

SERIE  
PROGRAMA DE  
PROSPERIDAD DEL  
REINO UNIDO EN  
COLOMBIA.

Desarrollo  
del Seguro  
Agropecuario

# Alistamiento de la información requerida para el componente de riesgo agroclimático y de los mapas de riesgo actuarial en el Sistema de Información para la Gestión de Riesgos Agropecuarios (SIGRA).



Embajada Británica  
Colombia

**CAF** BANCO DE DESARROLLO  
DE AMÉRICA LATINA

Alistamiento de la información requerida para el componente de riesgo agroclimático y de los mapas de riesgo actuarial en el Sistema de Información para la Gestión de Riesgos Agropecuarios (SIGRA).

**Editores Académicos:**

Francisco Boshell, Consultor  
Miguel Arango, Especialista  
Senior DAETSP, VSP, CAF

**Autores:**

Unión Temporal Proyecto SIGRA

**Revisión:**

CAF  
Embajada Británica en Colombia

Noviembre del 2022

# Alistamiento de la información requerida para el componente de riesgo agroclimático y de los mapas de riesgo actuarial en el Sistema de Información para la Gestión de Riesgos Agropecuarios (SIGRA).



Embajada Británica  
Colombia



Proyecto administrado por CAF y financiado por el gobierno británico a través del Programa de Prosperidad del Reino Unido en Colombia



# Prólogo

El Reino Unido y Colombia han forjado una estrecha relación basada en objetivos y ambiciones comunes. Muestra de esta cooperación es nuestro Programa de Prosperidad, presente en Colombia desde el 2017. Nuestro objetivo es fomentar el desarrollo económico inclusivo del país en 3 ejes claves: agricultura, infraestructura y fortalecimiento institucional; siempre teniendo presente un componente de enfoque de género e inclusión social, elementos transversales en todos nuestros proyectos.

Para avanzar en el logro de nuestro propósito, el Programa de Prosperidad ha establecido con CAF – banco de desarrollo de América Latina una alianza estratégica para la implementación de varios de sus proyectos en el país. Esta serie de publicaciones tienen como objetivo resaltar y compartir ampliamente algunos de los hallazgos y resultados más importantes que hemos obtenido en el marco de la cooperación entre el Reino Unido y Colombia, buscando que sean del mayor beneficio para el país; y animarles a explorar más a profundidad la contribución del gobierno británico y su Programa de Prosperidad al futuro de Colombia.

En particular, con la serie de Gestión Integral de Riesgos Agropecuarios, queremos difundir una visión integral de las herramientas y oportunidades para identificar, reducir y transferir los riesgos agropecuarios, con un enfoque diferencial.

En esta serie de proyectos, se priorizó el riesgo agroclimático y los cultivos de papa, maíz tecnificado, arroz seco y la ganadería de carne/leche. La alta exposición de los productores colombianos, en particular mujeres y población vulnerable, a eventos de variabilidad climática por su ubicación geográfica, y su alta vulnerabilidad a este riesgo, han sido una de las principales razones para priorizar estos cultivos. Esperamos que los resultados y los datos entregados en el programa continúen fortaleciendo la interinstitucionalidad de la agricultura colombiana y se conviertan en herramientas de política pública que permitan gestionar otros riesgos en diferentes sistemas productivos, incluyendo cacao, frutales y forestales.

**George Hodgson**  
*Embajador Británico en Colombia*

## Tabla de contenido

Índice de tablas	08
Tabla de ilustraciones	11
Acrónimos	12
<b>1 Introducción</b>	<b>14</b>
<b>2 Objetivo y alcance</b>	<b>18</b>
2.1. Objetivo general	19
2.2. Alcance del documento	19
<b>3 Riesgo Agroclimático</b>	<b>20</b>
3.1. Amenaza	21
3.1.1 Contexto histórico	21
3.1.1.1 Series climáticas históricas del IDEAM	23
3.1.1.2 Series climáticas históricas del portal datos abiertos	39
3.1.1.3 Series históricas de precipitación de CHIRPS	42
3.1.1.4 Datos de elevación de la superficie de la tierra Shuttle Radar Topography Mission (SRTM)	42
3.1.2 Contexto de monitoreo	43
3.1.2.1 Información diaria del portal de datos abiertos	44
3.1.2.2 Información diaria del portal de OSPA	55
3.1.2.3 Datos satelitales CHIRPS	59
3.1.2.4 Datos de lluvia satelital del hidroestimador	60
3.1.2.5 Información de precipitación de Fedearroz	60
3.1.3 Contexto de pronóstico	62
3.1.3.1 WRF - Pronóstico	63
3.1.4 Contexto de predicción climática	63
3.1.4.1 Datos del modelo CPT	63
3.2 Vulnerabilidad	64

3.2.1 Parámetros conservativos utilizados en las modelaciones de cultivos involucradas en análisis del componente vulnerabilidad	64
3.2.1.1 Parámetros de calibración de cultivos en las modelaciones de rendimientos	64
3.2.1.2 Fechas de siembra o calendarios de cultivo	66
3.2.1.3 Atributo de textura del suelo - Capa de correlación	66
3.2.1.4 Datos de fenología (coeficientes de cultivos kc)	67
3.2.1.5 Parámetros de calibración forrajes, como referencia para estimación de vulnerabilidad en sistemas productivos ganaderos (caso pasturas)	68
3.2.2 Contexto histórico	69
3.2.2.1 Concentración de dióxido de carbono (CO <sub>2</sub> )	69
3.2.2.2 Evaluaciones agropecuarias Municipales (EVA) - Estadísticas de productividad agrícola	69
3.2.3 Contexto de monitoreo	77
3.2.3.1 Datos de concentración de dióxido de carbono atmosférico (gas carbónico o CO <sub>2</sub> )	77
3.3 Exposición	77
3.3.1 Matriz de coeficientes técnicos	79
3.3.2 Vectores de precios de factores de producción	79
<b>4 Conclusiones</b>	<b>80</b>
Glosario	82
Bibliografía	87

## Índice de tablas

<b>Tabla 1:</b> resumen de datos priorizados para el cálculo de riesgos agroclimáticos en el contexto histórico	<b>22</b>
<b>Tabla 2:</b> comparación de estaciones de UPRA (VPN), IDEAM y Otras Entidades con el Catálogo de datos abiertos.	<b>25</b>
<b>Tabla 3:</b> controles de calidad de datos diarios de precipitación con estaciones automáticas del IDEAM.	<b>26</b>
<b>Tabla 4:</b> controles de calidad de datos diarios de precipitación con estaciones convencionales del IDEAM.	<b>27</b>
<b>Tabla 5:</b> controles de calidad a datos diarios de temperatura máxima del IDEAM.	<b>28</b>
<b>Tabla 6:</b> controles de calidad a datos diarios de temperatura mínima del IDEAM.	<b>29</b>
<b>Tabla 7:</b> controles de calidad a datos diarios de humedad relativa 1pm	<b>31</b>
<b>Tabla 8:</b> Controles de calidad a datos diarios de humedad relativa promedio.	<b>32</b>
<b>Tabla 9:</b> controles de calidad a datos diarios de brillo solar.	<b>33</b>
<b>Tabla 10:</b> controles de calidad a datos subhorarios de velocidad de viento.	<b>34</b>
<b>Tabla 11:</b> cantidad de datos para la calcular datos de viento a escala horaria	<b>35</b>
<b>Tabla 12:</b> cantidad de datos para calcular la velocidad máxima del viento absoluto	<b>35</b>
<b>Tabla 13:</b> controles de calidad a datos horarios de radiación solar global estaciones automáticas.	<b>36</b>
<b>Tabla 14:</b> transformación de los datos de radiación solar global.	<b>37</b>
<b>Tabla 15:</b> resultados de estaciones después de la homogenización	<b>38</b>
<b>Tabla 16:</b> controles de calidad aplicados a los datos históricos de velocidad de viento del portal de datos abiertos.	<b>40</b>
<b>Tabla 17:</b> cantidad de datos para la calcular datos de viento a escala horaria	<b>41</b>
<b>Tabla 18:</b> cantidad de datos para calcular la velocidad máxima del viento absoluto	<b>41</b>



<b>Tabla 19:</b> resultados de estaciones después de la homogenización de datos históricos de datos abiertos.	<b>42</b>
<b>Tabla 20:</b> resumen de datos priorizados para el cálculo de riesgos en el contexto de monitoreo	<b>43</b>
<b>Tabla 21:</b> enlaces de descarga de las variables meteorológicas a través de datos abiertos.	<b>44</b>
<b>Tabla 22:</b> cantidad de datos para acumular registro de precipitación a escala horaria	<b>46</b>
<b>Tabla 23:</b> criterios para acumular registros de precipitación a escala diaria	<b>46</b>
<b>Tabla 24:</b> controles de calidad propuestos para ser aplicados a los datos de precipitación provenientes del portal de datos abiertos y agregados a escala diaria.	<b>47</b>
<b>Tabla 25:</b> cantidad de estaciones para cada grupo de disponibilidad de datos de precipitación.	<b>48</b>
<b>Tabla 26:</b> criterio para seleccionar los grupos de representatividad para la temperatura diaria.	<b>49</b>
<b>Tabla 27:</b> controles de calidad propuestos para los datos de monitoreo de temperatura máxima y mínima del portal de datos abiertos.	<b>49</b>
<b>Tabla 28:</b> cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura máxima.	<b>50</b>
<b>Tabla 29:</b> cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura mínima.	<b>50</b>
<b>Tabla 30:</b> controles de calidad propuestos para los datos de monitoreo de la variable velocidad de viento del portal de datos abiertos.	<b>51</b>
<b>Tabla 31:</b> cantidad de datos para generar acumulados de viento a escala horaria	<b>51</b>
<b>Tabla 32:</b> cantidad de datos para generar la velocidad máxima absoluta del viento.	<b>52</b>
<b>Tabla 33:</b> cantidad de estaciones para cada base de datos creada con velocidad de viento.	<b>52</b>
<b>Tabla 34:</b> criterio para seleccionar los grupos de representatividad para la humedad diaria	<b>53</b>
<b>Tabla 35:</b> controles de calidad propuestos para los datos de monitoreo de humedad relativa del portal de datos abiertos.	<b>54</b>
<b>Tabla 36:</b> cantidad de estaciones para cada grupo de disponibilidad de datos de humedad relativa.	<b>55</b>

---

<b>Tabla 37:</b> controles de calidad propuestos para los datos de precipitación de OSPA.	<b>56</b>
<b>Tabla 38:</b> cantidad de estaciones para cada grupo de disponibilidad de datos de precipitación de la OSPA.	<b>57</b>
<b>Tabla 39:</b> controles de calidad propuestos para los datos de monitoreo de temperatura máxima y mínima de OSPA.	<b>58</b>
<b>Tabla 40:</b> cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura máxima de OSPA.	<b>59</b>
<b>Tabla 41:</b> cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura mínima de OSPA.	<b>59</b>
<b>Tabla 42:</b> controles de calidad aplicados a los datos de precipitación diarios de Fedearroz.	<b>62</b>
<b>Tabla 43:</b> intervalos de confianza para el rendimiento (t/ha)	<b>75</b>
<b>Tabla 44:</b> resultado final análisis EVA cruzado con el índice ONI	<b>76</b>

---

## Tabla de ilustraciones

<b>Ilustración 1:</b> resultados de Climatol para el conjunto de datos de Radiación Solar	<b>38</b>
<b>Ilustración 2:</b> solo para fines ilustrativos, Atributo de textura a nivel Nacional	<b>67</b>
<b>Ilustración 3:</b> boxplot datos de rendimiento (t/ha) y gráfico de correlación de Maíz entre CNA y EVAs	<b>71</b>
<b>Ilustración 4:</b> matriz de correlación Papa	<b>72</b>
<b>Ilustración 5:</b> boxplot de rendimientos (t/ha) y gráfico de correlación para Papa en CNA y en EVA	<b>72</b>
<b>Ilustración 6:</b> matriz de correlación Arroz	<b>73</b>
<b>Ilustración 7:</b> boxplot de rendimientos (t/ha) y gráfico de correlación para Arroz en CNA y en EVA	<b>73</b>
<b>Ilustración 8:</b> boxplots de distribución de rendimientos (t/ha)	<b>74</b>
<b>Ilustración 9:</b> distribución conceptual de las matrices de coeficientes técnicos	<b>78</b>

## ACRÓNIMOS

**AGROSAVIA:** Corporación colombiana de investigación agropecuaria.

**ANDI:** Asociación Nacional de Empresarios de Colombia.

**BMC:** Bolsa Mercantil de Colombia.

**CAF:** Banco de Desarrollo de América Latina.

**CENICAFE:** Centro de Investigación del Café.

**CENICAÑA:** Centro de Investigación de la Caña de Azúcar de Colombia.

**CGIAR:** Grupo Consultivo para la Investigación Agrícola Internacional.

**CHIRPS:** Climate Hazards Group InfraRed Precipitation with Station.

**CIAT:** Centro Internacional de Agricultura Tropical.

**DANE:** Departamento Administrativo Nacional de Estadística.

**DEM:** Modelo Digital de Elevaciones.

**ENOS:** El Niño Oscilación del Sur.

**ET0 o ET<sub>o</sub>:** Evapotranspiración de Referencia.

**EVA:** Evaluaciones Municipales Agropecuarias.

**ESRL:** Earth System Research Laboratory de NOAA.

**FAO:** Food and Agriculture Organization - Organización de las Naciones Unidas para la Alimentación y la Agricultura.

**FEDEARROZ:** Federación Nacional de Productores de Arroz.

**FEDEGAN:** Federación Colombiana de Ganaderos.

**FEDEPAPA:** Federación Colombiana de Productores de Papa.

**FENALCE:** Federación Nacional de Cereales y Leguminosas.

**FENALCE:** Federación Nacional de Cultivadores de Cereales, Leguminosas y Soya.

**FINAGRO:** Fondo para el Financiamiento del Sector Agropecuario.

**FNC:** Federación Nacional de Cafeteros de Colombia.

**GML:** ESRL's Global Monitoring Laboratory.

**ha:** Hectárea.

**IDEAM:** Instituto de Hidrología, Meteorología y Estudios Ambientales.

**IDH:** Índice de Disponibilidad Hídrica.

**IGAC:** Instituto Geográfico Agustín Codazzi.

**IRI:** International Research Institute for Climate and Society.

**ISRIC:** International Soil Reference and Information Centre.

**MADR:** Ministerio de Agricultura y Desarrollo Rural.

**NASA:** Administración Nacional de Aeronáutica y el Espacio.

**NCEP:** National Center for Environmental Prediction.

**NOAA:** National Oceanic and Atmospheric Administration.

**ODEPA:** Oficina de Estudios y Políticas Agrarias.

**OMM:** Organización Meteorológica Mundial.

**OSPA:** Oficina de Servicios, Pronósticos y Alertas de IDEAM.

**PDECI:** Índice de la Condición de la diferencia entre Evapotranspiración y Precipitación.

**PDSI:** Índice de Severidad de Sequía de Palmer.

**RPA:** Robotic Process Automation.

**SIATA:** Sistema de Alerta Temprana del Valle de Aburrá.

**SIGRA:** Sistema de Información para la Gestión de Riesgos Agropecuarios.

**SIPRA:** Sistema de Información para la Planificación Rural Agropecuaria.

**SPI:** Índice de Precipitación Estandarizada.

**UPME:** Unidad de Planeación Minero-Energética.

**UPRA:** Unidad de Planificación Rural Agropecuaria.

**USDA:** United States Department of Agriculture - Departamento de Agricultura de los Estados Unidos.

**WRSI:** Índice de Satisfacción de Requerimientos Hídricos.



# 01 Introducción

En el año 2015 el gobierno británico creó el Fondo de Prosperidad como parte de su plan de apoyo a los Objetivos de Desarrollo Sostenible de las Naciones Unidas, así como para promover el crecimiento y la prosperidad en países en vías de desarrollo. Por su parte, en noviembre de 2017 CAF - banco de desarrollo de América Latina y el entonces Fondo de Prosperidad del Reino Unido en Colombia, celebraron un convenio de Asistencia Técnica y Colaboración, a través del cual se acordó la administración de recursos del Reino Unido por parte de CAF.

