguno de los códigos de banderas de calidad correspondientes a las reglas establecidas según las pruebas estadísticas anteriormente descritas.

Para ver el SCRIPT de ejecución y su explicación detallada ver

ANEXO 3_SCRIPT_co_tesisdimarUD_R_temperatura_0m_etapa4

ANEXO 6_SCRIPT_co_tesisdimarUD_R_OD_0m_etapa4

ANEXO 9_SCRIPT_co_tesisdimarUD_R_NO3_0m_etapa4

6.9. Etapa 5- Banderas de Calidad de Clasificación Final

Las tres etapas anteriores contribuyen a la labor de asignación de banderas de calidad. En este punto se asumen las clasificaciones anteriores como banderas de calidad parciales anteriormente aplicadas para otros proyectos como “Diseño e implementación de una herramienta computacional para el control de calidad y validación de datos oceanográficos” (Sánchez Escobar, 2006). El paso final consiste en evaluar el resultado de cada dato en las 3 etapas a las que fue sometido y asignar una bandera de calidad de clasificación, que permita pasar la equivalencia desde GTSPP a IODE para la presentación final de los archivos. Cabe resaltar que en este punto del proceso cada documento se organiza y nombra según los parámetros establecido por la Guía de Normalización.

Así las reglas para la clasificación final de los datos según los resultados conjuntos de las tres etapas, se rigen bajo las recomendaciones hechas por la GTSPP. Las banderas de calidad de la

GTSP consideran los códigos 1 (dato bueno) y 2 (dato probablemente bueno) como aceptables y debido a la rigurosidad de cada prueba estadística cualquiera de las dos clasificaciones cuenta como aprobada. De 2 en adelante se considera como que el dato no paso la prueba correspondiente. Así la clasificación final en orden de prioridad queda definida como:

- Si el dato pasa 3 pruebas se clasifica con “1”.
- Si el dato obtuvo “4” en la primera etapa se clasifica automáticamente con “4” sin considerar su resultado en la etapa 3 y 4.
- Si el dato pasa 2 pruebas se clasifica con “2”.
- Si el dato falla en 2 pruebas se califica con “3”.
- Si el dato falla en 3 pruebas se califica con “4”.
- Si hay alguna etapa queda calificada con “0” el dato se califica con el código de mayor valor.
- Los datos ausentes se marcan con “9”.

Tabla XIV. Ejemplo de clasificación de banderas de calidad para TSM y su equivalencia en banderas IODE. Adaptado de Centro Colombiano de Datos Oceanográficos (2016)

Banderas de calidad asignadas según GSTPP				Equivalencia en banderas IODE
QF_ETAPA2	QF_ETAPA3	QF_ETAPA4	QF_CLASIFICACIÓN	
1	2	1	1	1
4	3	1	4	4
1	2	3	2	1
1	3	3	3	3
1	0	3	3	3
1	2	0	2	1
9	9	9	9	9

Para los parámetros de **oxígeno disuelto y nitratos** la asignación de banderas de calidad varia debido al comportamiento de cada variable y a que se utiliza un sistema de nomenclatura distinto, igualmente se tiene en cuenta que a estas dos variables solo es posible aplicar dos pruebas, la etapa 2 y 4.

Así las reglas para la clasificación final de los datos según los resultados conjuntos de las dos etapas, se rigen bajo las recomendaciones hechas por la OceanSites. Las banderas de calidad de la OceanSites consideran los códigos 1 (dato bueno) y 2 (dato probablemente bueno) como aceptables y debido a la rigurosidad de cada prueba estadística cualquiera de las dos clasificaciones cuenta como prueba aprobada. De 2 en adelante se considera como que el dato no paso la prueba correspondiente. Así la clasificación final en orden de prioridad queda definida como:

- Si el dato pasa 2 pruebas se clasifica con “1”.
- Si el dato pasa 1 prueba se clasifica con “2”.
- Si el dato obtuvo “4” o “3” en la primera etapa se clasifica automáticamente con “4” o “3” según sea el caso sin considerar los resultados de las etapas posteriores.
- Si el dato falla en 2 pruebas se califica con “4”.
- Si hay alguna etapa queda calificada con “0” el dato se califica con el código de mayor valor.
- Los datos ausentes se marcan con “9”.

Un ejemplo se puede observar en el Tabla XV.

Tabla XV. Ejemplo de clasificación de banderas de calidad para oxígeno disuelto y nitratos con su equivalencia en banderas IODE. Adaptado de Centro Colombiano de Datos Oceanográficos (2016).

Banderas de calidad asignadas según OceanSites			Equivalencia en banderas IODE
QF_ETAPA2	QF_ETAPA4	QF_CLASIFICACIÓN	
1	1	1	1
1	3	2	1
4	1	4	4
4	3	4	4
1	2	1	1
2	3	2	1
2	0	2	1
1	0	1	1
9	9	9	9

6.10. Etapa 6- Prueba de Detección de Cambios Entre Períodos Normal Vs. El Niño Y Normal Vs. La Niña

La etapa 6 no designa banderas de calidad a los datos sino que sirve a propósitos de análisis. De cada punto de muestreo se pretende someter los datos que han aprobado exitosamente las pruebas estadísticas de las etapas anteriores.

La prueba de Mann-Whitney es imparcial y consiste en un test de hipótesis precedentes el cual propone $P(X > Y)$. Aunque no siempre ocurre que es cierto para la hipótesis involucrada, el test pretende hacer una verificación o rechazo de la hipótesis propuesta (Sidney & Castellan, 2009).

Para asegurar que la prueba permanezca consistente e imparcial para las hipótesis que implican es suficiente agregar otra asunción al modelo anterior, es decir una hipótesis alternativa. La prueba requiere de datos de dos muestras aleatorias. Así $x_1, x_2, x_3 \dots x_n$ denota la muestra aleatoria de tamaño n de la población 1 y $y_1, y_2, y_3 \dots y_m$ denota la muestra aleatoria de tamaño m de la población 2. Así se comparan las diferencias entre dos medianas, por lo que se basa en rangos en lugar de en los parámetros de la muestra (media, varianza). Se emplea cuando los datos no siguen la distribución normal, en lugar del test de la t de Student que es de carácter paramétrico (Sidney & Castellan, 2009)

Se debe contar con muestras independientes y al azar. La hipótesis formulada es la de dos colas, la cual establece que existen diferencias entre las medianas (M) de los dos grupos considerados, sin presuponer cuál de las dos medianas es mayor que la otra. La hipótesis nula establece que no existen diferencias entre dichas medianas, así:

$$H_0: M_1 = M_2$$

$$H_1: M_1 \neq M_2$$

En este caso M_1 y M_2 corresponden a los datos de una misma estación pero de periodos diferentes, época Normal, Niño y Niña; la naturaleza de la prueba hace que solo se puedan comparar parejas de grupos. Por lo tanto se ensaya la época Normal con periodo El Niño para un caso y para el otro periodo Normal con La Niña, es decir para cada estación se aplica dos veces la prueba y arroja dos resultados.

El cálculo de la prueba se hace asignando rangos a cada dato. Para ello se ordenan todos los datos (juntando los dos grupos) en orden creciente. El rango de cada dato será el número de orden que le corresponde a cada dato. Cuando se repita el mismo valor numérico, el rango que se asigna a esos datos es la media aritmética de los rangos que les corresponderían en función del número de orden que ocupan. Luego se suman los rangos de cada uno de los inventarios (grupos) y se calcula la suma de los rangos de los datos de cada uno de los grupos (R_1 y R_2) (Universidad de Alcalá, 2005).