Los recursos administrados están destinados a apoyar proyectos del ahora Programa de Prosperidad del Reino Unido en Colombia (PP), dirigidos a reducir la pobreza y generar igualdad a través del desarrollo económico inclusivo de mujeres y población vulnerable tradicionalmente excluidas; crear oportunidades comerciales para empresas británicas; y desarrollar proyectos de inclusión económica en áreas vulnerables de la República de Colombia.

Con este fin, el Reino Unido autorizó que los recursos administrados fueran destinados para desarrollar tres actividades fundamentales como son: i) desarrollar las capacidades de los gobiernos locales; ii) apoyar la creación de estándares y capacidad de Alianzas Público-Privadas (APP) y las agencias relacionadas; y iii) desarrollar la gestión de riesgos agropecuarios en Colombia. En particular, la iniciativa de desarrollo de seguros agropecuarios, de la cual hace parte este documento, busca incrementar la penetración del seguro agrícola en Colombia y expandir su mercado, para aumentar la resiliencia del sector.

En desarrollo de lo anterior, CAF - banco de desarrollo de América Latina, como una de las entidades implementadora de los recursos del Programa de Prosperidad del Reino Unido en Colombia, y con recursos de éste, diseñó e implementó el Sistema de Información para la Gestión de Riesgos Agropecuarios (SIGRA) con la UT-SIGRA conformada por las empresas EY, CEBAR y Servinformación.<sup>1</sup>

En todos los trabajos realizados en el marco del PP, se ha prestado especial atención al enfoque diferencial por tamaño de productor y género, con énfasis en las necesidades de capacitación y apoyo de productores pequeños y mujeres productoras. La CAF desarrolló un Marco de Género e Inclusión (MGI) con base en los lineamientos proporcionados por el PP, por lo cual, todos los proyectos adquirieron una perspectiva de género y atención especial de grupos vulnerables.

El Marco de Género e Inclusión reconoce la diversidad humana, la inequidad existente y la imperiosa necesidad de lograr el ejercicio de derechos por parte de toda la población con una perspectiva de equidad, en especial aquella que se encuentra en condiciones de vulnerabilidad, por lo que ofrece una conceptualización del enfoque diferencial especialmente de género e inclusión, que permite cerrar brechas de desigualdad por razones de género o grupos tradicionalmente excluidos. Por lo anterior, en los diversos proyectos se caracteriza de manera amplia la población y se identifica de forma particular los impactos en mujeres y grupos vulnerables, así como el desarrollo de acciones afirmativas relacionadas directamente con los productos o resultados de los proyectos en tres niveles como son: i) Voz, ii) Empoderamiento y iii) Transformación.

El resultado del marco de género e inclusión es que en los proyectos del PP se identifican los efectos que tienen en mujeres y grupos vulnerables, se incorpora la acción sin daño, se realizan diseños sensibles al género y se elaboran recomendaciones que permiten abonar el camino hacia la transformación institucional y el empoderamiento de grupos excluidos.

Con el presente documento se busca divulgar resultados importantes obtenidos en el desarrollo progresivo del proyecto SIGRA, en materia de alistamiento adecuado de la información básica climática, agroeconómica y agroclimática necesaria para desarrollar modelos de simulación y calcular componentes del riesgo agroclimático, como amenaza, vulnerabilidad y exposición, así como para estimar indicadores actuariales del riesgo agroclimático. De este modo, se presentan procesos de inventarios, control de calidad, homogenización, estandarización y establecimiento de bases de información para esos datos, según metodologías aplicadas en el proyecto SIGRA, considerando las condiciones de infraestructura computacional que se viene desarrollando para su implementación en UPRA.

1. Equipo de trabajo de la Unión Temporal conformado por las empresas EY, CEBAR y Servinformación.

El proceso de alistamiento de información cubre los análisis temporales (en escalas histórica, de monitoreo y de pronóstico/predicción) requeridos para evaluar de modo apropiado componentes del cálculo de riesgo agroclimático citados y para futuros procesamientos necesarios para la generación de productos de información de riesgo agroclimático y riesgo agroclimático con criterio actuarial, que contribuirán con la gestión de riesgos para el sector agropecuario.

Por lo anterior, se considera que este trabajo es un valioso aporte para conocer cómo se analiza y procesa la información agroclimática necesaria para evaluar el riesgo agroclimático y

para realizar análisis agroclimáticos con criterio actuarial (que se utilizan, por ejemplo, en estudios técnicos de seguros agropecuarios), y por ello amerita ser divulgado para conocimiento de actores gremiales, investigativos, académicos e institucionales asociados al sector agropecuario colombiano. Es necesario anotar que mayores detalles sobre los componentes del riesgo ya anotados y sobre los elementos de gestión del riesgo agroclimático (conocimiento, prevención, reducción, asunción, transferencia) se pueden consultar en otras publicaciones relacionadas con los proyectos desarrollados en el marco del Programa de Prosperidad del Reino Unido – CAF, sobre riesgos y seguros agropecuarios.





# 02 Objetivo y alcance

- 2.1 Objetivo general
- 2.2 Alcance del documento

## 2.1 Objetivo general

Describir los procesos que se implementaron para alistar la información agroclimática, de forma coherente con la información disponible y requerida para la construcción de los productos del SIGRA y con las características de la infraestructura computacional que se viene desarrollando para su implementación.

## 2.2 Alcance del documento

En este documento se describen diversos procesos utilizados para el alistamiento, incluido normalización, control de calidad y análisis de homogeneidad, de información climática y agroclimática básica, en diversas escalas temporales (Histórica, Monitoreo y Pronóstico), relacionada con cálculos de componentes del riesgo agroclimático, como amenaza, vulnerabilidad y exposición, y con análisis del riesgo agroclimático con criterio actuarial en municipios priorizados.

# 03 Riesgo Agroclimático

- 3.1 Amenaza
- 3.2 Vulnerabilidad
- 3.3 Exposición

**Los procesos de alistamiento de los datos y productos de información agroclimática dependen de las características de los datos, los mecanismos de entrega y de los formatos en los que son entregados por parte de las fuentes de información.**

En el contexto del SIGRA se han utilizado diferentes tipos de datos y productos de información, como datos y series de datos georreferenciados, datos y series de datos espaciales en grillas Ráster y/o polígonos en formatos espaciales Ráster y Vector, entre otros.

Los procesos de alistamiento de información que a continuación se relacionan corresponden a los conjuntos de datos disponibles para el cálculo del riesgo

agroclimático, desagregados en sus componentes de amenaza, vulnerabilidad y exposición. Los datos fueron sometidos a diferentes procesos generales. El primero consistió en el alistamiento básico de la información, donde los datos fueron sometidos a procesos de descarga, ordenamiento, estandarización y normalización; y el segundo, correspondió a controles de calidad, homogenización y carga en los repositorios de BigQuery y CloudStorage para su disposición.

## 3.1 Amenaza

Para este componente del riesgo agroclimático se describen los conjuntos de datos utilizados según las amenazas priorizadas como son el déficit y el exceso hídrico, las heladas y los vientos fuertes, para cada contexto temporal propuesto, que será incluido en el SIGRA, como el histórico, el de monitoreo, el de pronóstico (corto plazo) y el de predicción (escala estacional).

### 3.1.1 Contexto histórico

En el caso del contexto histórico de cálculo de las amenazas agroclimáticas, las fuentes de información

consisten principalmente en series históricas de datos climáticos registrados por estaciones meteorológicas terrenas y series de datos espaciales en puntos de grilla provenientes de modelos climáticos y observaciones satelitales. También se seleccionaron series de mapas de índices agroclimáticos en formatos de imagen generados por el IDEAM<sup>2</sup>. Como principal fuente de información climática para el SIGRA, se presentan los procesos requeridos para el alistamiento de los conjuntos de datos climáticos históricos de la red del IDEAM. La siguiente lista resume los conjuntos de datos para el cálculo de riesgo en el contexto histórico:

**Tabla 1: resumen de datos priorizados para el cálculo de riesgos agroclimáticos en el contexto histórico**

Fuentes de datos priorizadas	Temporalidad	Periodo
Series climáticas históricas del IDEAM.	Brillo Solar, Humedad Relativa: horaria; Precipitación, Temperatura máx, Temperatura mín: diaria. Velocidad de viento: subhoraria Radiación global convencional: diaria Radiación global automática: horaria	Depende de cada variable y estación, en el periodo 1981 - 2019.
Series climáticas históricas del IDEAM descargadas del portal DATOS ABIERTOS <sup>3</sup> .	Vientos: Subhorarios	2004-2020.
Series históricas de precipitación CHIRPS.	Pentadal, decadal, mensual.	1981 - 2021.
Datos de elevación de la superficie de la tierra Shuttle Radar Topography Mission (SRTM).	N/A	N/A

Fuente: Elaboración propia

Es importante mencionar que las series históricas de datos climáticos registrados en estaciones meteorológicas georreferenciadas pueden contener datos erróneos asociados con eventos ajenos a la variabilidad climática natural de cada elemento climático. Estos errores pueden deberse a problemas en la toma, almacenamiento o transmisión de los registros, a cambios o mal funcionamiento en el instrumental utilizado, por la ubicación de la estación o por su entorno. Estos últimos pueden ser cambios bruscos, como el incendio de un bosque colindante, o inundación del emplazamiento de la estación, o graduales como la posterior recuperación de la vegetación o la urbanización en áreas circundantes a la estación. Estas alteraciones de las series, denominadas “inhomogeneidades” o “heterogeneidades”, enmascaran verdaderas fluctuaciones del clima y hacen que el estudio de las series conduzca a conclusiones erróneas (Gujarro, 2018).

Para abordar este problema se han desarrollado metodologías de homogeneización que buscan eliminar o reducir en lo posible estas alteraciones indeseadas. Estas metodologías se basan principalmente en construir una serie de referencia homogénea a partir del promedio de otras series seleccionadas por su proximidad o elevada correlación, comparar y ajustar cada serie con base en las series de referencia, diluyendo así sus posibles inhomogeneidades y posteriormente, aplicar procesos de homogeneización y complementación de datos faltantes mediante herramientas como Climatol (Gujarro, 2018), a las series históricas diarias provenientes de estaciones meteorológicas suministradas por IDEAM mediante la VPN o descargadas del portal de datos abiertos de dicha entidad.

3 [www.datos.gov.co](http://www.datos.gov.co). La iniciativa de datos abiertos busca que todas las entidades gubernamentales publiquen su información en formatos estructurados, para ponerla a disposición de los usuarios con criterios de pertinencia y calidad. La información es compartida públicamente en formatos digitales estandarizados con una estructura de fácil comprensión.

Por otra parte, se deben tener en cuenta los siguientes aspectos para lograr la transformación de la información desde el contexto de monitoreo (contexto de tiempo actual o cercano al actual) al contexto histórico:



La información para el monitoreo obtenida del portal de Datos Abiertos o suministrada por la Oficina de Servicios y Pronósticos del IDEAM (OSPA) se somete a controles básicos de calidad, basados en los controles aplicados rutinariamente por el IDEAM.



Posteriormente, esta información se utiliza para dar continuidad a las series completas históricas hasta el año más reciente con la connotación de "información preliminar".



Una vez dicha información preliminar se actualice por IDEAM, asegurando su oficialidad, las series de datos y productos relacionados se actualizan, y su denominación de preliminar se retira.

### 3.1.1.1 Series climáticas históricas del IDEAM

El proceso de alistamiento de los datos históricos suministrados por el IDEAM incluye varias etapas, empezando por el reconocimiento y organización de los archivos y datos, que hacen parte del proceso de normalización y estandarización. Posteriormente, se hace la aplicación de los controles de calidad de los datos, lo cual se describe a continuación:

#### **A. Alistamiento de datos:**

##### **i. Normalización: reconocimiento y organización de la información**

El proceso empieza con la revisión de las carpetas suministradas por el IDEAM, identificando archivos de texto con datos para cada estación meteorológica y variable climática en los archivos planos (.data), así como del catálogo de estaciones y el glosario. Luego, se construye un script en Python para recorrer las carpetas de la VPN entregada por la UPRA<sup>4</sup>, entidad beneficiaria del proyecto SIGRA, , para cada variable cli-

mática (IDEAM/Historicos/VariableClimatica/ID3026). Este script toma el nombre del archivo, lo separa y de ahí extrae el nombre de la variable, tipo de estación, periodicidad del dato y el código de la estación. Del contenido del archivo extrae la fecha y el valor del dato, luego, toda esta información se consolida, definiendo una estructura de campos. Finalmente, se procede a validar formatos de los campos y se cargan a la herramienta BigQuery.

Los datos recibidos corresponden a escala diaria para las variables de precipitación, temperatura máxima y mínima, escala horaria para brillo solar, y humedad relativa y escala subhoraria (menor a una hora), para velocidad de viento. Cada registro cuenta con un campo de fecha-hora de reporte, el cual fue revisado para encontrar la cantidad de registros para cada una de las horas del día y para identificar la frecuencia de reporte de las estaciones en los datos subhorarios. A continuación, se describen los resultados encontrados:



Se validó que el 99.99% de los registros de precipitación corresponden a reportes de las 7 horas de cada día (7:00 a.m.), lo que implica que cada dato corresponde a la cantidad de lluvia acumulada desde las 7:00 a.m. del día anterior a las 7:00 a.m. de la fecha del reporte.



El 99.98% de los datos de temperatura máxima se reportan a las 19 horas (7:00 p.m.), mientras que el 99.99% de los datos de temperatura mínima se reportan a las 7 horas (7:00 a.m.).



Los datos de brillo solar se reportan en escala horaria, distribuidos proporcionalmente desde las 5 horas (5:00 a.m.) hasta las 18 horas (6:00 p.m.), con un aproximado de 7% de los datos para cada hora del día. Se encontraron reportes de brillo solar para las 4:00 a.m. y las 19 horas, que representan apenas un 0.01% de los datos.



El 34,3% de los datos de humedad relativa tienen hora de reporte a las 7 horas, el 32,7% a las 13 horas (1:00 p.m.), el 23,6% a las 19 horas, el 7,2% a las 18 horas (6:00 p.m.) y el 2,1% se distribuye en el resto de las horas del día, con muy pocos datos en horas nocturnas.



Para los datos de velocidad de viento subhorarios reportados a lo largo del día, se encontró que el 83.93% de los datos corresponden a 326 estaciones cuyos datos se procesaron originalmente en IDEAM en períodos de 10 minutos y el 16.07% de los datos corresponden a 212 estaciones con reportes en períodos de 60 minutos.

## ii. Estandarización

Para aplicar los controles de calidad a los datos históricos, se agregan campos nuevos a todas las tablas de cada variable en Big Query. Se agrega el campo 'control' que tiene un valor inicial de 'NA'; luego de ejecutar los controles de calidad, se identifican los datos marcados en cada control, mediante los valores 'control-1', 'control-2', según corresponda. El segundo campo que se agrega es el 'estado', que tiene un valor inicial de 'ok' para todos los registros; luego de aplicar los controles se conserva el valor de 'ok' si el dato pasa todos los controles de calidad, o toma los valores de 'Error', 'Sospechoso', 'Atípico mensual', 'Super atípico mensual', según como se identifique el dato en los controles. El tercer campo es el 'valor procesado' que tiene un valor inicial igual al campo 'valor' de la tabla; luego de la ejecución de los controles, toma el valor de 99999 cuando el dato se identifica como 'Error' en algún control. En caso contrario, mantiene el mismo valor inicial.

## iii. Catálogo de estaciones del IDEAM:

Los datos climáticos corresponden a series históricas de estaciones georreferenciadas cuyas ubicaciones y descripción están contenidas en el catálogo de estaciones suministrado por el IDEAM, archivo *CNE\_IDEAM*. En el contexto del proyecto SIGRA fue necesario concatenar las series de datos climatológicos, según el código de estación que titulaba cada archivo, con los códigos de las estaciones disponibles en el catálogo del IDEAM. La cantidad de códigos disponibles en los archivos de datos suministrados por IDEAM es diferente a los disponibles en otros catálogos, como el suministrado a la Consultora por UPRA en la VPN. Por lo tanto, fue necesario repetir este proceso, con el catálogo de estaciones descargado del portal de datos abiertos (<https://www.datos.gov.co/api/views/hp9rjxuu/rows.CSV?accessType=DOWN-LOAD>) para relacionar todos los datos suministrados u obtenidos de IDEAM por diversos canales. Este proceso permitió relacionar los códigos de estacio-



nes con los registros, entre horarios y diarios, para las variables precipitación de estaciones convencionales y automáticas, temperatura máxima y mínima, brillo solar, humedad relativa y velocidad del viento.

Se concluyó que el catálogo descargado desde el portal de datos abiertos contiene la mayoría de es-

taciones de los catálogos de UPRA (VPN), IDEAM y Otras Entidades (disponible en el Sistema de Información de Datos Hidrológicos y Meteorológicos del IDEAM-DHIME). En la Tabla 2 se muestra el porcentaje de coincidencia de estaciones de los 3 catálogos con el catálogo de datos abiertos.

**Tabla 2: comparación de estaciones de UPRA (VPN), IDEAM y Otras Entidades con el Catálogo de datos abiertos.**

Nombre Catálogo	Cantidad de estaciones que cruzaron espacialmente	Total de estaciones	Porcentaje de coincidencia de estaciones en Catálogo Datos Abiertos
Catálogo UPRA (VPN)	4.436	4.442	99.8%
Catálogo IDEAM	4.436	4.486	98.8%
Catálogo Otras Entidades	4.529	4.587	98.7%

Fuente: Elaboración propia

## B. Procesamiento de datos

### i. Control de calidad de los datos

Se implementó un proceso de control de calidad de los datos meteorológicos, orientado a identificar datos atípicos y sospechosos que fueron marcados para posteriores análisis. Los datos identificados como errores fueron retirados de las series climáticas.

El proceso fue desarrollado originalmente por Sorza (2014), en el marco del convenio interinstitucional entre AGROSAVIA<sup>5</sup> e IDEAM, en el proyecto "Reducción del Riesgo y Adaptación al Cambio Climático", para mejorar la calidad de las series meteorológicas de la base de datos del IDEAM para el periodo de 1981-2011. Este protocolo fue discutido y construido con base en aportes de profesionales de la subdirección de meteorología del IDEAM e incluyó las siguientes etapas para cada una de las variables:

### • Conjunto de datos de precipitación

**Descripción:** conjunto de datos de precipitación para estaciones automáticas y convencionales disponibles en Colombia a escala diaria.

#### 1. Etapa de Revisión y preparación

**Paso 1:** agregación de registros de fecha-hora no existente y su respectivo valor con la convención de dato faltante (99999).

**Paso 2:** identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha u hora.

*II. Etapa de Control de calidad para  
datos de precipitación en escala diaria*

A continuación, se muestran los controles de calidad aplicados y la cantidad de datos marcados por cada uno de estos, para precipitación registrada en estaciones automáticas en la **Tabla 3** y para precipitación medida en estaciones convencionales en la **Tabla 4**.

**Tabla 3: controles de calidad de datos diarios de precipitación con estaciones automáticas del IDEAM.**

CONTROLES DIARIOS PRECIPITACIÓN (Estaciones Automáticas)					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Control 1 (Revisión de rangos)	Error	ppt<0 mm o ppt>350mm	control='control-1', estado='Error', valor_procesado= 99999	64	0.013%
	Sospechoso	200 mm<ppt<_ 350 mm	Control='control-1', estado=' Sospechoso', valor_procesado= valor	57	0.011%
Control 3 (Valores atípicos respecto al mes)	Atípico mensual	Dato>Cuartil 3+RIC*R Dato<Cuartil1 – RIC*R donde R=5,RIC=Q3 –Q1	control='control-3', estado='Atípico mensual', valor_procesado=valor	3775	0.765%
Control 4 (identificación de consecutivos sin tener en cuenta (0))	Error	Más de 3 días consecutivos	control='control-4', estado='Error', Valor_procesado= 99999	1727	0.350%
	Sospechoso	2 y 3 días consecutivos	control= 'control-4', estado= 'Sospechoso', Valor_procesado= valor	11355	2,303%
Control 6 (Valores súper atípicos del mes)	Súper atípico mensual	Dato>Q3 + RIC*R Dato<Q1 – RIC *R Donde R=6,RIC=Q3–Q1	Control= 'control-6', estado= 'Súper atípico mensual', valor_procesado= valor	32889	6.669%
Datos limpios	-	-	Control=', NA', estado='ok', valor_procesado=valor	443257	89,887%

Fuente: Elaboración propia

**Tabla 4: controles de calidad de datos diarios de precipitación con estaciones convencionales del IDEAM.**

CONTROLES DIARIOS PRECIPITACIÓN (Estaciones convencionales)					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Control 1 (Revisión de rangos)	Error	ppt<0 mm o ppt>350mm	control=' control-1', estado='Error', valor_procesado= 99999	38	1,03e-4%
	Sospechoso	200 mm<ppt<_ 350 mm	control='control-1', estado='Sospechoso', valor_procesado= valor	113	3,06e-4%
Control 3 (Valores atípicos respecto al mes)	Atípico mensual	Dato>Cuartil 3+RIC*R Dato<Cuartil1 - RIC*R donde R=5, RIC=Q3 -Q1	control='control-3', estado='Atípico mensual', Valor_procesado= valor	216379	0,587%
Control 4 (identificación de consecutivos sin tener en cuenta (0)	Error	Más de 3 días consecutivos	control=' control-4', estado='Error', valor_procesado= 99999	8803	0,024%
	Sospechoso	2 y 3 días consecutivos	control='control-4', estado='Sospechoso', Valor_procesado= valor	260394	0,706%
Control 6 (Valores súper atípicos del mes)	Súper atípico mensual	Dato>Q3 + RIC*R Dato<Q1 - RIC *R Donde R=6, RIC=Q3-Q1	Control='control-6', estado='Súper atípico mensual', valor_procesado= valor	2209751	5,991%
Datos limpios	-	-	Control='NA', estado'ok', Valor_procesado= valor	34189443	92,692%

Fuente: Elaboración propia

• Conjunto de datos de temperatura máxima y mínima

**Descripción:** conjunto de datos de temperatura máxima y mínima para estaciones convencionales disponibles en Colombia a escala diaria.

I. Etapa de Revisión y preparación

**Paso 1:** agregación de registros de fecha-hora no existente y su respectivo valor con la convención de dato faltante (99999).

**Paso 2:** identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha u hora.

II. Etapa de Control de calidad para datos de Temperatura máxima y mínima en escala diaria

A continuación, se muestran los controles de calidad aplicados y la cantidad de datos marcados por cada uno de estos, para temperatura máxima en la **Tabla 5** y para temperatura mínima en la **Tabla 6**.

**Tabla 5: controles de calidad a datos diarios de temperatura máxima del IDEAM.**

CONTROLES DIARIOS TEMPERATURA MÁXIMA					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Control 1 (Revisión de rangos)	Error	Temp > 50OC Temp < 10oC	control=' control-1', estado ='Error', valor_procesado=99999	0	0%
Control 2 (Consistencia Tmin y Tmax)	Error	Tmin > Tmax	control='control-2', estado='Error', valor_procesado=99999	10	2,19e -4%
Control 3 (Valores atípicos respecto al mes)	Atípico mensual	Dato>Cuartil3 RIC*R Dato< Cuartil1 – RIC*R donde R=4, RIC=Q3-Q1	control='control-3', estado='Atípico mensual', valor_procesado= valor	406	0.009%
Control 4 (Identificación de consecutivos)	Error	Más de 4 días consecutivos	control='control-4', estado=Error', valor_procesado=99999	3179	0.069%
	Sospechoso	3 y 4 días consecutivos	Control='control-4', estado='Sospechoso', valor_procesado= valor	73080	1.597%
Control 5 (Diferencias entre datos)	Error	Diferencias > 16°C de 2 datos diarios consecutivos	control='control-5', estado='Error', valor_procesado=99999	24	5.25e-4%
	Sospechoso	Diferencia entre 7°C y 16°C entre 2 datos diarios consecutivos	control='control-5', estado='Sospechoso', valor_procesado=valor	12020	0.2635%
Datos limpios	-	-	control='NA', estado='ok', valor_procesado=valor	4486498	98.061%

**Tabla 6: controles de calidad a datos diarios de temperatura mínima del IDEAM.**

CONTROLES DIARIOS TEMPERATURA MÍNIMA					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Control 1 (Revisión de rangos)	Error	Temp > 50OC Temp < 10oC	control=' control-1', estado ='Error', valor_procesado= 99999	0	0%
Control 2 (Consistencia Tmin y Tmax)	Error	Tmin > Tmax	control='control-2', estado='Error', valor_ procesado=99999	10	1,95e-4%
Control 3 (Valores atípicos respecto al mes)	Atípico mensual	Dato>Cuartil3 RIC*R Dato< Cuartil1 – RIC*R donde R=4,RIC=Q3-Q1	control='control-3', estado='Atípico mensual', valor_procesado= valor	2238	0,043%
Control 4 (Identificación de consecutivos)	Error	Más de 4 días consecutivos	control='control-4', estado=Error', valor_procesado= 99999	6946	0,135%
	Sospechoso	3 y 4 días consecutivos	Control='control-4', estado='Sospechoso', valor_procesado= valor	119790	2,332%
Control 5 (Diferencias entre datos)	Error	Diferencias > 16°C de 2 datos diarios consecutivos	control='control-5', estado='Error', valor_procesado= 99999	94	0,002%
	Sospechoso	Diferencia entre 7°C y 16°C entre 2 datos diarios consecutivos	control='control-5', estado='Sospechoso', valor_procesado= valor	8432	0,164%
Datos limpios	-	-	control='NA', estado='ok', valor_procesado= valor	4999049	97,323%

Fuente: Elaboración propia

• Conjunto de datos de humedad relativa

**Descripción:** conjunto de datos de humedad relativa para estaciones convencionales a escala horaria.

I. Etapa de agrupación de datos a escala diaria.

Se generaron dos bases de datos de humedad relativa diaria descritas a continuación:

- Humedad relativa 1:00 p.m. diaria: se seleccionó para cada día el dato de las 13 horas, como dato diario y se adicionó el campo numero\_registros =1.
- Humedad relativa promedio diaria: se construyó con el promedio de los valores de humedad a las 7, 13 y 19 horas, si no está el dato exacto de la hora, se tomó el dato de 1 hora antes o después de la hora indicada. Si hace falta alguno de los 3 datos no se puede calcular el promedio y queda como dato faltante. Adición del campo numero\_registros,=3 para los datos generados con promedio y para los datos faltantes numero\_registros=1 y valor=99999.

II. Etapa de revisión y preparación

**Paso 1:** agregación de registros de fecha-hora no existente y su respectivo valor con la convención de dato faltante (99999) a las dos series diarias construidas de humedad relativa diaria.

**Paso 2:** identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha u hora.

III. Etapa de Control de calidad para datos de humedad relativa 1pm y humedad relativa promedio

A continuación, se muestran los controles de calidad aplicados y la cantidad de datos marcados por cada uno, para humedad relativa 1pm en la **Tabla 7** y para humedad relativa promedio en la **Tabla 8**.