Para el cálculo de los estadísticos que permiten confirmar o rechazar la hipótesis propuesta se usa fórmulas:

$$M_1 = n_1 * n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

$$M_2 = n_1 * n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

Siendo n el tamaño de muestra respectivo de cada grupo y R la suma de los rangos de los datos de cada uno de los grupos. Así, con el resultado M_{cal} escogiendo el valor más grande entre M_1 y M_2 . Se debe comprobar la significancia estadística de M_{cal} comparando este valor con el valor de un estadístico M_{crit} obtenido a partir de las tablas correspondientes.

Si $M_{cal} \geq M_{crit}$ ($\alpha = 0.05$ o inferior) \rightarrow se rechaza H_0 (las medianas son diferentes)

Si $M_{cal} < M_{crit}$ ($\alpha = 0.05$) \rightarrow No se rechaza H_0 (las medianas son iguales)

Así la aplicación de la Etapa 6 consta de los siguientes pasos:

1. Filtrar los datos y seleccionar solo aquellos marcados con la bandera de calidad “1” producto de la clasificación final en la Etapa 5.
2. Dividir el conjunto de datos en grupos según época Normal, Niño y Niña.
3. Aplicar la prueba de Mann-Whitney para los grupos Normal Vs. Niño
4. Aplicar la prueba de Mann-Whitney para los grupos Normal Vs. Niña

(Universidad de Alcalá, 2005)

NOTA: Para esta etapa se escribió un SCRIPT en el software R que ejecuta las operaciones estadísticas anteriormente descritas de manera automática para cada conjunto de datos. Al final de esta fase el programa arroja el resultado de la prueba estadística para cada estación y cada caso.

Para ver el SCRIPT de ejecución y su explicación detallada ver

ANEXO 4_SCRIPT_co_tesisdimarUD_R_temperatura_0m_etapa6

ANEXO 7_SCRIPT_co_tesisdimarUD_R_OD_0m_etapa6

ANEXO 10_SCRIPT_co_tesisdimarUD_R_NO3_0m_etapa6

Los resultados obtenidos en esta etapa revelan si existe una diferencia en las medianas comparando los grupos Normal Vs. Niño y Normal Vs. Niña de cada estación en cada parámetro oceanográfico, como lo muestra la Tabla XVI.

Tabla XVI. Resultados de la prueba de Mann-Whitney por estación para TSM, oxígeno disuelto y nitratos.

ESTACIÓN	Valor p TSM				Valor p Oxígeno Disuelto				Valor p Nitratos	
	Normal Vs Niño	Medianas diferentes	Normal Vs Niña	Medianas diferentes	Normal Vs Niño	Medianas diferentes	Normal Vs Niña	Medianas diferentes	Normal Vs Niña	Medianas diferentes
10	0.106	NO	0.925	NO	0.643	NO	0.143	NO	0.191	NO
12	0.046	SI	0.499	NO	0.100	NO	0.699	NO	0.800	NO
14	0.195	NO	0.272	NO	0.800	NO	0.147	NO	0.800	NO
16	0.055	NO	0.138	NO	0.250	NO	0.202	NO	0.400	NO
17	0.164	NO	0.775	NO	0.764	NO	1.000	NO	--	--
19	0.244	NO	0.869	NO	0.250	NO	1.000	NO	--	--
21	0.842	NO	0.945	NO	0.730	NO	0.250	NO	--	--
25	0.015	SI	0.883	NO	0.024	SI	0.610	NO	0.800	NO
27	0.294	NO	0.68	NO	1.000	NO	1.000	NO	0.400	NO
29	0.037	SI	0.887	NO	0.222	NO	0.222	NO	0.400	NO
31	0.073	NO	0.012	SI	0.085	NO	0.833	NO	0.333	NO
33	0.198	NO	0.858	NO	0.500	NO	0.106	NO	0.071	NO
43	0.019	SI	0.555	NO	0.776	NO	0.533	NO	0.857	NO
45	0.144	NO	0.978	NO	0.333	NO	1.000	NO	0.533	NO
47	0.054	NO	0.366	NO	0.033	SI	0.927	NO	NA	NO
61	0.12	NO	0.222	NO	0.500	NO	0.700	NO	NA	NO
63	0.008	SI	1	NO	0.800	NO	0.730	NO	--	--
65	0.42	NO	0.5	NO	0.429	NO	0.610	NO	--	--
77	0.311	NO	1	NO	NA	NO	0.476	NO	0.571	NO
79	0.009	SI	0.267	NO	0.571	NO	0.841	NO	0.800	NO
81	0.031	SI	0.206	NO	0.333	NO	0.421	NO	0.267	NO

Para el caso de TSM de las 21 estaciones sólo 7 (33.3%) mostraron diferencias significativas de medianas entre los periodos Normal Vs. Niño y para los periodos Normal Vs. Niña solo una estación (4.76%) reveló diferencias significativas. Para oxígeno disuelto 2 estaciones (9.52%) mostraron cambios en periodos Normal VS Niño y ninguna para Normal VS Niña; finalmente la disponibilidad de datos para nitratos solo contaba con mediciones efectuadas durante época Normal y Niña, para esta comparación ninguna de las estaciones reveló diferencias entre las medianas según la prueba de Mann-Whitney.

Ya que mediante la prueba de Mann-Whitney no fue posible evidenciar la diferencia entre los valores de los datos según pedidos de medición; se decidió calcular las anomalías para cada estación y variable que permita observar como los eventos El Niño y La Niña influyen en el comportamiento de cada parámetro oceanográfico como se ve en la Tabla XVII y XVIII.

Tabla XVII. Anomalías por estación para TSM, oxígeno disuelto y nitratos para periodos El Niño y La Niña.

Estos resultados se encuentran en vista protegida
por petición del autor

Tabla XVIII. Resumen general de las anomalías para TSM, oxígeno disuelto y nitratos para periodos El Niño y La Niña.

Parámetro	TSM		OD		NITRATOS	
	Positiva	Negativa	Positiva	Negativa	Positiva	Negativa
El Niño	20 (95.23%)	1 (4.76%)	3 (14.28%)	17 (80.95%)	--	--
La Niña	12 (57.14%)	9 (42.85%)	5 (23.80%)	16 (76.19%)	3 (18.75%)	11 (68.75%)

La distribución espacial de las anomalías promedio para cada parámetro oceanográfico se observa en las de la figura 53 a la 58.



Figura 52. Distribución espacial de las anomalías promedio de TSM por estación para la CPC.
Anomalías El Niño (rojo) y La Niña (azul).



Figura 53. Valor de las anomalías promedio de TSM por estación para época El Niño.



Figura 54. Valor de las anomalías de TSM por estación para época La Niña.



Figura 55. Valor de las anomalías de oxígeno disuelto (0m) por estación para época El Niño.



Figura 56. Valor de las anomalías de oxígeno disuelto (0m) por estación para época La Niña



Figura 57. Valor de las anomalías de nitratos (0m) por estación para época La Niña

6.11. Resultados Generales

TSM

Para temperatura superficial del mar de las 21 estaciones seleccionadas se procesaron un total de 768 datos y cuyo porcentaje de clasificación por banderas de calidad para cada etapa aparece en la Tabla XIX.

Tabla XIX. Resultado de asignación de banderas de calidad por etapas para TSM.

Código de bandera de calidad	QF_GSTPP (%)				QF_IODE (%)
	QF_ETAPA2	QF_ETAPA3	QF_ETAPA4	QF_CALIFICACIÓN	QF_FINAL
1	92.45	0.00	61.98	46.74	74.09
2	0.00	56.90	0.00	27.21	0.00
3	0.00	36.85	36.59	18.88	18.75
4	7.42	0.00	0.00	7.03	7.03
9	0.13	0.00	0.00	0.13	0.13
0	0.00	6.25	1.43	0.00	0.00

Del total de datos procesados 392 (51,04%) correspondían a época Normal, 219 (28,51%) a evento El Niño y 157 (20,44%) a La Niña según el Índice Oceánico El Niño (ONI) de la NOAA. Para la determinación de los promedios se tomaron los datos que finalmente fueron calificados con la bandera 1 (dato bueno) según la clasificación IODE y se calculó para cada una de las épocas el promedio global de la zona de estudio (Tabla XX).