**Tabla 7: controles de calidad a datos diarios de humedad relativa 1pm**

CONTROLES DIARIOS HUMEDAD RELATIVA 1PM					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Control 1 (Revisión de rangos)	Error	hr<0% o hr > 100%	control='control-1', estado='Error', valor_procesado=99999	0	0%
Control 3 (valores atípicos respect al mes)	Atípico mensual	Dato > Cuartil3 +RIC *R Dato < Cuartil1 – RIC *R donde R=3, RIC=Q3-Q1	control='control-3', estado='Atípico mensual', valor_procesado=valor	285	0,046%
Control 4 (Identificación de consecutivos)	Error	Más de 7 días consecutivos	control='control-4', estado='Error', valor_procesado=99999	16	0,002%
	Sospechoso	Entre 4 – 7 días consecutivos	control='control-4', estado='Sospechoso', valor_procesado=valor	2265	0,363%
Control 5 (Diferencias entre datos)	Error	Diferencias >55% de 2 Datos diarios consecutivos	control='control-5', estado='Error', valor_procesado=valor	255	0,041%
	Sospechoso	Diferencias entre 40 y 55% entre 2 diarios consecutivos	Control='control-5', estado=Sospechoso', valor_procesado=valor	5599	0,898%
Control 6 (valores super atípicos del mes)	Súper atípico mensual	Dato >Cuartil3 + RIC * R Dato < Cuartil1 – Ric * R donde R=3.5, RIC =Q3-Q1	control='control-6', estado=Súper atípico mensual', valor_procesado=valor	288	0,046%
Datos limpios	-	-	control='NA',estado='ok', valor_procesado=valor	614522	98,603%

Fuente: Elaboración propia

**Tabla 8: controles de calidad a datos diarios de humedad relativa promedio.**

CONTROLES DIARIOS HUMEDAD RELATIVA PROMEDIO					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Control 1 (Revisión de rangos)	Error	$hr < 0\%$ o $hr > 100\%$	control='control-1', estado='Error', valor_procesado=99999	2	3,64e-4%
Control 3 (valores atípicos respect al mes)	Atípico mensual	Dato > Cuartil3 + RIC * R Dato < Cuartil1 - RIC * R donde R=3, RIC=Q3-Q1	control='control-3', estado='Atípico mensual', valor_procesado=valor	249	0,045%
Control 4 (Identificación de consecutivos)	Error	Más de 7 días consecutivos	control='control-4', estado='Error', valor_procesado=99999	8	0,001%
	Sospechoso	Entre 4 - 7 días consecutivos	control='control-4', estado='Sospechoso', valor_procesado=valor	172	0,031%
Control 5 (Diferencias entre datos)	Error	Diferencias >55% de 2 Datos diarios consecutivos	control='control-5', estado='Error', valor_procesado=valor	0	0%
	Sospechoso	Diferencias entre 40 y 55% entre 2 diarios consecutivos	Control='control-5', estado='Sospechoso', valor_procesado=valor	48	0,009%
Control 6 (valores super atípicos del mes)	Súper atípico mensual	Dato > Cuartil3 + RIC * R Dato < Cuartil1 - Ric * R donde R=3.5, RIC =Q3-Q1	control='control-6', estado='Súper atípico mensual', valor_procesado=valor	263	0,048%
Datos limpios	-	-	control='NA', estado='ok', valor_procesado=valor	548911	99,865%

Fuente: Elaboración propia



• Conjunto de datos de brillo solar

**Descripción:** conjunto de datos de brillo solar para estaciones convencionales a escala horaria.

ción de dato faltante (99999) a la serie diaria de brillo solar.

I. Etapa de agrupación de datos a escala diaria.

Asignación del dato diario con la suma de los datos horarios del brillo solar de cada día. Adición del campo numero\_registros, con el número de datos horarios sumados para el dato del día.

**Paso 2:** identificación de duplicidades por fecha, el objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha u hora.

III. Etapa de Control de calidad para datos diarios de brillo solar

II. Etapa de revisión y preparación

**Paso 1:** agregación de registros de fecha-hora no existente y su respectivo valor con la conven-

A continuación, se muestran los controles de calidad aplicados y la cantidad de datos marcados por cada uno, para brillo solar diarios en la **Tabla 9**.

**Tabla 9: controles de calidad a datos diarios de brillo solar.**

CONTROLES DIARIOS BRILLO SOLAR					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Control 1 (Revisión de rangos)	Error	bs<0 hr o bs>13 hr	control='control-1', estado='Error', valor_procesado=99999	51	0,001%
	Sospechoso	11 hr < bs > 13 hr	control='control-1', estado=sospechoso', valor_procesado=valor	5516	0,157%
Control 3 (Valores atípicos respecto al mes)	Atípico mensual	Dato > Cuartil3 + RIC * R Dato > Cuartil1 - RIC * R Donde R=3, RIC=Q3-Q1	control='control-3', estado='Atípico mensual', valor_procesado=valor	2324	0,066%
Contol 4 (Identificación de consecutivos)	Error	Más de 3 días consecutivos	control='control-4', estado='Error', valor_procesado=99999	14862	0,422%
	Sospechoso	3 días consecutivos	control='control-4', estado='Sospechoso', valor_procesado=valor	16329	0,464%
Control 5 (Diferencias entre datos)	Error	Diferencias > 13 hr de 2 datos diarios consecutivos	control='control-5', estado='Error', valor_procesado=99999	0	0%
	Sospechoso	Diferencias > 12 y 13 hr entre 2 diarios consecutivos.	control='control-5', estado='Sospechoso', valor_procesado=valor	1	2,8e-5 %
datos limpios	-	-	control='NA', estado='ok', valor_procesado=valor	3483111	98,89%

Fuente: Elaboración propia

• *Conjunto de datos de velocidad de viento*

**Descripción:** conjunto de datos de velocidad de viento (VV) para estaciones convencionales y automáticas a nivel subhorario (inferior a 1 hora) para todo Colombia.

*I. Etapa de Revisión y preparación*

**Paso 1:** clasificación de estaciones en el campo "frecuencia", de acuerdo con la escala temporal:

- Cincominutal (5mins)
- Diezminutal (10mins)
- Quinceminutal (15mins)
- Horaria (60mins)

**Paso 2:** agregación de registros de fecha-hora no existente y su respectivo valor con la convención de dato faltante (99999).

**Paso 3:** identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha-hora.

*II. Etapa de Control de calidad*

A continuación, se muestran los controles de calidad aplicados y la cantidad de datos marcados para velocidad de vientos.

**Tabla 10: controles de calidad a datos subhorarios de velocidad de viento.**

CONTROLES MINUTALES VELOCIDAD DE VIENTO					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Identificación de rango	Error	$VV < 0 \text{ m/s}$ o $VV > 33 \text{ m/s}$	control='control-rangos', estado='Error', valor_procesado=99999	190702	0,226%
	Sospechoso	$VV > 17 \text{ m/s}$	control='control-rangos', estado='Sospechoso', valor_procesado=valor	49207	0,058%
Valores atípicos (respecto al mes)	Sospechoso	Dato > Cuartil3 * RIC * R Dato < Cuartil1 - RIC * R Donde R=5, RIC=Q3-Q1	control='control-atípicos', estado='Atípicos mensual', valor_procesado=valor	235132	0,279%
Identificación de consecutivos (sin tener en cuenta 0)	Error	3 o más datos (5, 10 o 15) minutales consecutivos	control='control-consecutivos', estado='Error', valor_procesado=valor 99999	3475964	4,126%
	Sospechoso	2 datos (5,10 o 15) minutales consecutivos	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor	11479850	13,627%
Diferencias entre datos	Sospechoso	Diferencias mayores a 10 m/s entre dos datos consecutivos	control='control-diferencias', estado='Sospechoso', valor_procesado=valor	56328	0,067%
Datos limpios	-	-	contro='NA', estado=ok, valor_procesado=valor	68753757	81,616%

III. Etapa de agrupación de  
datos a escala horaria y diaria

Luego de los controles realizados, se establecen las siguientes bases de datos:

1. Promedios horarios: se considera la cantidad de datos en cada hora a ser promediados.

**Tabla 11: cantidad de datos para la calcular datos de viento a escala horaria**

Escala minutal	Cantidad de datos	Promedio con mínimo
5	12	8
10	6	4
15	4	3

Fuente: Elaboración propia

2. Valor *máximo promedio* horario para cada día, con la condición de tener al menos 16 datos de los 24 posibles.
3. Para cada serie 5, 10 ó 15 minutal, Velocidad máxima del viento absoluto para cada día con su respectivo registro de fecha-hora.

**Tabla 12: cantidad de datos para calcular la velocidad máxima del viento absoluto**

Escala minutal	Cantidad de datos	Obtención de máximo absoluto con cantidad mínima de datos
5	288	202
10	144	101
15	96	67
60	24	17

Fuente: Elaboración propia

4. Promedio de velocidad del viento diario a partir de los promedios horarios, es decir del cálculo 1 anterior.
  - **Grupo 1:** > 70% ( $\geq 17$  datos)
  - **Grupo 2:** > 50% y < 69% (entre 12 y 16 datos)
  - **Grupo 3:** < 50% (< 11 datos)

Se generan grupos según representatividad, es decir según la cantidad de datos horarios.

• *Conjunto de radiación solar global*

**Descripción:** conjunto de datos de radiación solar global para estaciones convencionales a nivel diario y para las automáticas a nivel horario.

*1. Estaciones automáticas*

Fueron reportados 5.995.813 datos por 119 estaciones automáticas en el periodo de 2005/02/02 a 2020/10/01.

*I. Etapa de Revisión y preparación*

**Paso 1:** Agregación de registros de fecha-hora no existente y su respectivo valor con la convención de dato faltante (99999).

**Paso 2:** Identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha-hora.

*II. Etapa de Control de calidad*

A continuación, se presentan los controles de calidad aplicados y la cantidad de datos marcados para los datos de radiación solar global de estaciones automáticas.

**Tabla 13: controles de calidad a datos horarios de radiación solar global estaciones automáticas.**

CONTROLES HORARIOS RADIACIÓN SOLAR (Estaciones automáticos)					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Identificación de rango	Error	$RS < 0$ o $RS > 1361 \text{ Wh/m}^2$	control='control-rangos', estado='Error', valor_procesado=99999	185	0,003%
	Sospechoso	$1200 \leq RS \leq 1361 \text{ Wh/m}^2$	control='control-rangos', estado='Sospechoso', valor_procesado=valor	3	5 e-5%
Control noche	Error	Datos !=0 desde las 8pm hasta las 5 am	control='control-noche', estado='Error', valor_procesado=0	93913	1,566%
Datos limpios	-	-	control=NA, estado='ok', valor_procesado=valor	5901712	98,430%

Fuente: Elaboración propia

Finalmente, de los datos procesados se pudieron sacar algunas métricas:

- De los 185 datos que fueron identificados como error en el control de identificación de rango, 169 datos (91.35%) corresponden a la estación con código 23105070, con nombre MACEO en el municipio de Yolombó en Antioquia.
- De las 119 estaciones del conjunto de datos, tan sólo 4 reportaron en el año 2020.
- De las 119 estaciones, solo 16 reportan durante 10 años continuos o más y cuentan con al menos el 70% de los datos en este periodo.

**III. Etapa de agrupación de datos a escala diaria**

Se calculó el acumulado diario de radiación solar global a partir de los datos horarios con controles de calidad desde las 8AM hasta las 5PM.

Se generaron grupos de representatividad, según la cantidad de datos horarios.

- **Grupo 1:** > =70% (>= 7 datos)
- **Grupo 2:** < 70% (< 7 datos)

**2. Unión de datos de estaciones automáticas y convencionales**

Después de revisar duplicados en los valores de radiación solar convencional diarios, se procedió a la unión de estos datos con los de radiación solar automáticos diarios del grupo 1, para generar una sola base de datos de radiación solar diaria.

**3. Resumen**

**Tabla 14: transformación de los datos de radiación solar global.**

Datos en escala horaria	Aplicación controles de calidad	Horarios con controles de calidad	Agrupación a escala diaria	Datos en escala diaria	Unión de datos diarios	Datos consolidados
SIGRA_Clima_VPN. Radiacion_global_automáticas_vpn		SIGRA_Clima_VPN. Radiacion_global_automáticas_vpn_ controles		SIGRA_Clima_VPN. Radiacion_global_automáticas_vpn_ diarios		SIGRA_Clima_VPN. Radiacion_global_ consolidado_diarios
	NA			SIGRA_Clima_VPN. Radiacion_global_ convencionales_vpn		

Fuente: Elaboración propia

**ii. Homogenización de los datos**

Para completar los datos faltantes de las variables de humedad relativa, brillo solar y velocidad del viento, se realizó una homogenización usando la herramienta Climatol<sup>6</sup>, por medio de un script construido en R. Inicialmente se realizó una homogenización exploratoria para revisar la disponibilidad de los datos y obtener los valores del parámetro Standard Normal Homogeneity Test (SNHT), los cuales se usaron posteriormente para la homogenización completa. Las salidas de este proceso incluyeron un archivo en PDF con unas gráficas que permitieron el

análisis de los datos de la variable homogenizada, así como, un archivo TXT con los logs del proceso, y otros archivos de Excel en donde se encuentran:

- Los cortes de las series procesadas
- Un archivo que incluye para cada registro, una bandera que indica si el dato es correcto, fue generado o fue un dato faltante.
- Archivo de series en donde se encuentran los datos completados.

6 <https://www.climatol.eu/>

A continuación, se detallan los resultados obtenidos de la homogenización para cada variable climática:

**Tabla 15: resultados de estaciones después de la homogenización**

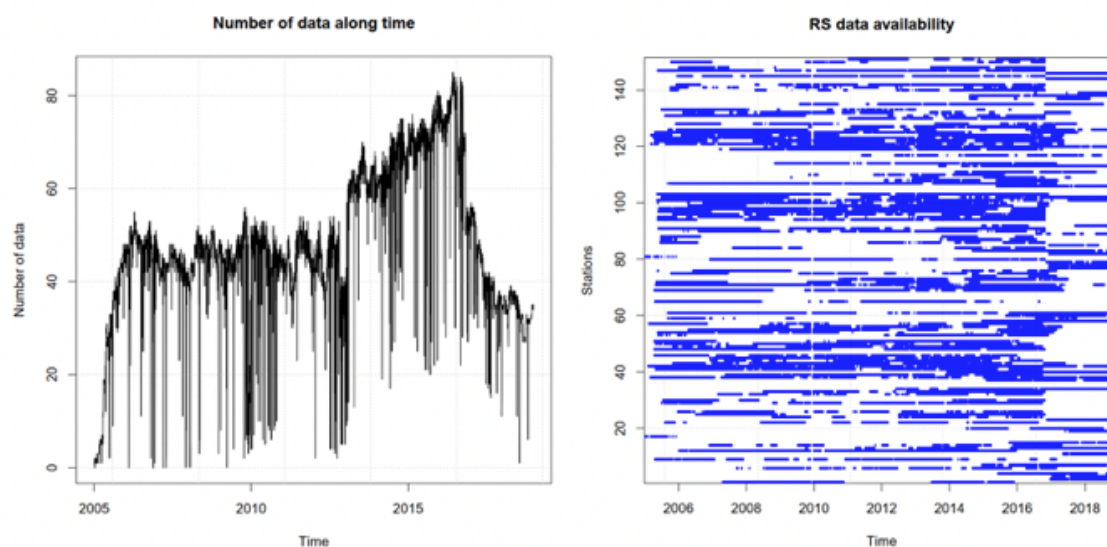
Variable	Periodo	Número de estaciones
Brillo solar	1980-01-01 a 2019-12-31	448
Humedad relativa promedio	1985-01-01 a 2019-12-31	221
Velocidad de viento promedio	2008-01-01 a 2020-12-31	293
Velocidad de viento máxima absoluta	2008-01-01 a 2020-12-31	293
Temperatura máxima	1980-01-01 a 2019-12-31	642
Temperatura mínima	1980-01-01 a 2019-12-31	641

Fuente: Elaboración propia

Por otra parte, la variable de radiación solar presentó dificultades durante el proceso de homogenización debido a que no cumplió con la cantidad de datos mínima de estaciones para realizar una homogenización completa, como se puede observar en la siguiente ilustración, la tabla de la izquierda

presenta líneas negras verticales indicando datos vacíos en la serie de tiempo. Mientras que la gráfica de la derecha presenta la cantidad de estaciones en cada año, en ella se puede identificar una gran cantidad de espacios blancos que representan la ausencia de información a lo largo de toda la serie.