Tabla XX. Resultado de promedio de TSM por épocas Normal, El Niño y La Niña.

Estadísticos TSM	Normal	El Niño	La Niña
Media	26.849 °C	27.579 °C	26.958 °C
Varianza	0.604	0.873	0.625
Desviación estándar	0.778	0.937	0.794
Mediana	27.00	27.312	27.061
Coefficiente de variación	0.029	0.034	0.029

Antes del procesamiento de los datos el valor de TSM más bajo registrado fue de 0.016 °C en la estación 65 (2° N, -81°W) en septiembre de 2009; mientras que el valor más alto fue de 30.14 °C en la estación 77 (4°N, -82°W) en mayo de 1998; ambos valores corresponden a épocas de evento El Niño.

Tras el procesamiento de los datos, para el conjunto de mediciones clasificadas con 1 (bueno) en la calificación final el valor más bajo de TSM es de 23.83 °C en la estación 43 (5°N, -80°W) en marzo de 2002 para un período Normal y el valor más alto es de 29,75 °C para la estación 14 (4°N, -78°W) en mayo de 1998 durante la ocurrencia de una evento El Niño.

OXIGENO DISUELTO

Para la concentración de oxígeno disuelto en las 21 estaciones seleccionadas se procesaron un total de 352 datos y cuyo porcentaje de clasificación por banderas de calidad para cada etapa aparece en la Tabla XXI.

Tabla XXI. Resultado de asignación de banderas de calidad por etapas para Oxígeno Disuelto a 0m de profundidad.

Código de bandera de calidad	QF_OceanSites (%)			QF_IODE (%)
	QF_ETAPA2	QF_ETAPA4	QF_CALIFICACIÓN	QF_FINAL
1	78.13	33.52	31.82	78.13
2	0.00	0.00	46.31	0.00
3	0.00	66.19	0.00	0.00
4	19.89	0.00	19.89	19.89
9	1.99	0.00	1.99	1.99

Del total de datos procesados 176 (50,00%) correspondían a época Normal, 61 (17,33%) a evento El Niño y 114 (32,39%) a La Niña según el Índice Oceánico El Niño (ONI) de la NOAA. Para la determinación de los promedios se tomaron los datos que finalmente fueron calificados con la bandera 1 (dato bueno) según la clasificación IODE y se calculó para cada una de las épocas el promedio global de la zona de estudio (Tabla XXII)

Tabla XXII. Resultado de promedio de oxígeno disuelto a 0m por épocas Normal, El Niño y La Niña.

Estadísticos Oxígeno Disuelto a 0 m	Normal	El Niño	La Niña
Media	5.843 mg/l	5.128 mg/l	5.748 mg/l
Mediana	5.892	5.020	5.821
Desviación estándar	0.888	0.440	0.822
Varianza	0.784	0.190	0.667
Coficiente de variación	0.152	0.086	0.143

Antes del procesamiento de los datos el valor de oxígeno disuelto a 0 m más bajo registrado fue de -1.8456 mg/l en la estación 14 (4° N, -78°W) en octubre de 2011; mientras que el valor más alto fue de 19.7140 mg/l en la estación 10 (6°N, -78°W) en mayo de 1999; ambas mediciones efectuadas durante un evento La Niña.

Tras el procesamiento de los datos, para el conjunto de mediciones clasificadas con 1 (bueno) en la calificación final el valor más bajo de TSM es de 3.6606 mg/l en la estación 27 (5°N, -79°W) en marzo de 2009 para un período Normal y el valor más alto es de 7.7443 mg/l para la estación 16 (3°N, -78°W) en marzo de 2009 también durante una época Normal.

NUTRIENTES NITRATOS (NO3)

Para la concentración de nitratos a 0 m de profundidad en las 16 estaciones seleccionadas se procesaron un total de 100 datos y cuyo porcentaje de clasificación por banderas de calidad para cada etapa aparece en la Tabla XXIII.

Tabla XXIII. Resultado de asignación de banderas de calidad por etapas para Nitratos a 0m de profundidad.

Código de bandera de calidad	QF_OceanSites (%)			QF_IODE (%)
	QF_ETAPA2	QF_ETAPA4	QF_CALIFICACIÓN	QF_FINAL
1	77	27	35	77
2	0	47	43	0
3	11	0	10	10
9	12	12	12	12
0	0	14	0	0

Del total de datos procesados 72 (72 %) correspondían a época Normal y 28 (28 %) a época La Niña. No se encontraron registros durante la ocurrencia de EL Niño, según el Índice Oceánico El Niño (ONI) de la NOAA. Para la determinación de los promedios se tomaron los datos que finalmente fueron calificados con la bandera 1 (dato bueno) según la clasificación IODE y se calculó para cada una de las épocas el promedio global de la zona de estudio (Tabla XXIV).

Tabla XXIV. Resultado de promedio de oxígeno disuelto a 0m por épocas Normal y La Niña.

Estadísticos Nitratos a 0 m	Normal	La Niña
Media	0.866 $\mu\text{mol/l}$	0.523 $\mu\text{mol/l}$
Mediana	0.600	0.454
Desviación estándar	0.856	0.304
Varianza	0.719	0.089
Coeficiente de variación	1.426	0.670

Antes del procesamiento de los datos el valor de nitratos a 0 m más bajo registrado fue de 0.0563 $\mu\text{mol/l}$ en la estación 12 (5° N, -78°W) en septiembre de 2012; mientras que el valor más alto fue de 20.0905 $\mu\text{mol/l}$ en la estación 27 (5°N, -79°W) en marzo de 2009; ambas mediciones efectuadas durante una época Normal.

Tras el procesamiento de los datos, para el conjunto de mediciones clasificadas con 1 (bueno) en la calificación final el valor más bajo de TSM es el mismo de 0.0563 $\mu\text{mol/l}$ y el valor más alto es de 4.8450 $\mu\text{mol/l}$ para la estación 47 (3°N, -80°W) en marzo de 2009 durante una época Normal.

****Para ver en detalle los resultados de cada variable para cada estación ver el ANEXO 12**

7. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

La siguiente sección se centra en la discusión de los métodos estadísticos desarrollados en la metodología propuesta centrándose en las etapas que contribuyeron a la asignación de banderas de calidad para datos oceanográficos; y en el análisis de los resultados de acuerdo al comportamiento de las variables estudiadas (TSM, oxígeno disuelto, nitratos).

7.1. Etapa 2- Identificación de Datos Atípicos

La etapa para la identificación de datos atípicos propuesta para los datos oceanográficos de la CPC permite detectar datos cuyos valores están por fuera del rango determinado para el conjunto de datos para cada variable en cuestión (TSM, oxígeno disuelto, NO₃). Esta etapa recrea lo que otras metodologías de control de calidad de datos oceanográficos proponen para identificar los valores atípicos. En la GTSP (UNESCO, 2010), los valores atípicos se marcan como aquellos por fuera de rangos ya establecidos según estudios previos. Las pruebas de “Valores de parámetros globales imposibles” y “Valores de parámetros regionales imposibles” fijan rangos según datos de Atlas climatológicos y oceanográficos construidos a lo largo de los años producto de múltiples investigaciones. De forma similar Sánchez Escobar (2006) utiliza un algoritmo para asignar banderas de calidad, donde los rangos de valores son fijados según estándares internacionales. Sin embargo según el nivel de exigencia de los procedimientos estos rangos a veces pueden resultar muy amplios; como en el caso de los valores globales de temperatura del mar establecidos por la GTSP (UNESCO, 2010) que están entre -2° a 40°C.