**Ilustración 1: resultados de *Climatol* para el conjunto de datos de Radiación Solar**



Fuente: Elaboración propia

### 3.1.1.2 Series climáticas históricas del portal datos abiertos

Para el cálculo de la amenaza por vientos fuertes fue necesario complementarlos con información proveniente de Datos Abiertos en escalas subhorarias (cincominutal 5mins, diezminutal 10mins y quinceminutal 15mins) descargados del link: <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Velocidad-Viento/sgfv-3yp8>. Los datos de vientos fuertes en escalas de tiempos cortos son importantes para aplicaciones agroclimáticas y trabajos de seguros agropecuarios, ya que están relacionados con eventuales daños físicos en cultivos como maíz, arroz y banano, entre otros.

A estos datos, se aplican los siguientes procesos:

#### **A. Alistamiento de datos**

##### **i. Normalización: reconocimiento y organización de la información**

Los datos se descargan mediante un script de Python que consume el servicio de datos abiertos para diferentes periodos de fechas. De la información entregada se toman sólo los campos de interés, definiendo una estructura de campos. Luego, se procede a unir los datos de los diferentes rangos de fechas consultados, se eliminan filas duplicadas, se validan formatos de los campos y se cargan los datos a la herramienta BigQuery. Con esta información se realiza un cruce con el catálogo de las estaciones para vincular la información correspondiente, como ubicación (coordenadas), estado de la estación y demás información complementaria.

##### **ii. Estandarización**

Para aplicar los controles de calidad a los datos diarios del portal de datos abiertos, se agregan nuevos campos a las tablas en Big Query. Se agrega el campo 'control' que tiene un valor inicial de 'NA'; luego de ejecutar los controles de calidad se identifican los datos que fueron marcados en cada control, mediante los valores 'control-rangos',

'control-consecutivos', según corresponda. El segundo campo agregado es el 'estado' que tiene un valor inicial 'ok' para todos los registros; luego de aplicar los controles, se conserva el valor 'ok' si el dato pasa todos los controles de calidad, o toma los valores de 'Error' y 'Sospechoso', según se identifique el dato en los controles. El tercer campo es el 'valor\_procesado' que tiene un valor inicial igual al campo 'valor' de la tabla; luego de la ejecución de los controles, toma el valor de 99999 cuando el dato se identifica como 'Error' en algún control; en caso contrario, mantiene el mismo valor inicial.

#### **B. Procesamiento**

##### **i. Control de calidad de los datos de velocidad de viento**

Las series de datos históricas descargadas del portal de datos abiertos están disponibles desde el año 2005 y en escala subhoraria.

El proceso de calidad aplicado para los datos de velocidad de viento es el siguiente:

##### **I. Etapa de Revisión y preparación**

**Paso 1:** identificación de los conjuntos de estaciones de acuerdo a su escala temporal: 1. Cincominutal (5 min), 2. Diezminutal (10 min) y 3. Quinceminutal (15 min).

**Paso 2:** agregación de registros de fecha-hora no existentes y su respectivo valor con la convención de dato faltante (99999).

**Paso 3:** identificación de duplicidades por fecha-hora. El objetivo es tener la fecha-hora de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha-hora.

##### **II. Etapa de Control de calidad**

La siguiente tabla presenta los controles de calidad realizados a los datos subhorarios de velocidad de viento:

**Tabla 16: controles de calidad aplicados a los datos históricos de velocidad de viento del portal de datos abiertos.**

CONTROLES MINUTALES VELOCIDAD DE VIENTO					
Nombre	Estado	Condición	Marcación en Big Query	# datos	%
Identificación de rangos	Error	$VV < 0 \text{ m/s}$ o $VV > 33 \text{ m/s}$	control='control-rangos', estado='Error', valor_procesado=99999	165707	0,162%
	Sospechoso	$VV > 17 \text{ m/s}$	control='control-rangos', estado='Sospechoso', valor_procesado=valor	143445	0,140%
Valores atípicos (respecto al mes)	Sospechoso	Dato > Cuartil3 * RIC * R Dato < Cuartil1 - RIC * R Donde R=5, RIC=Q3-Q1	control='control-atípicos', estado='Atípicos mensual', valor_procesado=valor	366790	0,358%
Identificación de consecutivos (sin tener en cuenta 0)	Error	3 o más datos (5, 10 o 15) minutales consecutivos	control='control-consecutivos', estado='Error', valor_procesado=valor 99999	4340621	4,241%
	Sospechoso	2 datos (5, 10 o 15) minutales consecutivos	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor	11948603	11,675%
Diferencias entre datos	Sospechoso	Diferencias mayores a 10 m/s entre dos datos consecutivos	control='control-diferencias', estado='Sospechoso', valor_procesado=valor	69598	0,068%
Datos limpios	-	-	contro='NA', estado=ok, valor_procesado=valor	85304989	83,355%

Fuente: Elaboración propia

### III. Etapa de Agrupación de datos a escala horaria y diaria

Luego de los controles realizados, se establecen cuatro (4) bases de datos, así:

1. Obtención de los promedios horarios. Se considera la cantidad de datos mínimos en cada hora a ser promediados, así:



**Tabla 17: cantidad de datos para la calcular datos de viento a escala horaria**

Escala minutal	Cantidad de datos	Mínimo de datos para promedio
5	12	8
10	6	4
15	4	3

Fuente: Elaboración propia

- Obtención del valor máximo promedio horario para cada día, con la condición de tener al menos 16 datos de los 24 posibles.
- Para cada serie 5, 10 o 15 minutos se obtuvo el valor de Velocidad máxima del viento absoluto para cada día, con su respectivo registro de fecha-hora.

**Tabla 18: cantidad de datos para calcular la velocidad máxima del viento absoluto**

Escala minutal	Cantidad de datos	Cantidad mínima de datos para obtención de máximo absoluto
5	288	202
10	144	101
15	96	67
60	24	17

Fuente: Elaboración propia

- Se calculó el promedio de velocidad del viento diario a partir de los promedios horarios, es decir del cálculo 1 anterior.  
Se generaron grupos de representatividad, es decir la cantidad de datos horarios.
  - Grupo 1:** > 70% ( $\geq 17$  datos)
  - Grupo 2:** > 50% y < 69% (entre 12 y 16 datos)
  - Grupo 3:** < 50% (< 11 datos)

## ii. Homogenización de los datos

Para completar los datos faltantes de velocidad del viento promedios diarios, se realizó una homogenización usando la herramienta Climatol, por medio de un script construido en R. Inicialmente

se realiza una homogenización exploratoria para revisar la disponibilidad de los datos y obtener los valores del parámetro Standard Normal Homogeneity Test (SNHT), los cuales se usan posteriormente para la homogenización completa. Las salidas de este proceso incluyen un archivo en PDF con unas gráficas que permiten el análisis de los datos de la variable homogenizada, así como, un archivo TXT con los logs del proceso, y otros archivos de Excel en donde se encuentran:

- Los cortes de las series procesadas
- Un archivo que incluye para cada registro, una bandera que indica si el dato es correcto, fue generado o fue un dato faltante.
- Archivo de series en donde se encuentran los datos completados.

A continuación, se detalla la cantidad de estaciones resultantes y el periodo de tiempo, para los datos luego de la homogenización:

**Tabla 19: resultados de estaciones después de la homogenización de datos históricos de datos abiertos.**

Variable	Periodo	Número de estaciones
Velocidad de viento promedio diario	2010-01-01 a 2020-12-31	476

Fuente: Elaboración propia

### 3.1.1.3 Series históricas de precipitación de CHIRPS

La potencialidad en la implementación de los datos satelitales CHIRPS en el contexto histórico del SIGRA, consiste en el uso del producto global en escala temporal mensual, pentadal, decadal, disponibles desde enero de 1981 hasta el mes de noviembre de 2020. El uso principal del producto CHIRPS en el contexto histórico es para el cálculo de amenazas por excesos y/o deficiencias hídricas a escala mensual. El proceso propuesto para el alistamiento de los datos CHIRPS para el contexto histórico de SIGRA se describe a continuación:

#### A. Alistamiento de datos

##### i. Obtención de datos CHIRPS<sup>7</sup>:

Se realiza la descarga de los datos CHIRPS mensuales en formato ráster (\*.tif), desde enero de 1981 hasta el último mes disponible, por ejemplo, noviembre de 2020, que se encuentran publicados en el enlace: [https://data.chc.ucsb.edu/products/CHIRPS-2.0/global\\_monthly/](https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_monthly/). Esta descarga se realiza por medio de un script en R, que los proyecta al sistema de coordenadas WGS84.

##### ii. Información de precipitación CHIRPS históricas por pixel:

Este proceso de extracción se implementa a través de un script en R que organiza en una matriz las coordenadas y las fechas asociadas a cada dato para obtener la precipitación en cada pixel. Esta información se almacena pixel a pixel en una matriz y posteriormente se descarga en formato CSV, en el repositorio Cloudstorage.

Adicionalmente, se realiza la validación de posibles errores sistemáticos que pueden presentarse en las capas ráster para los datos de precipitación. Esta verificación consiste en asegurar que la zona de cobertura de la capa corresponda a Colombia, así como asegurar que el sistema de coordenadas se encuentra en WGS84. También se realiza la identificación de la resolución actual, la cual es de 5 km x 5 km, que se ajusta una vez se define el tamaño de la grilla óptima. Finalmente, se confirma que la corrección de datos, de estaciones nacionales, ya se ha realizado por la fuente de datos antes de la publicación de la capa, razón por la cual no se realiza en este caso dicha corrección en el proceso de alistamiento de datos.

#### B. Controles de calidad

Para las series históricas de precipitación de CHIRPS no fue necesario realizar procesos de tratamiento de la información adicionales, controles de calidad o procesos de homogenización, ya que la fuente oficial de estos datos es el Climate Hazard Center de la Universidad de California en Santa Bárbara, que ya aplica robustos procesos de calidad previos a su publicación. Estos controles se encuentran descritos en el enlace <https://chc.ucsb.edu/data/chirps/diagnostics>.

### 3.1.1.4 Datos de elevación de la superficie de la tierra Shuttle Radar Topography Mission (SRTM)<sup>8</sup>

La misión SRTM provee modelos digitales de elevación (DEM) con cobertura para Colombia con diferentes niveles de procesamiento y resoluciones espaciales. A continuación, se describen los procesos que se realizan a los datos:

7 [www.chc.ucsb.edu/data/chirp](https://data.chc.ucsb.edu/data/chirp)

8 <https://www2.jpl.nasa.gov/srtm/>

**A. Alistamiento de datos**

*i. Descarga del modelo digital de elevación*

Se realiza la descarga de los datos de SRTM en formato ráster (\*.tif), que se encuentran publicados en el enlace: <https://srtm.csi.cgiar.org/srtmdata/> . Esta descarga se realiza de forma manual teniendo en cuenta que sólo se requiere la descarga por una única vez.

*ii. Recorte a las zonas de interés*

Posterior a la descarga del modelo de elevación, se procede a recortarlo para dejar sólo las zonas de interés para el SIGRA, empezando por delimitar la zona de Colombia, proceso que se lleva a cabo mediante la herramienta QGis.

*iii. Re-proyección a la resolución requerida*

Finalmente, haciendo uso de la herramienta QGis se procede a re proyectar al sistema de coordenadas WGS84, que es el sistema de coordenadas con el que se trabajan los datos espacializados, en el contexto del proyecto SIGRA.

**B. Controles de calidad**

Para los datos de elevación de la superficie de la tierra no fue necesario realizar procesos de tratamiento de la información adicionales, controles de calidad o procesos de homogenización, ya que la fuente oficial de estos datos es el Grupo Consultivo para la Investigación Agrícola Internacional (CGIAR por sus siglas en inglés), que ya aplica robustos procesos de calidad previos a su publicación. Estos controles se encuentran descritos en el enlace <https://srtm.csi.cgiar.org/>.

**3.1.2 Contexto de monitoreo**

Las fuentes de información priorizadas para el monitoreo de amenazas climáticas consisten principalmente en datos meteorológicos de actualización diaria en el portal de datos abiertos para las variables precipitación, temperatura, humedad relativa, velocidad del viento, así como datos de precipitación y temperatura de la Oficina del Servicio de Pronósticos y Alertas del IDEAM (OSPA) y datos de lluvia provenientes de satélites (Hidroestimador y CHIRPS). La siguiente tabla resume los datos priorizados para el contexto de monitoreo de amenazas agroclimáticas:

**Tabla 20: resumen de datos priorizados para el cálculo de riesgos en el contexto de monitoreo.**

Fuentes de datos priorizadas	Temporalidad	Frecuencia	Fuente
Información meteorológica diaria del IDEAM en el portal de datos abiertos	Precipitación, humedad del aire, velocidad del viento, temperatura máxima y mínima: <b>subhoraria.</b>	Actualización diaria.	IDEAM
Oficina de Pronostico y Alertas de IDEAM (OSPA)	Precipitación, temperatura: <b>diaria.</b>	Actualización diaria.	IDEAM
Datos satelitales CHIRPS	Precipitación CHIRPS: <b>mensual / decadal / pentadal y diarios.</b>	Actualización mensual.	Climate Hazards Center - UC Santa Bárbara.
Datos de lluvia satelital del hidroestimador	Precipitación <b>diaria</b>	Actualización diaria.	NOAA GOES (STAR Satellite Rainfall Estimates)

Fuente: Elaboración propia

### 3.1.2.1 Información diaria del portal de datos abiertos

Como se indicó antes, con la iniciativa de datos abiertos se busca que todas las entidades gubernamentales publiquen su información en formatos estructurados, para ponerla a disposición de los usuarios con criterios de pertinencia y calidad.