Se logra observar que esta etapa desarrollada para la presente investigación construye los rangos extremos a partir del mismo conjunto de datos que se quiere evaluar; el investigador tiene un papel como revisor de que los límites calculados por el método estadístico sean coherentes con la variabilidad de la variable en cuestión, para ello se tiene en cuenta los valores para los diferentes parámetros oceanográficos que otras metodologías y estudios en la región de interés manejan. El método propuesto permite abarcar en un solo procedimiento la evaluación correspondiente a la detección de valores extremos tanto global como regionalmente, se obtienen rangos particulares para cada conjunto de datos y estos a su vez obedecen al comportamiento de la variable según la forma como se disponen los datos, en este caso la clasificación de las mediciones entre grupos Normal, Niño y Niña permitió clasificar los datos teniendo en cuenta la influencia que dichos fenómenos puede tener en los valores de las series.

También se debe observar que el construir los rangos a partir del mismo conjunto de datos que se quiere evaluar representa construir límites de valores extremos de diferentes variables incluso si no se tiene información de referencia que permita corroborar tales rangos, ya que puede ocurrir que para ciertas variables y secciones del océano no haya límites aún definidos en la literatura. También puede ocurrir que el método no presente los resultados esperados debido a que la presencia de datos atípicos en tamaños de muestra muy pequeños puede influir en los rangos y evitar que se detecten datos atípicos, como ocurrió en el caso de los datos de oxígeno disuelto, sin embargo se propuso una metodología alternativa para abordar este tipo de problema que puede repetirse en otros conjuntos de datos y otro tipo de variables; ante tales casos el investigador tiene a la mano una metodología que cumple con la misma función que el procedimiento inicial.

En esta etapa, mediante el procedimiento de percentiles la asignación de banderas de calidad con clasificación de dato bueno (QF 1) corresponde al 92.06% de los datos de TSM y 77% de los datos de NO₃; mientras que aplicando el método de la media recortada (metodología alternativa) 78.13 % de los datos de oxígeno disuelto paso la prueba. Pudiéndose identificar una

eficacia equivalente para ambos métodos, con resultados similares teniendo en cuenta la diferencia de parámetros oceanográficos evaluados y las características del conjunto de datos al que pertenece cada uno.

7.2. Etapa 3- Prueba de Coherencia Temporal

La etapa que aplica la prueba de coherencia temporal toma elementos aplicados por la Global Temperature Salinity Profile Programme (GTSP) (UNESCO, 2010), desde la prueba de climatología mensual de Levitus para la comparación de los datos a nivel mensual, paralelamente las metodologías de gestión de datos a nivel climatológico, trabaja las series de tiempo a nivel mensual y anual y por estaciones de monitoreo. Debido a la disponibilidad de datos se escogió una escala temporal mensual para evaluar los diferentes parámetros oceanográficos, aplicándola a una serie de tiempo interrumpida. (Las series de tiempo por estación se pueden observar en el ANEXO 13).

En lugar de utilizar la climatología de Levitus se identificó un referente climatológico particular suficientemente detallado para la región, con el cual hacer las comparaciones mensuales del área de estudio. El Atlas de los Datos Oceanográficos de Colombia (Andrade Amaya et al., 2015) permitió utilizar una referencia oceanográfica regional y aplicando el intervalo de confianza de Student (Ospina Botero, 2001) se obtuvo un 56.90 % de coincidencia entre los datos de TSM de crucero y la información de referencia, ello significa que la asignación de banderas de calidad en esta logró este valor porcentual. Este nivel de coincidencia puede atribuirse en parte debido a que del Atlas no se tuvo acceso a los datos primarios utilizados para su construcción, si no a estadísticos ya calculados; para ciertos puntos de muestreo no había información de todos los meses del año; y a que el proceso de análisis de los datos no se centraba en la posible influencia que eventos como El Niño o La Niña pudieran tener en los valores de las series. Así mismo las banderas de calidad aplicadas son más flexibles para intervalos de confianza que en ocasiones pudieron resultar muy estrechos para ciertas estaciones en ciertos meses del año.

Desde la climatología se entiende la importancia de la coherencia temporal para los datos, la decisión de manejar periodos mensuales se refuerza con los resultados arrojados por el Atlas, el mapa de isotermas para la CPC revela la fluctuación del campo de temperatura a lo largo del año (variabilidad intra-anual), y como de un mes a otro las condiciones pueden cambiar significativamente para un mismo punto de muestreo. A diferencia de la etapa anterior, la cantidad de datos disponibles por mes y por estación (tamaño de muestra), no resultaba suficiente para construir los intervalos a partir del mismo conjunto de datos con un nivel de confianza estadísticamente adecuado. Pero, el recurrir a fuentes externas de información como lo resultó el Atlas de los Datos Oceanográficos de Colombia resultó una herramienta útil que permite hacer comparaciones coherentes a la climatología de la zona de estudio. Es posible que en casos futuros el mismo procedimiento pueda ser aplicado recurriendo a otras fuentes de referencia como datos de satélite u otras bases de datos internacionales. Para lo cual se debe tener en cuenta tanto las características de la información de referencia como de los datos a evaluar, esto permite fijar condiciones a la asignación de banderas de calidad que no descarten un gran porcentaje de los datos sino que los clasifiquen en rangos de calidad que sean útiles para el investigador.

Para este caso el Atlas de los Datos Oceanográficos de Colombia sólo proporcionaba información de temperatura, salinidad y densidad a varias profundidades. En una aproximación más precisa el contar con datos de referencia de las demás variables analizadas en el presente trabajo (oxígeno disuelto y NO₃) habría permitido respaldar la metodología en parámetros oceanográficos diferentes de TSM con sus propias particularidades.

7.3. Etapa 4- Prueba de Coherencia Espacial

La aplicación de una prueba de coherencia espacial para el caso de datos oceanográficos utilizó elementos desde el análisis de variables climatológicas. El empleo de los datos a nivel estacional y la identificación de estaciones vecinas propuestas por (Bernal S et al., 2011) en la Homogenización de series de tiempo mensuales de precipitación, permitió a nivel oceanográfico construir rangos de vecindad definidos en Km para cada variable de acuerdo a las características de cada conjunto de datos. A su vez que la inter-comparación con estaciones vecinas permitió evaluar los valores de los datos a nivel espacial con las mismas mediciones de crucero.

En el ensayo y desarrollo de esta etapa, se identificó cómo la disponibilidad de datos y los tamaños de muestra, pueden influir en los valores significativos del Índice de Moran y los resultados que se pueden obtener con el correlograma. El etapa abordada revela que el contar con mediciones de todos los meses y todas las estaciones, además de coordenadas geográficas homogéneamente distribuidas, permite con el software *PASSAGE* construir resultados más coherentes. Por ello debido a la equidistante distribución de las estaciones oceanográficas fue posible definir un único rango de vecindad para todas las estaciones, según cada variable. De no ser así, habría sido necesario calcular un rango de vecindad para cada estación en particular, como ocurre para el caso de estaciones meteorológicas propuesto por Bernal S et al (2011), Lombana González et al(n.d.) y Barrios Moreno (2013).

En el caso de TSM los datos disponibles permitieron asignar un coeficiente de correlación (Índice de Moran) para casi cada mes del año, lo cual refuerza los resultados obtenidos en la prueba anterior y muestra cómo la correlación espacial entre las estaciones puede cambiar mes a mes debido a los movimientos del campo de temperatura que se evidencian en las gráficas de isotermas para la CPC obtenidas del Atlas de los Datos Oceanográficos de Colombia (Ver figura 31 a la 42); existen meses del año en que la correlación espacial es más fuerte que en otros según la estación. Los resultados obtenidos para TSM, exponen como la correlación espacial de ciertos parámetros oceanográficos no depende únicamente de la ubicación geográfica sino también del tiempo. Como se puede observar en detalle en el ANEXO 12, a medida que los puntos de muestreo se alejan de la costa el porcentaje de coincidencia espacial entre los datos de crucero y su rango regional aumenta, esto puede deberse a que cerca de la costa las variaciones de los parámetros oceanográficos son mayores debido a las interacciones con el continente (aportes fluviales, corrientes, mareas, entre otros), caso específico que se tuvo en consideración para la elaboración del Atlas de los Datos Oceanográficos de Colombia (Andrade Amaya et al., 2015) en donde se designaron valores de tolerancia mayores para las zonas costeras que para las oceánicas; en este aspecto sería posible perfeccionar la metodología propuesta para considerar las diferencias entre la correlación espacial a nivel costero y oceánico.

Incluir una función que considera la correlación espacial para construir un intervalo de confianza, como lo propone Cressie (Li, 2009) y a la vez tomar los datos de las estaciones vecinas dentro del rango establecido en el software *PASSAGE* demuestra una coincidencia espacial entre los datos de crucero de 61.98% para TSM, para oxígeno disuelto de 33.52% y para nitratos (NO₃) de 27%; es decir estos valores porcentuales corresponden a la asignación de bandera de calidad “1” (dato bueno) después de aplicar esta etapa de coherencia espacial.

En el caso de oxígeno disuelto, éste estableció el rango de correlación más grande (280 Km) lo cual supone una menor variación espacial en función de la distancia para este parámetro, una disponibilidad mayor de mediciones para todos los meses y las coordenadas de interés permitiría corroborar el resultado para esta variable. Mientras que para TSM y nitratos los rangos fueron similares pero por causas diferentes. En el caso de TSM debido a que se contaba con información completa de todas las estaciones; por el contrario para nitratos se contaba con menos estaciones, menos datos y menos meses, el modificar esta etapa designando un área de

cálculo más uniforme permitió al software *PASSAGE* arrojar un Índice de Moran significativo y un rango de vecindad con el que se pudieran evaluar los datos. El porcentaje de éxito en esta etapa para oxígeno disuelto fue el más reducido, posiblemente debido a las variaciones que presentan los valores de parámetros químicos y cuyo rango queda excluido según los intervalos de confianza construidos con el planteamiento de Cressie (1993), que si bien se identificaron como estrechos con tamaños de muestra (*n*) grandes. El comportamiento particular del oxígeno disuelto y los nitratos permite explorar el estudio de la correlación espacial con parámetros cuya variabilidad es amplia, y la posibilidad de estudiarlo a profundidad de forma tal que permita definirlo matemáticamente para esta y otras variables.

7.4. Etapa 6- Prueba de Detección de Cambios Entre Períodos Normal Vs. El Niño Y Normal Vs. La Niña

En los resultados producto de la etapa 6 para la detección de cambios entre períodos Normal Vs. El Niño y Normal Vs. La Niña se identificó para muy pocas estaciones la diferencia de medianas. Estadísticamente tales contrastes no fueron significativos a nivel de la prueba de Mann-Whitney, sin embargo al calcular las anomalías para cada estación en cada período, se observó una influencia en los valores de los datos por efecto de eventos como El Niño y La Niña.

Para la TSM en la mayoría de los casos la anomalía absoluta fue mayor para eventos El Niño que para la Niña; el comportamiento esperado se presentó en el caso de las anomalías para las épocas El Niño en donde 95.23% de los casos resultaron ser positivos revelando un significativo aumento de la temperatura superficial del mar comparado con las épocas normales según el Índice Oceánico de El Niño (ONI) para la zona de estudio; mientras que para el caso de La Niña las anomalías que se esperarían ser negativas por efecto de un descenso en la TSM presentaron el comportamiento contrario (positivo) en el 57.14% de los casos.

En primer lugar se debe tener en cuenta que la magnitud de las anomalías de temperatura superficial del mar para el fenómeno de La Niña, no es proporcionalmente inversa a la que se observa durante El Niño, puesto que no hay simetría en las anomalías ni en la ocurrencia de estos fenómenos en relación con los eventos cálidos (Pabón Caicedo & Montealegre Bocanegra, 2017). Los resultados de la Tabla XVII y la Figura 53, sobre las anomalías de El Niño se presentan en mayor magnitud que las de La Niña, ya sean positivas o negativas, asunto que según lo explicado por Pabón Caicedo & Montealegre Bocanegra (2017), la TSM del Pacífico centro oriental que es donde se efectúan las mediciones para el cálculo del ONI, las anomalías positivas pueden llegar a 5° C en tanto que en los episodios fríos las anomalías negativas no pasan de 2,5°C. Tratándose de una sección diferente del océano (Pacífico oriental) puede que las anomalías no alcanzaran tales niveles, sin embargo la equivalencia en sus diferencias concuerda con la literatura.

Las anomalías revelaron un Niño significativamente cálido para el caso de TSM, sin embargo los datos correspondientes a períodos La Niña muestran un comportamiento no esperado; solo el 42.85% de las anomalías resultó negativa en comparación con los períodos normales. La Niña, se comportó con una tendencia cálida en una magnitud mucho menor a la observada durante época EL Niño.

Para evaluar tal comportamiento se debe tener en cuenta que los datos suministrados por la Dimar comprenden un período de tiempo de 1970 a 2016; de los utilizados para TSM en las 21 estaciones seleccionadas van de 1982 a 2014. Un lapso de tiempo en el que se registraron nueve eventos La Niña, de los cuales tres se catalogaron como fuertes, 3 como moderados y 3 como débiles, considerando que dicha clasificación se hace según las mediciones hechas en la región Niño 3 efectuadas por la NOAA en el Pacífico oriental central, a aproximadamente 4440 Km de

distancia de la CPC la persistencia de los cambios en la temperatura de las masas de agua sería diferente para las aguas del Pacífico colombiano. La forma como están distribuidas espacialmente las anomalías positivas y negativas para la época La Niña podría deberse a que durante la fase de debilitamiento del Fenómeno de La Niña es posible encontrar fraccionamiento en la franja de máximas anomalías negativas, que favorece la redistribución del calor tras la situación fría anómala en la región (Pabón Caicedo & Montealegre Bocanegra, 2017) así, es posible encontrar remanentes de aguas frías que se desplazan por la costa suramericana hacia el norte y hacia el sur, y coincide con las anomalías negativas de mayor magnitud que se detectaron en las estaciones 61, 65, 63, 81 (ver Figura 56), siendo las más al sur de la cuenca como fracciones de aguas más frías que vienen desde las aguas de Ecuador. En otros casos podría tratarse de un enfriamiento confinado en el Pacífico central denominado La Niña modoki, un evento Niña caracterizado por enfriamiento de las aguas en el Pacífico central (donde se realizan las mediciones para el ONI) a la vez que se presenta un calentamiento o condiciones cercanas a lo normal al occidente y oriente del océano Pacífico (Pabón Caicedo & Montealegre Bocanegra, 2017), zona de donde provienen las mediciones utilizadas para la presente investigación.