El proceso para la utilización de las series de datos meteorológicos publicados diariamente en el portal de datos abiertos se compone de 4 pasos, que se describen a continuación:

#### A. Alistamiento de datos:

##### i. Normalización: reconocimiento y organización de la información

La descarga de la información por variable se realiza directamente en el portal datos abiertos<sup>9</sup> para las variables: precipitación, humedad del aire, velocidad del viento y temperatura máxima y mínima. Los enlaces específicos de descarga por variable se muestran en la **Tabla 21**.

**Tabla 21: enlaces de descarga de las variables meteorológicas a través de datos abiertos.**

No.	Variable	Nombre de Conjunto de datos	Actualización	Serie histórica	Descripción del conjunto de datos
1	Precipitación	Precipitación	Diaria	Si, dependiendo de cada estación registrada	<a href="https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg">https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg</a> Documento API: <a href="https://dev.socrata.com/foundry/www.datos.gov.co/s54a-sgyg">https://dev.socrata.com/foundry/www.datos.gov.co/s54a-sgyg</a>
2	Precipitación	Precipitaciones 2017-2020	Diaria	2017 a 2020 para estaciones meteorológicas en Bogotá	<a href="https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaciones-2017-20/uwim-sqxg">https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaciones-2017-20/uwim-sqxg</a> Documento API: <a href="https://dev.socrata.com/foundry/www.datos.gov.co/uwim-sqxg">https://dev.socrata.com/foundry/www.datos.gov.co/uwim-sqxg</a>
3	Temperatura mínima	Temperatura mínima del aire a 2m	Diaria	Si, dependiendo de cada estación registrada	<a href="https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Temperatura-M-nima-del-Aire/afdg-3zpb">https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Temperatura-M-nima-del-Aire/afdg-3zpb</a> Documento API: <a href="https://dev.socrata.com/foundry/www.datos.gov.co/afdg-3zpb">https://dev.socrata.com/foundry/www.datos.gov.co/afdg-3zpb</a>
4	Temperatura máxima	Temperatura máxima del aire a 2m	Diaria	Si, dependiendo de cada estación registrada	<a href="https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Temperatura-M-xima-del-Aire/ccvq-rp9s">https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Temperatura-M-xima-del-Aire/ccvq-rp9s</a> Documento API: <a href="https://dev.socrata.com/foundry/www.datos.gov.co/ccvq-rp9s">https://dev.socrata.com/foundry/www.datos.gov.co/ccvq-rp9s</a>

<sup>9</sup> <https://www.datos.gov.co>

No.	Variable	Nombre de Conjunto de datos	Actualización	Serie histórica	Descripción del conjunto de datos
5	Velocidad del viento	Velocidad del viento cada 10 minutos.	Diaria	Si, dependiendo de cada estación registrada	<a href="https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Velocidad-Viento/sgfv-3yp8">https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Velocidad-Viento/sgfv-3yp8</a> Documento API: <a href="https://dev.socrata.com/foundry/www.datos.gov.co/sgfv-3yp8">https://dev.socrata.com/foundry/www.datos.gov.co/sgfv-3yp8</a>
6	Humedad del aire	Humedad del aire a 2m horaria	Diaria	Si, dependiendo de cada estación registrada	<a href="https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Humedad-del-Aire-2-metros/uext-mhny">https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Humedad-del-Aire-2-metros/uext-mhny</a> Documento API: <a href="https://dev.socrata.com/foundry/www.datos.gov.co/uext-mhny">https://dev.socrata.com/foundry/www.datos.gov.co/uext-mhny</a>

Fuente: Elaboración propia

Los datos se descargaron mediante un script de Python que consume el servicio de datos abiertos para diferentes periodos. De la información descargada se toman sólo los campos de interés, definiendo una estructura de campos. Luego, se procede a unir los datos de los diferentes rangos de fechas consultados, se eliminan filas duplicadas, se validan formatos de los campos y se cargan los datos a la herramienta BigQuery. Con esta información, se realiza un cruce con el catálogo de las estaciones para vincular la información correspondiente, como ubicación (coordenadas), estado de la estación y demás información.

## ii. Estandarización

Para aplicar los controles de calidad a los datos diarios de monitoreo de datos abiertos, se agregan nuevos campos a las tablas en BigQuery. Se agrega el campo 'control' que tiene un valor inicial de 'NA'; luego de ejecutar los controles de calidad se identifican los datos que fueron marcados en cada control, mediante los valores 'control-rangos', 'control-consecutivos', según corresponda. El segundo campo agregado es el 'estado' que tiene un valor inicial de 'ok' para todos los registros; luego de aplicar los controles, se conserva el valor de 'ok' si el dato pasó todos los controles de calidad, o toma los valores de 'Error' y 'Sospechoso', según como haya sido identificado el dato en los controles. El tercer campo es el 'valor procesado' que tiene un valor inicial igual al campo 'valor' de la tabla; luego de la ejecución

de los controles, toma el valor de 99999 cuando el dato es identificado como 'Error' en algún control, en dado caso que no, mantiene el mismo valor inicial.

## B. Procesamiento

### i. Controles de calidad aplicados a las variables meteorológicas

El control de calidad de los datos descargados del portal de datos abiertos consiste en la verificación de rangos, la identificación de datos atípicos y consecutivos, así como el control de diferencias entre datos consecutivos según la temporalidad de las variables. El propósito de este control es evitar que ingresen datos erróneos y sean almacenados en la base de datos del SIGRA para el módulo de monitoreo. Los controles propuestos para los datos descargados del portal de datos abiertos son los siguientes:

#### • Conjunto de datos de precipitación

**Descripción:** conjunto de datos de precipitación para estaciones automáticas disponibles en Colombia a nivel cincominutal y diezminutal. El conjunto de datos puede ser consultado en el siguiente enlace: <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitacion/s54a-sgyg>

El proceso propuesto para la generación de acumulados diarios de precipitación es el siguiente:

*I. Etapa de Revisión y preparación*

**Paso 1:** identificar los conjuntos de estaciones de acuerdo con la escala temporal:

Conjunto 1. Cincominutal (5 min),

Conjunto 2. Diezminutal (10 min)

**Paso 2:** introducir registros de fecha-hora no existente. Identificación de valores faltantes.

El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal.

**Paso 3:** identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha u hora.

*II. Etapa de agrupación de datos a escala horaria.*

*Generación de Base de Datos de precipitación*

*Horaria (BDPH)*

Los datos de precipitación provenientes de las estaciones automáticas, deben ser agrupados en escala horaria. Para ello, se deben acumular los datos para cada hora, teniendo en cuenta los grupos de datos mínimos mostrados en la **Tabla 22:**

**Tabla 22:** cantidad de datos para acumular registro de precipitación a escala horaria

Escala minutal	Cantidad de datos	Acumulado con mínimo	Bandera de calidad (representatividad)
5	12	8	Grupo 1 > 8 datos Grupo 2: Entre 1 y 7 datos
10	6	4	Grupo 1 > 4 datos Grupo 2: Entre 1 y 3 datos

Fuente: Elaboración propia

*III. Etapa de agrupación de datos a escala diaria.*

*Generación de Base de Datos de Precipitación*

*Diaria (BDPD)*

Luego de la generación de la BDPH y la revisión de las series de tiempo en registros inexistentes y duplicados,

se construye una base de datos de precipitación acumulada diaria (BDTD), considerando la cantidad de datos horarios a ser acumulados, como se muestra en la siguiente tabla:

**Tabla 23:** criterios para acumular registros de precipitación a escala diaria

Cantidad de datos horarios a acumular	Grupo de representatividad
>=22	Grupo 1 (>= 90%)
12 y 21	Grupo 2 (>= 50% y < 89%)
<=11	Grupo 3 (<49%)

Fuente: Elaboración propia

*IV. tapa de Control de calidad para datos de Precipitación en escala diaria*

La **Tabla 24** presenta los controles realizados a cada serie climática:

**Tabla 24: controles de calidad propuestos para ser aplicados a los datos de precipitación provenientes del portal de datos abiertos y agregados a escala diaria.**

Nombre	Tipificación	Condición 1	Condición 2	Marcación en Big Query
Identificación de rangos	Error	ppt < 0 mm o ppt > 350 mm	N/A	control='control-rangos', estado='Error', valor_procesado=99999
	Sospechoso	200 mm < ppt ≤ 350 mm	N/A	control='control-rangos', estado='Sospechoso', valor_procesado=valor
Identificación de consecutivos	Error	4 o más días consecutivos.	Sin tener en cuenta valor 0. El primer día de los 4 se marca como sospechoso y el resto como error.	control='control-consecutivos', estado='Error' (excepto el 1er que se marca como 'Sospechoso'), valor_procesado=99999
	Sospechoso	2 y 3 días consecutivos.	Sin tener en cuenta valor 0.	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor
Diferencias entre días consecutivos.	Sospechoso	Diferencia >150 mm entre 2 días consecutivos.	Se marca como sospechoso el día actual.	control='control-diferencias', estado='Sospechoso', valor_procesado=valor

Fuente: Elaboración propia

Se aplican 5 controles de calidad para los datos diarios, lo que permite obtener datos diarios de calidad para los productos de monitoreo para excesos y deficiencias hídricas. Es de resaltar, que el control de datos atípicos para datos diarios de precipitación debe ser explorado con mayor profundidad, debido a que la distribución de dichos datos diarios no es ajustable a una curva normal, por lo cual su análisis, requiere otras metodologías o un ajuste previo. Por otro lado, el control de diferencias de días consecutivos no es muy usado para la variable precipitación, debido a la alta variabilidad entre dos días consecutivos, sin

embargo, se plantea inicialmente un valor de 150 mm, que puede ser ajustado para cada región de monitoreo, estudiando previamente las diferencias históricas entre dos consecutivos.

*V. Resultado para precipitación en Datos Abiertos*

Después de los procesos descritos anteriormente para la variable de precipitación, finalmente se dispone de 476 estaciones. A continuación, se describe la cantidad de estaciones para cada rango de disponibilidad de datos en el periodo de análisis 2021-02-01 a 2021-05-25.

**Tabla 25:** cantidad de estaciones para cada grupo de disponibilidad de datos de precipitación.

Disponibilidad de datos	Cantidad Estaciones
0 – 50%	186
50 – 70%	66
70 – 100%	224

Fuente: Elaboración propia

• **Conjunto de datos de temperatura máxima y mínima:**

Descripción: conjunto de datos de Temperatura para estaciones automáticas a nivel subhorario para toda Colombia. Los conjuntos de datos pueden ser consultados en los siguientes enlaces:

► <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Temperatura-Minima-del-Aire/afdg-3zpb>

<https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Temperatura-Maxima-del-Aire/ccvq-rp9s>

I. Etapa de agrupación de datos a escala horaria.  
Generación de Base de Datos de Temperatura Horaria (BDTH)

Los datos de temperatura máxima y mínima provienen de estaciones automáticas y por diversas razones, pueden existir 1, 2 o más datos para una misma hora, por lo cual se debe obtener un único valor para cada hora, por medio del valor más alto de los reportados (2 datos o más) para la temperatura máxima y por medio del valor más bajo de los reportados (2 datos o más) para la temperatura mínima o bien el dato

reportado en la hora. Con lo anterior, se construyó una Base de Datos de Temperatura Máxima Horaria (BDTmaxH) y otra base de datos de Temperatura Mínima Horaria (BDTminH).

II. Revisión y preparación

**Paso 1:** introducir registros de fecha-hora no existente a la BDTmaxH y a BDTminH, es decir tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal. Luego, se identifican los valores faltantes que se denotan con el código 99999.

**Paso 2:** identificación de duplicidades por fecha, es decir, tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha-hora.

III. Etapa de agrupación de datos a escala diaria.  
Generación de Base de Datos de Temperatura Mínima, Máxima y promedio Diario (BDTD)

Luego de la generación de la BDTmaxH y BDTminH, la revisión de las series de tiempo en registros no existentes y duplicados, se debe armar una Base de Datos de Temperatura Máxima Diaria (BDTmaxD) y Temperatura Mínima Diaria (BDTminD), considerando la cantidad de datos horarios, como se muestra en la siguiente tabla:



**Tabla 26: criterio para seleccionar los grupos de representatividad para la temperatura diaria.**

Cantidad de datos horarios de temperatura	Grupo de representatividad
≥16	Grupo 1 (≥70%)
entre 12 y 16	Grupo 2 (≥50% y < 70%)
≤11	Grupo 3 (<50)

Fuente: Elaboración propia

#### IV. Etapa de Control de calidad para datos de temperatura

La siguiente tabla presenta los controles de calidad a realizar a los datos diarios del grupo 1 de BDTmaxD y BDTminD :

**Tabla 27: controles de calidad propuestos para los datos de monitoreo de temperatura máxima y mínima del portal de datos abiertos.**

Nombre	Tipificación	Condición 1	Condición 2	Marcaación en Big Query
Identificación de rangos	Error	Temp > 50°C o Temp < -10°C	N/A	control='control-rangos', estado='Error', valor_procesado=99999
Comparaciones	Error	Tmin > Tmax	N/A	control='control-comparaciones', estado='Error', valor_procesado=99999
Valores atípicos	Sospechoso (Atípico)	Dato > Cuartil3 + RIC * R // Dato < Cuartil1 - RIC * R, donde R=4	RIC = Q3 - Q1	control='control-atípicos', estado='Sospechoso', valor_procesado=valor
Identificación de consecutivos	Error	4 o más datos iguales consecutivos	El primer día de los 4 se marca como sospechoso y el resto como error.	control='control-consecutivos', estado='Error' (excepto el 1er que se marca como 'Sospechoso'), valor_procesado=99999
Identificación de consecutivos	Sospechoso	3 datos iguales consecutivos	Todos los 3 días son sospechosos	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor
Diferencias entre datos	Error	Diferencias mayores a 16,0°C entre dos datos consecutivos	N/A	control='control-diferencias', estado='Error', valor_procesado=99999
	Sospechoso	Diferencias entre 7,0°C a 15,9°C entre dos datos consecutivos	N/A	control='control-diferencias', estado='Sospechoso', valor_procesado=valor

Fuente: Elaboración propia

*V. Resultado temperatura en Datos Abiertos*

Después de los procesos descritos anteriormente, para la variable de temperatura máxima finalmente se

tienen 220 estaciones y para temperatura mínima 219. A continuación, se describe la cantidad de estaciones para cada rango de disponibilidad de datos en el periodo de análisis 2021-02-01 a 2021-05-25.