Para el caso de oxígeno disuelto las anomalías muestran tanto para época El Niño como La Niña un comportamiento predominantemente negativo, es decir, detecta una disminución de las concentraciones de oxígeno disuelto en el agua a nivel superficial en comparación con los periodos en condiciones normales. El Niño presenta concentraciones más reducidas que La Niña debido a que a mayor temperatura la solubilidad del oxígeno disuelto en el agua disminuye y así mismo la concentración de éste. Consecuentemente los niveles de nitratos en aguas superficiales también revelaron anomalías negativas en el 68.75% de los casos por estaciones para el evento de La Niña lo cual puede ser en respuesta a la disminución en las concentraciones de oxígeno disuelto correspondiente a la misma época ya que el proceso de nitrificación que transforma el amonio en nitritos y luego en nitratos requiere en cuerpos de agua de condiciones con un alto contenido de oxígeno disuelto y un pH neutral (Fuentes & Massol-Deyá, 2002a), la reducción en los niveles de oxígeno disuelto durante los periodos Niña por efectos de un significativo calentamiento de las aguas se refleja a su vez en la disminución en la concentración de nitratos para las mismas condiciones, esto es observable en los resultados de los gráficos de caja producto de la etapa 2.

Mediciones de nitratos efectuadas durante eventos El Niño, permitirían corroborar si dicha relación entre estos tres parámetros oceanográficos (físicos y químicos) se mantiene para condiciones más cálidas.

8. CONCLUSIONES

- La composición y distribución de los datos oceanográficos analizados presenta retos a la hora de plantear metodologías de control de calidad basadas en técnicas estadísticas para la homogenización de los mismos. Debido a los requerimientos logísticos para la recolección de estos, se trabaja con series de tiempo interrumpidas, pese a contar con una cantidad de mediciones mucho menores de las que se trabaja en climatología, los elementos derivados de algunas metodologías utilizadas en este campo pudieron ser adaptadas a datos oceanográficos para la interpretación del comportamiento de los parámetros escogidos y la observación de sus variaciones en el tiempo y en el espacio.
- La metodología desarrollada permitió plantear un protocolo general para la evaluación de datos oceanográficos en las variables de temperatura, oxígeno disuelto y nitratos, y su posterior asignación de banderas de calidad. La disponibilidad de datos para cada variable presentó sus propias particularidades a la hora de aplicar cada etapa del proceso, siendo la temperatura superficial del mar el parámetro que permitió la aplicación de las etapas de forma más completa y arrojó los resultados base que luego fueron adaptados para cada variable. Así, la eficacia de la metodología propuesta se manifiesta en los promedios de cada parámetro oceanográfico calculados para la época Normal a partir de las mediciones aprobadas por el control de calidad. La TSM presentó un promedio de 26.82 °C para la época Normal, dentro de los rangos de temperatura esperados para la zona (25-27°C); para oxígeno disuelto a 0 m el promedio en época normal fue de 5.85 mg/L en concordancia con el comportamiento de este parámetro (5.8- 6.8 mg/L) y finalmente las concentraciones de nitratos a 0m en condiciones normales 0.88 μmol/L se mantienen dentro del rango de concentraciones para la zona cuenca pacífica colombiana (0.02- 3.2 μmol/L). Valores comparados con la Compilación Oceanográfica de la Cuenca Pacífica Colombiana (CCCP, 2002).
- Mediante los datos aprobados por los controles de calidad desarrollados con la asignación de banderas de calidad, es posible identificar la relación que existe entre la temperatura superficial del mar, el oxígeno disuelto (0m) y los nitratos (0m) debido a los cambios que experimentan las variables no solo en condiciones normales sino también por efecto de fenómenos océano- atmosféricos como lo son El Niño y La Niña; observándose una relación inversamente proporcional entre la temperatura y las concentraciones de oxígeno disuelto y nitratos en el agua.
- El análisis de la variabilidad de los parámetros oceanográficos por efecto de eventos El Niño y La Niña en aguas del Pacífico colombiano, revelaron que a nivel oceánico El Niño se comporta típicamente con un significativo calentamiento de la capa superficial de agua. Mientras que La Niña en lugar de presentar un enfriamiento de las aguas se manifestó como un ligero calentamiento en el promedio de la series comparado con las condiciones normales, especialmente en la zona centro y noroeste de la cuenca. Observándose que el comportamiento de las masas de agua para la Cuenca Pacífica Colombiana obedece a dinámicas propias a las que ya se han estudiado en la sección del Pacífico oriental.
- La metodología desarrollada en la presente investigación permite aplicar una evaluación estadística a datos oceanográficos para el aseguramiento de la calidad de los mismos. De forma tal que puedan ser utilizados como información base para el diagnóstico y

observación del estado del recurso marino a partir de las mediciones de sus diferentes parámetros oceanográficos a nivel temático, espacial y temporal. De forma tal que atribuye un mayor nivel de confiabilidad a los análisis y decisiones que tomen con base en la información procesada.

9. RECOMENDACIONES

- Es preciso continuar trabajando en las metodologías para el control de calidad de datos oceanográficos, no solo desde el perfeccionamiento de las técnicas estadísticas aplicadas, sino también la profundización en el tema oceanográfico y climatológico que permitan el extender los procesos de control a diferentes parámetros oceanográficos de tipo físico, químico y biológico; que permita contar con una metodología estándar que integre los aportes de diferentes campos de investigación en los que son requeridos datos de calidad a nivel oceanográfico.
- Es importante seguir trabajando en formas y mecanismos que permitan la automatización de los procesos de control de calidad para datos oceanográficos. Con la ayuda de programación en software es posible extender el alcance para procesar una gran cantidad de información de forma práctica y eficiente, que permita a las diferentes entidades y actores involucrados en la gestión de datos adoptar las metodologías de forma más eficaz, a la vez que se estandarizan los procesos de control de calidad.
- Para poder obtener la mayor utilidad de la propuesta metodológica es esencial integrar este tipo de iniciativas enfocadas a garantizar la calidad de datos oceanográficos en el contexto de los diferentes programas y redes de monitoreo que existen para el recurso marino, de tal forma que centros de investigación y entidades a cargo de estos cuenten con un procedimiento estándar para evaluar y clasificar su propia producción de información según parámetros nacionales e internacionales.

10. BIBLIOGRAFÍA

- A.Colubi, A.Lubiano, & P.Terán. (n.d.). Estadística Administrativa I (pp. 1–7). Universidad de Oviedo. Retrieved from <http://ocw.uniovi.es/course/view.php?id=167§ion=4>
- Andrade Amaya, C. A., Rangel Parra, O. E., & Herrera Vásquez, É. (2015). *Atlas de los Datos Oceanográficos de Colombia 1922-2013. Temperatura, Salinidad, Densidad, Velocidad Geostrofica*. (DIMAR- Dirección General Marítima & E. S.A., Eds.). Bogota, D. C.
- Artiola, J., Pepper, I. L., & Brusseau, M. L. (2004). *Environmental Monitoring and Characterization*. Elsevier Science & Technology Books.
- Barrios Moreno, J. S. (2013). *Aporte al Proceso de Toma de Decisiones en la Planificación Ambiental de la Región Climática del Bajo Magdalena con Base en Análisis Estadístico Empleando una Metodología para la Homogenización de Series mensuales de Precipitación*. Universidad Distrital Francisco Jose de Caldas.
- Bernal, N. R., Barrios, J. S., Ramos, M. A., Velásquez, C., & Ibarra, Y. C. (2012). Propuesta Metodológica para la Homogenización de Series de Tiempo de Precipitación Mensual y su utilidad en procesos de toma de decisiones , estudio de caso Región Climatológica del Bajo Magdalena.
- Bernal S, N. R., Correa, R. L., & Rangel M, E. (2011). Homogenización de series de tiempo mensuales de precipitación. *IX Congreso de Meteorología*, 26.
- Cassanova Rosero, R. F., Zambrano Ortiz, M. M., Latandret Solana, S. A., Suárez Vargas, N. P., Albán Illera, C. B., & Centro de Investigaciones Oceanográficas e Hidrológicas Pacífico CCCP. (2012). Variabilidad de parámetros fisicoquímicos en una estación oceánica frente a la Bahía de Tumaco. *Bol. Cient. CIOH*, (February 2011).
- CCCP, C. C. C. del P., & DIMAR, D. G. M. (2002). *Compilación Oceanográfica de la Cuenca Pacífica Colombiana*.
- CCCP, C. de I. O. e H. P., & DIMAR, D. G. M. (2016). Programas de Investigación- Oceanografía Operacional- Antecedentes. Retrieved from <http://www.cccp.org.co/index.php/programas-de-investigacion/oceanografia-operacional/106-antecedentes>
- CCCP, C. de I. O. e H. P., & Dirección de Climatología. (2002). Equipos de investigación. Retrieved from <http://www.cccp.org.co/index.php/component/joomgallery/equipos?page=2&catpage=1#category>
- CCO, C. C. del O. (2015). Eventos Extremos. Retrieved from www.cco.gov.co/eventos-extremos.html
- CECOLDO, C. C. de D. O., & DIMAR, D. G. M. (2016). Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceanográficos. Bogotá D.C.: Comisión Colombiana del Océano Comité Técnico Nacional de Coordinación de Datos e Información Oceánica.
- CENICAÑA, C. de I. de la C. e A. de C. (2017). Publicaciones fenómeno El Niño-CENICAÑA. Retrieved from [http://www.cenicana.org/web/climatologia/itemlist/tag/fenomeno El Niño](http://www.cenicana.org/web/climatologia/itemlist/tag/fenomeno%20El%20Ni%C3%B1o)
- Centro Colombiano de Datos Oceanográficos. (2016). *Guía para la Normalización de Conjuntos de Datos Oceanográficos*. EDITORIAL DIMAR.

- Centro Control Contaminación del Pacífico. (2002). *Compilación Oceanográfica de la Cuenca Pacífica Colombiana*.
- Centro Internacional para la Investigación del Fenómeno de El Niño. (2016). CIIFEN - Ciclos Biogeoquímicos. Retrieved from http://www.ciifen.org/index.php?option=com_content&view=article&id=580%253Aciclos-biogequimicos&catid=98%253Acontenido-1&Itemid=131&lang=es
- Cifuentes Lemus, J. L., Torres García, M. del P., & Frías M., M. (1997a). *El océano y sus recursos. II. Las ciencias del mar: oceanografía geológica y oceanografía química*. Fondo de Cultura Económica.
- Cifuentes Lemus, J. L., Torres García, M. del P., & Frías M., M. (1997b). *El Océano Y Sus Recursos. II. Las Ciencias Del Mar: Oceanografía Geológica Y Oceanografía Química*. (Fondo de Cultura Económica, Ed.) (Segunda Ed).
- COI, I. O. C., & UNESCO, I. C. for S. (n.d.). The Global Ocean Observing System. Retrieved from http://www.goosocean.org/index.php?option=com_content&view=article&id=119&Itemid=120
- CPPS, C. P. del P. S. (2015a). Historia y Trabajo de la CPPS. Retrieved from <http://cpps-int.org/index.php/home/cpps-historia>
- CPPS, C. P. del P. S. (2015). *Informe de la XXV Reunión del Comité Científico Regional del Programa ERFEN*. Santiago de Chile.
- CPPS, C. P. del P. S. (2015b). *Protocolo de Uso de CTD Sea-Bird y Procesamiento de Datos*. Guayaquil, Ecuador.
- CPPS, C. P. del P. sur. (2015). Generalidades sobre el Protocolo ERFEN. Retrieved from <http://cpps-int.org/index.php/sobre-erfen>
- DIMAR, D. G. M. (2016). Sistema de Medición de Parámetros Oceanográficos y de Meteorología Marina - SMPOMM. Retrieved from <https://www.dimar.mil.co/content/sistema-de-medicion-de-parametros-oceanograficos-y-de-meteorologia-marina-smpomm>
- DIMAR, D. G. M., & CECOLDO, C. C. de D. O. (1973). *Informe Crucero Oceanográfico Pacífico III-1972*. Bogotá D.C.
- DIMAR, D. G. M., & CECOLDO, C. C. de D. O. (1992). *Informe Crucero Oceanográfico Pacífico IX-1986*. Bogotá D.C.
- DIMAR, D. G. M., CECOLDO, C. C. de D. O., & CIOH, C. de I. O. e H. (1993). *Informe Crucero Oceanográfico ERFEN XXI-1993*. Cartagena de Indias-Colombia.
- Fuentes, F., & Massol-Deyá, A. (2002a). Tercera Parte Nutriente y Gases: Oxígeno Disuelto. In *MANUAL DE LABORATORIOS ECOLOGÍA DE MICROORGANISMOS*. Universidad de Puerto Rico. Retrieved from <http://www.uprm.edu/biology/profs/massol/manual/p3-oxigeno.pdf>
- Fuentes, F., & Massol-Deyá, A. (2002b). Tercera Parte Nutrientes y Gases: Nitrógeno. In *MANUAL DE LABORATORIOS ECOLOGÍA DE MICROORGANISMOS*. Universidad de Puerto Rico.
- García Torres, E. (2009). *Desarrollo de un Sistema Electrónico para el almacenamiento, Procesamiento y Control de Calidad de Datos en Tiempo Real Proveniente de*

- Mareografos para El Control de Alerta de Tsunamis*. Universidad industrial de Santander.
- Giraldo Henao, R. (n.d.). *Introducción a la Geoestadística. Teoría y Aplicación*. Bogotá D.C.: Universidad Nacional de Colombia.
- Hurtado Moreno, G., González, O. C., & IDEAM, I. de H. M. y E. A. (2015). Evaluación e la Afectación Territorial de los Fenómenos El Niño/La Niña y Análisis de la Confiabilidad de la Predicción Climática Basada en la Presencia de un Evento. Retrieved from <http://www.ideam.gov.co/documents/21021/440517/Evaluación+de+la+Afectación+Territorial.pdf/6b3fe5b4-2f99-40ab-ba47-ebacaec5ed91>
- IDEAM, I. de H. M. y E. A. (2014a). ASUNTOS METEREOLÓGICOS - IDEAM. Retrieved from <http://www.ideam.gov.co/web/atencion-y-participacion-ciudadana/asuntos-metereologicos>
- IDEAM, I. de H. M. y E. A. (2014b). PREDICCIÓN CLIMÁTICA - IDEAM. Retrieved from http://www.ideam.gov.co/web/tiempo-y-clima/prediccion-climatica/-/document_library_display/ljPLJWRaQzCm/view/413023?_110_INSTANCE_ljPLJWRaQzCm_redirect=http%3A%2F%2Fwww.ideam.gov.co%2Fweb%2Ftiempo-y-clima%2Fprediccion-climatica%3Fp_p_id%3D110_INSTANCE_ljPLJWRaQzCm%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-1%26p_p_col_count%3D1
- IDEAM, I. de H. M. y E. A., & Melo Franco, J. Y. (2005). Metodología para la Realización de Cruceros Oceanográficos en la Cuenca del Pacífico Suroriental (Pacífico-ERFEN). Bogotá D.C.
- Instituto de Tecnologías Educativas. (n.d.). Caracterización de las masas de agua. Retrieved from https://fjferreer.webs.ull.es/Apuntes3/Leccion04/1_caracterizacin_de_las_masas_de_agua.html
- INVEMAR, I. de I. M. y C. J. B. V. de A. (2003). *Manual de Técnicas Analíticas para la Determinación de Parámetros Fisicoquímicos y Contaminantes Marinos*. Santa Marta.
- INVEMAR, I. de I. M. y C. J. B. V. de A. (2017). REDCAM -INVEMAR Colombia 50% Mar. Retrieved from <http://www.invemar.org.co/redcam>
- INVEMAR, I. de investigaciones M. y C. J. B. V. de A. (2016). *Diagnóstico y Evaluación de la Calidad de las Aguas Marinas y Costeras del Caribe y Pacífico Colombianos*. Santa Marta, Colombia.
- Jaramillo Robledo, Á., Arcila Pulgarín, J., & CENICAFE, F. N. de C. (2009). Variabilidad Climática en la Zona Cafetera Colombiana Asociada al Evento de El Niño y su Efecto en la Caficultura.
- Kurtz, D. B., Perucca, A. R., Ybarra, D. D., Fernández López, C., Rey Montoya, S., & Barrios, R. (1998). El Fenómeno el Niño, Comparación de superficies afectadas por excesos hídricos en Corrientes. *Grupo Recursos Naturales EEA Corrientes- INTA*, 1–18.
- Li, Y. (2009). Chapter 4 Modeling and Analysis of Spatially Correlated Data.
- Lombana González, L., Bernal Suárez, N. R., & Barrios Moreno, J. S. (n.d.). *GUÍA PARA EL PROCESAMIENTO DE SERIES DE TIEMPO DE PRECIPITACIÓN Y TEMPERATURA : ESTIMACIÓN DE DATOS FALTANTES , DETECCIÓN DE CAMBIOS Y HOMOGENIZACIÓN*. Bogotá D.C.: Universidad Distrital Francisco José de Caldas.