**Tabla 28: cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura máxima.**

Disponibilidad de datos	Cantidad Estaciones
0 – 50%	103
50 – 70%	38
70 – 100%	79

Fuente: Elaboración propia

**Tabla 29: cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura mínima.**

Disponibilidad de datos	Cantidad Estaciones
0 – 50%	103
50 – 70%	42
70 – 100%	74

Fuente: Elaboración propia

• *Conjunto de datos de viento:*

Descripción: el conjunto presenta datos de velocidad de viento (VV) de 323 estaciones automáticas a nivel cincominutal (5 min), diezminutal (10 min) y quinceminutal (15 min) para toda Colombia. El conjunto de datos puede ser consultado en el siguiente enlace: <https://www.datos.gov.co/resource/sgfv-3yp8.CSV>

El proceso de calidad propuesto para los datos de monitoreo en escala horaria y diaria para velocidad de viento es el siguiente:

*I. Etapa de Revisión y preparación*

**Paso 1:** identificar los conjuntos de estaciones de acuerdo a su escala temporal: 1. Cincominutal (5 min), 2. Diezminutal (10 min) y 3. Quinceminutal (15 min).

**Paso 2:** introducir registros de fecha-hora no existente. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, es decir, que no falte ninguna fecha-hora. El código de identificación para datos faltantes es 99999.

**Paso 3:** identificación de duplicidades por fecha-hora. El objetivo es tener la fecha-hora de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha-hora.

*II. Etapa de Control de calidad*

La siguiente tabla presenta los controles de calidad a realizar a los datos subhorarios para velocidad de viento:

**Tabla 30: controles de calidad propuestos para los datos de monitoreo de la variable velocidad de viento del portal de datos abiertos.**

Nombre	Tipificación	Condición 1	Condición 2	Marcación en Big Query
Identificación de rangos	Error	$VV < 0 \text{ m/s}$ o $VV > 33 \text{ m/s}$	NA	control='control-rangos', estado='Error', valor_procesado=99999
Identificación de rangos	Sospechoso	$VV > 17 \text{ m/s}$	NA	control='control-rangos', estado='Sospechoso', valor_procesado=valor
Valores atípicos	Sospechoso (atípico)	$\text{Dato} > \text{Cuartil3} + \text{RIC} * R$ // $\text{Dato} < \text{Cuartil1} - \text{RIC} * R$ , donde $R=5$	$\text{RIC} = Q3 - Q1$	control='control-atípicos', estado='Sospechoso', valor_procesado=valor
Identificación de consecutivos	Error	3 o más datos (5, 10 o 15) minutales consecutivos	Sin tener en cuenta valor 0. El primer dato es sospechoso y los otros 2 son errores	control='control-consecutivos', estado='Error'(excepto el 1er que se marca como 'Sospechoso'), valor_procesado=99999
	Sospechoso	2 datos (5, 10 o 15) minutales consecutivos	Sin tener en cuenta valor 0. Ambos datos son sospechosos.	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor
Diferencias entre datos	Sospechoso	Diferencias mayores a 10 m/s entre dos datos consecutivos	NA	control='control-diferencias', estado='Sospechoso', valor_procesado=valor

Fuente: Elaboración propia

### III. Etapa de agrupación de datos a escala horaria y diaria

Luego de los controles realizados, se armaron tres (3) bases de datos, así:

1. Obtención de promedios horarios. Se debe considerar la cantidad de datos en cada hora a ser promediados:

**Tabla 31: cantidad de datos para generar acumulados de viento a escala horaria.**

Escala minutal	Cantidad de datos	Promedio con mínimo
5	12	8
10	6	4
15	4	3

Fuente: Elaboración propia

2. Obtener el valor máximo promedio horario para cada día, con la condición de tener al menos 16 datos de los 24 posibles.
3. Para cada serie 5, 10 o 15 minutos obtener el valor de *Velocidad máxima del viento absoluto* para cada día, con su respectivo registro de fecha-hora.

**Tabla 32: cantidad de datos para generar la velocidad máxima absoluta del viento.**

Escala minutal	Cantidad de datos	Obtención de máximo absoluto con cantidad mínima de datos
5	288	202
10	144	101
15	96	67
60	24	17

Fuente: Elaboración propia

4. Calcular el promedio de velocidad del viento diario a partir de los promedios horarios, o sea del cálculo 1, teniendo en cuenta como mínimo una representatividad horaria de 70%.  
  
Generación de grupos de representatividad, es decir la cantidad de datos horarios.
  - **Grupo 1:** > 70% (>= 21 datos)
  - **Grupo 2:** > 50% y < 69% (entre 15 y 20 datos)
  - **Grupo 3:** < 50% (< 14 datos)
5. Calcular el promedio de velocidad del viento mensual a partir de los promedios diarios, es decir del cálculo 4, teniendo en cuenta, como mínimo una representatividad diaria de 70% (grupo 1 diarios).  
  
Generación de grupos de representatividad, es decir la cantidad de datos horarios.
  - **Grupo 1:** > 70% (>= 21 datos)
  - **Grupo 2:** > 50% y < 69% (entre 15 y 20 datos)
  - **Grupo 3:** < 50% (< 14 datos)

*IV. Resultado para velocidad de viento en Datos Abiertos*

Después de los procesos descritos anteriormente para la variable de velocidad de viento, a continuación, se describe la cantidad de estaciones para cada una de las bases de datos creadas.

**Tabla 33: cantidad de datos para generar la velocidad máxima absoluta del viento.**

Base de datos	Cantidad Estaciones
Promedios horarios	322
Valor máximo promedio horario	308
Velocidad máxima del viento absoluto	308
Promedios diarios	322
Promedios mensuales	280

Fuente: Elaboración propia

• Conjunto de datos de humedad relativa:

Descripción: el conjunto de datos de Humedad relativa del aire a 2m para estaciones automáticas a nivel subhorario y horario para Colombia. Los conjuntos de datos pueden ser consultado en los siguientes enlaces:

<https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Humedad-del-Aire-2-metros/uext-mhny>

I. Etapa de agrupación de datos a escala horaria. Generación de Base de Humedad Relativa Horaria (BDHRH)

Los datos de Humedad Relativa provienen de estaciones automáticas y son reportados en escala cincominutal (5mins) y horarios dependiendo de la estación, por lo cual se debe obtener un único valor para cada hora, por medio del cálculo del promedio de la hora o bien el dato reportado durante la misma. Se deben armar una Base de Datos de Humedad Relativa Horaria (BDHRH).

II. Revisión y preparación

**Paso 1:** introducir registros de fecha-hora no existente a la BDHRH. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, es decir, que no falte ninguna fecha-hora e identificando los valores faltantes con el código 99999.

**Paso 2:** identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha-hora.

III. Etapa de Agrupación de datos a escala diaria. Generación de Base de Datos de Humedad Relativa Diaria (BDHRD)

Luego de la generación de la BDHRH y la revisión de registros no existentes y duplicados, se debe armar una Base de Datos de Humedad Relativa Diaria (BDHRD), considerando la cantidad de datos horarios, así:

**Tabla 34: criterio para seleccionar los grupos de representatividad para la humedad diaria.**

Cantidad de datos horarios de temperatura	Grupo de representatividad
$\geq 16$	Grupo 1 ( $\geq 70\%$ )
entre 12 y 15	Grupo 2 ( $\geq 50\%$ y $< 70\%$ )
$\leq 11$	Grupo 3 ( $< 50\%$ )
Reporte a horas específicas (Nota 1)	Grupo 4

Fuente: Elaboración propia

En relación con el conjunto de datos horarios de humedad relativa, es necesario resaltar los siguientes aspectos que se deben tener en cuenta.

- Se pueden encontrar datos horarios de humedad relativa en horas específicas, normalmente en horas de la mañana (tipo 6, 7 y 8am), a comienzo de tarde (12, 1 ó 2pm) y datos a final de tarde (5, 6 ó 7pm). Para obtener el dato del día, se tomará el dato reportado al medio día, es decir alrededor de las 12, 1 ó 2pm.
- Cuando se tengan datos alrededor de la 1pm, se obtendrá el promedio con los datos vecinos (12m, 1pm y 2pm). (Prioridad a grupo 4).
- En caso de reporte único en otros momentos del día diferentes al medio día, se tomará el promedio como grupo 3.

*IV. Etapa de Control de calidad para datos de Humedad Relativa diaria*

La siguiente tabla presenta los controles a realizar a los datos de monitoreo diario para la Humedad relativa, para el caso se tendrán en cuenta los datos provenientes de los grupos 1 y 4:

**Tabla 35: controles de calidad propuestos para los datos de monitoreo de humedad relativa del portal de datos abiertos.**

Nombre	Tipificación	Condición 1	Condición 2	Marcación en Big Query
Identificación de rangos	Error	HR < 0 y HR > 100%	N/A	control='control-rangos', estado='Error', valor_procesado=99999
Valores atípicos	Sospechoso (Atípico)	Dato > Cuartil3 + RIC * R // Dato < Cuartil1 - RIC * R, donde R=3	RIC = Q3 -Q1	control='control-atípicos', estado='Sospechoso', valor_procesado=valor
Identificación de consecutivos	Error	8 o más datos iguales consecutivos	El primer día de los 4 se marca como sospechoso y el resto como error.	control='control-consecutivos', estado='Error' (excepto el 1er que se marca como 'Sospechoso'), valor_procesado=99999
Identificación de consecutivos	Sospechoso	Entre 4 y 7 días iguales consecutivos	Todos los 4 o 7 días son sospechosos	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor
Diferencias entre datos	Error	Diferencias mayores a 55% entre dos datos consecutivos	NA	control='control-diferencias', estado='Error', valor_procesado=99999
Diferencias entre datos	Sospechoso	Diferencias entre 40 y 55% entre dos datos consecutivos	N/A	control='control-diferencias', estado='Sospechoso', valor_procesado=valor

Fuente: Elaboración propia

*V. Resultado para humedad relativa en Datos Abiertos*

Después de los procesos descritos anteriormente para la variable de humedad relativa, finalmente se tienen 335 estaciones. A continuación, se describe

la cantidad de estaciones para cada rango de disponibilidad de datos en el periodo de análisis 2021-02-01 a 2021-05-25.

**Tabla 36: cantidad de estaciones para cada grupo de disponibilidad de datos de humedad relativa.**

Disponibilidad de datos	Cantidad Estaciones
0 – 50%	152
50 – 70%	96
70 – 100%	87

Fuente: Elaboración propia

### 3.1.2.2 Información diaria del portal de OSPA

La oficina de servicios de pronósticos y alertas OSPA del IDEAM, consolida diariamente los datos meteorológicos de precipitación y temperatura, provenientes de la red nacional de estaciones meteorológicas reportados en escala diaria desde sus oficinas regionales. Estos datos, usados para el seguimiento del estado del tiempo y para los modelos de pronóstico y alerta meteorológica, son considerados preliminares, por no ser sometidos a procesos complementarios de calidad y oficialización. En las mesas de trabajo para el desarrollo del SIGRA, se acordó que IDEAM compartiría diariamente estos datos, que constituirían la principal fuente de información meteorológica de superficie para el monitoreo meteorológico requerido para la estimación de riesgos agroclimáticos en el SIGRA.

Se implementaron protocolos básicos de control de calidad que permiten garantizar la calidad de los datos y detectar errores en rangos físicos y estadísticos de las variables, antes de la integración de estos al SIGRA. A continuación, se detalla el proceso de control de calidad realizado a los datos en referencia.

#### A. Alistamiento de datos

##### [i. Normalización: descarga y organización de la información](#)

La descarga de los datos, específicamente aquellos de monitoreo diario para temperatura y precipitación, se hace en formato .xls, de la siguiente fuente: [http://181.225.72.46:45128/almacen/interno/estaciones/reportes\\_upra/](http://181.225.72.46:45128/almacen/interno/estaciones/reportes_upra/). Por ende, fue necesario desarrollar una macro que permitiera estructurar los datos en columnas, debido a que el formato dispuesto es una matriz que dificulta su procesamiento.

Al programar una macro que organiza los registros por filas y no por celdas, fue posible exportarlos en formato CSV para realizar el proceso anteriormente descrito. Luego, mediante un script en Python, se validaron los formatos de los campos del archivo CSV y posteriormente, se cargaron los datos a la herramienta BigQuery.

##### [ii. Estandarización](#)

Para aplicar los controles de calidad a los datos diarios de monitoreo OSPA, se definió agregar nuevos campos a las tablas en BigQuery de las variables precipitación, temperatura máxima y mínima. Se agregó el campo 'control' que tiene un valor inicial de 'NA'; luego de ejecutar los controles de calidad, se identificaron los datos que fueron marcados en cada control, mediante los valores 'control-rangos', 'control-consecutivos', según corresponda. El segundo campo agregado es el 'estado' que tiene un valor inicial de 'ok' para todos los registros; luego de aplicar los controles, se conservó el valor de 'ok' si el dato pasó todos los controles de calidad, o toma los valores de 'Error' y 'Sospechoso', según como haya sido identificado el dato en los controles. El tercer campo es el 'valor\_procesado' que tiene un valor inicial igual al campo 'valor' de la tabla; luego de la ejecución de los controles, toma el valor de 99999 cuando el dato es identificado como 'Error' en algún control, en caso contrario, se mantiene el valor inicial.

#### B. Procesamiento

##### [i. Control de calidad de los datos](#)

##### [• Conjunto de datos de precipitación:](#)

Descripción: conjunto de datos de Precipitación para estaciones a nivel diario, para toda Colombia.

*I. Revisión y preparación*

**Paso 1:** introducir registros de fecha no existentes, es decir asegurar que se cuente con la fecha de comienzo y fin de cada serie y de manera continua con su registro temporal. En las fechas agregadas, el dato de la variable toma el valor del código de identificación para datos faltantes que es el 99999.

**Paso 2:** identificación de duplicidades por fecha, es decir, tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha.

*II. Etapa de Control de calidad para datos de precipitación*

Teniendo los datos de precipitación en escala diaria, se procede a la aplicación de los controles de calidad, descritos en la siguiente tabla:

**Tabla 37: controles de calidad propuestos para los datos de precipitación de OSPA.**

Nombre	Tipificación	Condición 1	Condición 2	Marcación en Big Query
Identificación de rangos	Error	ppt < 0 mm o ppt > 350 mm	N/A	control='control-rangos', estado='Error', valor_procesado=99999
	Sospechoso	200 mm < ppt ≤ 350 mm	N/A	control='control-rangos', estado='Sospechoso', valor_procesado=valor
Identificación de consecutivos	Error	4 o más días consecutivos.	Sin tener en cuenta valor 0. El primer día de los 4 se marca como sospechoso y el resto como error.	control='control-consecutivos', estado='Error' (excepto el 1er que se marca como 'Sospechoso'), valor_procesado=99999
	Sospechoso	2 y 3 días consecutivos.	Sin tener en cuenta valor 0.	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor
Diferencias entre días consecutivos.	Sospechoso	Diferencia >150 mm entre 2 días consecutivos.	Se marca como sospechoso el día actual.	control='control-diferencias', estado='Sospechoso', valor_procesado=valor

Fuente: Elaboración propia

*III. Resultado para precipitación en OSPA*

Después de los procesos descritos anteriormente para la variable precipitación, finalmente se cuenta con 784 estaciones de las cuales 83 también se encuentran en los datos de monitoreo de

precipitación de datos abiertos. A continuación, se describe la cantidad de estaciones para cada rango de disponibilidad de datos en el periodo de análisis 2021-05-01 a 2021-06-20.