- Lovett, G. M., Burns, D. A., Driscoll, C. T., Jenkins, J. C., Mitchell, M. J., Rustad, L., ... Haeuber, R. (2007). Who needs environmental monitoring ?
- Mendenhall, Beaver, & Beaver. (2010). *Introducción a la probabilidad y estadística* (Décima ter).
- Minster, J.-F. (2004). *Los Océanos*. Siglo XXI. Retrieved from https://books.google.es/books?hl=es&lr=&id=J99xqOSzkucC&oi=fnd&pg=PA7&dq=temperatura+del+ocean+en+los+ciclos+biogeoquimicos&ots=buUMX8kj9S&sig=3M5DJS_RxkUcM5cewjRIu1R1ZcY#v=onepage&q&f=true
- Montealegre Bocanegra, J. E., & Pabón Caicedo, J. D. (2000). La Variabilidad Climática Interanual Asociada al Ciclo El Niño-La Niña Oscilación del Sur y su efecto en el Patrón Pluviométrico de Colombia, 7–21.
- Montealegre Bocanegra, J. E., & Zea Mazo, J. A. (1994). *Estudio sobre el fenómeno “El Niño.”* (I. de H. M. y E. A. IDEAM, Ed.) (Segunda Ed). Bogotá D.C.: IDEAM, Instituto de Hidrología Meteorología y Estudios Ambientales.
- NOAA, N. O. and A. A. (2015a). Climate Prediction Center - Monitoring & Data_ ENSO Impacts on the U. Retrieved from http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml
- NOAA, N. O. and A. A. (2015b). GTSP (Global Temperature and Salinity Profile Programme). Retrieved from <https://www.nodc.noaa.gov/GTSP/overview/index.html>
- NOAA, N. O. and A. A., & Operations, O. of S. and P. (n.d.). Sea Surface Temperature (SST) Contour Charts - Office of Satellite and Product Operations. Retrieved from <http://www.ospo.noaa.gov/Products/ocean/sst/contour/>
- OceanSites. (2010). OceanSites User`s Manual.
- Organización Meteorológica Mundial OMM. (2011). *Guía de prácticas climatológicas*. Organización Meteorológica Mundial.
- Ospina Botero, D. (2001). *Introducción al Muestreo*. (Departamento de Estadística & Facultad de Ciencias, Eds.) (Primera). Bogotá: Universidad Nacional de Colombia.
- Pabón Caicedo, J. D., & Montealegre Bocanegra, J. E. (2017). *Los Fenómenos de El Niño y de La Niña su efecto climático e impactos socioeconómicos* (Colección). Bogotá D.C.: Academia Colombiana de Ciencias Exactas, Físicas y Naturales.
- Poveda, G., Jaramillo, A., Gil, M. M., Quiceno, N., & I.Mantilla, R. (2001). Seasonality in ENSO-related precipitation , river discharges , soil moisture , and vegetation index in Colombia. *Water Resources Research*, 37(8), 2169–2178.
- Rangel, G. J. L. (n.d.). Guías- Apuntes de Métodos Estadísticos. Retrieved from http://webdelprofesor.ula.ve/economia/gudberto/materias/metodos_estadisticos1/apuntes_me1.html
- Rivas Ruiz, R., Moreno Palacios, J., & O. Talavera, J. (2013). Investigación Clínica XVI Diferencias de Medianas con la U de Mann-Whitney, 51(4).
- Rodríguez, L. A., & Ortiz Martínez, R. V. (2006). *ESTUDIO PILOTO DE LOS PROCEDIMIENTOS DE CONTROL DE CALIDAD OCEANOGRÁFICOS Y METEOROLÓGICOS PRODUCIDOS POR EL CENTRO DE CONTROL CONTAMINACIÓN DEL PACÍFICO*. Centro Control Contaminación del Pacífico, Universidad del Magdalena.

- Rodríguez, N., Armenteras, D., Bernal Suárez, N. R., & Rincón, A. (2007). Fundamentos conceptuales de monitoreo y su aplicación en biodiversidad. In *Monitoreo de los ecosistemas andinos 1985-2005: Síntesis*. Bogotá D.C.: Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.
- Rosenberg, M. S., & Corey D. Anderson. (2011). PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. Methods in Ecology and Evolution. Retrieved from <http://www.passagesoftware.net/index.php>
- Rúa, D. M. R. (2010). *Implementación de Estándares Internacionales para el Control de Calidad de Datos Oceanográficos Provenientes de una Boya de Oleaje Direccional Triaxys*. Universidad de Antioquia.
- Sánchez Escobar, R. E. (2006). *DISEÑO E IMPLEMENTACIÓN DE UNA HERRAMIENTA COMPUTACIONAL PARA EL CONTROL DE LA CALIDAD Y VALIDACIÓN DE DATOS OCEANOGRÁFICOS*. Universidad del Cauca.
- Schlitzer, R., Wegener, A., & Instituto for Poland and Marine Research. (2013). Oceanographic quality flag schemes and mappings between them. Bremerhaven, Germany.
- SENAMHI, S. N. de M. e H. del P., Dirección General de Meteorología, & Dirección de Climatología. (2014). *El Fenómeno El Niño en el Perú*. Lima, Perú.
- Sidney, S., & Castellan, N. J. (2009). *Estadística no Paramétrica aplicada a las Ciencias de la Conducta* (4^o).
- UNESCO, I. O. C. of. (2010). GTSP Real-Time Quality Control Manual. Paris: UNESCO.
- UNESCO, I. O. C. of. (2013a). IODE Quality Management Framework for National Oceanographic Data Centres.
- UNESCO, I. O. C. of. (2013b). Ocean Data Standards. Paris: UNESCO. Retrieved from https://www.nodc.noaa.gov/oceanacidification/support/MG54_3.pdf
- Universidad de Alcalá. (2005). Métodos de Análisis de Datos en Ecología (pp. 1–31). Universidad de Alcalá, Departamento de Ecología. Retrieved from <https://www.uco.es/servicios/informatica/windows/filemgr/download/ecolog/Metodos analisis datos.pdf>
- Villalobos Flores, R., & Reneta, J. A. (2000). El Niño: Una Revisión Bibliográfica. *Instituto Meteorológico Nacional Gestión de Desarrollo*, (Jiménez 1987), 1–9.
- Weston, S. (2011). An Overview of Environmental Monitoring and its Significance in Resource and Environmental Management, 1–9.