**Tabla 38:** cantidad de estaciones para cada grupo de disponibilidad de datos de precipitación de la OSPA.

Disponibilidad de datos	Cantidad Estaciones
0 – 50%	42
50 – 70%	60
70 – 100%	682

Fuente: Elaboración propia

• Conjunto de datos de temperatura máxima y mínima:

**Descripción:** conjunto de datos de Temperatura máxima y mínima para estaciones a nivel diario, para toda Colombia.

II. Revisión y preparación

**Paso 1:** introducir registros de fecha no existentes, es decir tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal. En las fechas agregadas, el dato de la variable tomó el valor del código de identificación para datos faltantes que es el 99999.

**Paso 2:** identificación de duplicidades por fecha, es decir, tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha.

III. Etapa de Control de calidad para datos de temperatura

Teniendo los datos de temperatura máxima y mínima en escala diaria, se procedió a la aplicación de los controles de calidad, descritos en la siguiente tabla:



**Tabla 39: controles de calidad propuestos para los datos de monitoreo de temperatura máxima y mínima de OSPA.**

Nombre	Tipificación	Condición 1	Condición 2	Marcación en Big Query
Identificación de rangos	Error	Temp > 50°C o Temp < -10°C	N/A	control='control-rangos', estado='Error', valor_ procesado=99999
Comparaciones	Error	Tmin > Tmax	N/A	control='control-comparaciones', estado='Error', valor_ procesado=99999
Valores atípicos	Sospechoso (Atípico)	Dato > Cuartil3 +RIC * R // Dato < Cuartil1 - RIC * R, donde R=4	RIC = Q3 -Q1	control='control-atípicos', estado='Sospechoso, valor_ procesado=valor
Identificación de consecutivos	Error	4 o más datos iguales consecutivos	El primer día de los 4 se marca como sospechoso y el resto como error.	control='control-consecutivos', estado='Error' (excepto el 1er que se marca como 'Sospechoso'), valor_procesado=99999
Identificación de consecutivos	Sospechoso	3 datos iguales consecutivos	Todos los 3 días son sospechosos	control='control-consecutivos', estado='Sospechoso', valor_ procesado=valor
Diferencias entre datos	Error	Diferencias mayores a 16,0°C entre dos datos consecutivos	N/A	control='control-diferencias', estado='Error', valor_ procesado=99999
	Sospechoso	Diferencias entre 7,0°C a 15,9°C entre dos datos consecutivos	N/A	control='control-diferencias', estado='Sospechoso', valor_ procesado=valor

Fuente: Elaboración propia

#### IV. Resultado para temperatura en OSPA

Después de los procesos descritos anteriormente, para la variable de temperatura máxima finalmente se tienen 247 estaciones de las cuales 46 también se encuentran en los datos de monitoreo de temperatura máxima de datos abiertos. Para temperatura mínima se tienen 246 estaciones de

las cuales 46 también se encuentran en los datos de monitoreo de temperatura mínima de datos abiertos. A continuación, se describe la cantidad de estaciones para cada rango de disponibilidad de datos en el periodo de análisis 2021-02-01 a 2021-05-25.

**Tabla 40: cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura máxima de OSPA.**

Disponibilidad de datos	Cantidad Estaciones
0 – 50%	25
50 – 70%	21
70 – 100%	201

Fuente: Elaboración propia

**Tabla 41: cantidad de estaciones para cada grupo de disponibilidad de datos de temperatura mínima de OSPA.**

Disponibilidad de datos	Cantidad Estaciones
0 – 50%	10
50 – 70%	24
70 – 100%	212

Fuente: Elaboración propia

### 3.1.2.3 Datos satelitales CHIRPS

La potencialidad en la implementación de los datos CHIRPS en el contexto de monitoreo del SIGRA, consiste en el uso del producto global en escala temporal mensual actualizada disponible. El producto CHIRPS mensual se publica generalmente durante la tercera semana del mes y corresponde a la precipitación acumulada durante el mes anterior a su publicación. El uso principal del producto CHIRPS en el contexto de monitoreo es para el cálculo de amenazas por excesos y/o deficiencias hídricas a escala mensual, así como para el modelamiento de cultivos requerido para determinar los índices de vulnerabilidad.

El proceso propuesto para el alistamiento de los datos CHIRPS para el contexto de monitoreo de SIGRA se describe a continuación:

#### A. Alistamiento de datos

##### i. Obtención de datos CHIRPS:

Se realiza la descarga de los datos CHIRPS mensuales en formato ráster (\*.tif), desde enero de 1981 hasta el último mes disponible, por ejemplo, noviembre de 2020, los cuales se encuentran publicados en el enlace: [https://data.chc.ucsb.edu/products/CHIRPS-2.0/global\\_monthly/](https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_monthly/). Esta descarga se realiza por medio de un script en R, que los reproyecta al sistema de coordenadas WGS84.

##### ii. Información de lluvias CHIRPS históricas por píxel:

Este proceso de extracción es implementado a través de un script en R que organiza en una matriz las coordenadas y las fechas asociadas a cada dato para obtener la precipitación en cada píxel. Esta información se guarda punto a punto en una matriz que posteriormente se descarga en formato CSV, para su almacenamiento en CloudStorage de Google.

Adicionalmente, se realiza la validación de posibles errores sistemáticos que pueden presentarse en las capas ráster para los datos de precipitación. En esta verificación se asegura que la zona de cobertura de la capa corresponda a Colombia, y que el sistema de coordenadas se encuentra en WGS84. También se realiza la identificación de la resolución actual, la cual es de 5 km x 5 km, la cual será ajustada una vez se defina el tamaño de la grilla óptima. Finalmente, se verifica que la corrección de datos para estaciones nacionales fue realizada por la fuente de datos antes de la publicación de la capa, por lo que no se realiza dicha corrección en el proceso de alistamiento de datos.

La base de datos generada y la información almacenada sirve, junto con los datos de lluvia diaria provenientes de datos abiertos (acumulados a 5 días) y del hidroestimador del satélite GOES 16, para complementar la información de precipitaciones disponible de las áreas de interés del SIGRA en el contexto del módulo de monitoreo. Posteriormente se utiliza en el cálculo de índices como el SPI, IDH, PEDCI, entre otros, con los cuáles se realiza el cálculo de amenazas por déficit y exceso hídrico. Adicionalmente, sirve para el cálculo de la vulnerabilidad, al ser la precipitación, una de las variables requeridas por los modelos de cultivos considerados en el marco del SIGRA.

### **B. Controles de calidad**

Para las series de monitoreo de precipitación de CHIRPS no fue necesario realizar procesos de tratamiento de la información adicionales, controles de calidad o procesos de homogenización, ya que la fuente oficial de estos datos es el Climate Hazard Center de la Universidad de California en Santa Bárbara, el cual ya aplica robustos procesos de calidad previos a su publicación. Estos controles se encuentran descritos en el enlace <https://chc.ucsb.edu/data/chirps/diagnostics>.

#### **3.1.2.4 Datos de lluvia satelital del hidroestimador**

La fuente de información corresponde al monitoreo de la precipitación diaria acumulada en 24 horas estimada por el satélite GOES-16, en formato (.tif) georreferenciado para Colombia. Este producto fue

suministrado por el IDEAM en la página web: [http://bart.ideam.gov.co/wrfideam/new\\_modelo/GOES/tif/](http://bart.ideam.gov.co/wrfideam/new_modelo/GOES/tif/).

### **A. Alistamiento de datos**

El alistamiento de los datos consiste en la descarga de la información a través de un script en R, donde se obtiene información desde el 01 de enero hasta una fecha que se define, información que se almacena en el CloudStorage de Google.

Adicionalmente se realiza la validación de los posibles errores sistemáticos que pueden presentarse en las capas ráster para los datos de precipitación. Esta verificación consiste en asegurar que la zona de cobertura de la capa corresponda a Colombia, y que el sistema de coordenadas se encuentra en WGS84. También se realiza la identificación de la resolución actual, la cual es de 4 km x 4 km, resolución que será ajustada una vez se defina el tamaño de la grilla óptima. Finalmente, se verifica que la corrección de datos de estaciones nacionales fue realizada por la fuente de datos antes de la publicación de la capa, por lo que no se realizó dicha corrección en el proceso de alistamiento de datos.

### **B. Controles de calidad**

Para las series de monitoreo de precipitación satelital del hidroestimador no fue necesario realizar procesos de tratamiento de la información adicionales, controles de calidad o procesos de homogenización, ya que la fuente oficial de estos datos es el NOAA STAR (Administración nacional oceánica y atmosférica por sus siglas en inglés), el cual ya aplica robustos procesos de calidad previos a su publicación. Estos controles se encuentran descritos en el enlace <https://www.star.nesdis.noaa.gov/smcd/emb/ff/HEtechnique.php>.

#### **3.1.2.5 Información de precipitación de Fedearroz**

Corresponde a información meteorológica suministrada por Fedearroz a la UPRA. Esta información consolida diariamente los datos meteorológicos de precipitación provenientes de estaciones meteorológicas automáticas desde el 2011 a un período más reciente, como 2021. A continuación, se detalla el proceso de control de calidad realizado a estos datos.

### A. Alistamiento de datos:

#### i. Normalización: organización de la información

Se construye un script en Python para recorrer las carpetas de la VPN. Según el nombre de archivo, se separa el que corresponda y se extrae el nombre de la variable, tipo de estación y el código de la estación. Del contenido del archivo se extrae la fecha y el valor del dato. Luego toda esta información se consolida, definiendo una estructura de campos. Finalmente, se procede a validar formatos de los campos y se cargan a la herramienta BigQuery.

#### ii. Estandarización

Para aplicar los controles de calidad a los datos diarios, se agregan nuevos campos a las tablas en BigQuery. Se agrega el campo 'control' que tiene un valor inicial de 'NA'; luego de ejecutar los controles de calidad, se identifican los datos que fueron marcados en cada control, mediante los valores 'control-rangos', 'control-consecutivos', según corresponda. El segundo campo agregado es el 'estado' que tiene un valor inicial de 'ok' para todos los registros; luego de aplicar los controles, se conserva el valor de 'ok' si el dato pasó todos los controles de calidad, o toma los valores de 'Error' y 'Sospechoso', según como haya sido identificado el dato en los controles. El tercer campo es el 'valor\_procesado' que tiene un valor inicial igual al campo 'valor' de la tabla; luego

de la ejecución de los controles, toma el valor de 99999 cuando el dato es identificado como 'Error' en algún control, en dado caso que no, mantiene el mismo valor inicial.

### B. Procesamiento

#### i. Controles de calidad aplicados a la variable de precipitación

**Descripción:** conjunto de 96.229 datos de precipitación de 40 estaciones automáticas a escala diaria, en el periodo de 2011/08/23 a 2021/05/05.

#### I. Etapa de Revisión y preparación

**Paso 1:** agregación de registros de fecha-hora no existente y su respectivo valor con la convención de dato faltante (99999).

**Paso 2:** identificación de duplicidades por fecha. El objetivo es tener la fecha de comienzo y fin de cada serie y de manera continua su registro temporal, sin que se repita ninguna fecha-hora.

#### II. Etapa de Control de calidad para datos de Precipitación en escala diaria

La Tabla 42 presenta los controles realizados a los datos:



**Tabla 42: controles de calidad aplicados a los datos de precipitación diarios de Fedearroz.**

Nombre	Tipificación	Condición 1	Condición 2	Marcación en Big Query
Identificación de rangos	Error	ppt < 0 mm o ppt > 350 mm	N/A	control='control-rangos', estado='Error', valor_procesado=99999
	Sospechoso	200 mm < ppt ≤ 350 mm	N/A	control='control-rangos', estado='Sospechoso', valor_procesado=valor
Identificación de consecutivos	Error	4 o más días consecutivos.	Sin tener en cuenta valor 0. El primer día de los 4 se marca como sospechoso y el resto como error.	control='control-consecutivos', estado='Error' (excepto el 1er que se marca como 'Sospechoso'), valor_procesado=99999
	Sospechoso	2 y 3 días consecutivos.	Sin tener en cuenta valor 0.	control='control-consecutivos', estado='Sospechoso', valor_procesado=valor
Diferencias entre días consecutivos.	Sospechoso	Diferencia >150 mm entre 2 días consecutivos.	Se marca como sospechoso el día actual.	control='control-diferencias', estado='Sospechoso', valor_procesado=valor

Fuente: Elaboración propia

Desde febrero de 2021, se registran datos de 40 estaciones y hay una disponibilidad del 97.32% de los datos en este periodo, por tanto, sólo falta el 2.68% de los datos.

### 3.1.3 Contexto de pronóstico

En el contexto del proyecto SIGRA se definió la información priorizada para el contexto de pronóstico del tiempo atmosférico (corto plazo) y de predicción climática (escala estacional).

Dicha información proviene de productos como modelos atmosféricos dinámicos y climáticos de tipo estadístico, productos que en la actualidad el IDEAM ofrece a los usuarios. En el marco de interés del SIGRA se priorizaron los siguientes:

#### Pronóstico de tiempo atmosférico:

#### I.

Precipitación pronosticada a 7 días por Modelo Dinámico WRF Colombia (10x10Km) en resolución temporal horaria y diaria.

#### II.

Velocidad de viento en superficie pronosticada a 7 días por Modelo Dinámico WRF para Colombia (10x10Km) en resolución temporal horaria.

#### III.

Velocidad de viento máximodiario (10m de altura) pronosticado a 7 días por Modelo Dinámico WRF para Colombia (10x10Km).

#### IV.

Temperatura a 2m pronosticada a 7 días por Modelo Dinámico WRF Colombia (10x10Km) en resolución temporal horaria.

#### V.

Temperatura mínima diaria pronosticada a 7 días por Modelo Dinámico WRF para Colombia (10x10Km).

#### VI.

Evapotranspiración de referencia (ETo) pronosticada a 7 días por Modelo Dinámico WRF para Colombia (10x10Km).

### 3.1.3.1 WRF - Pronóstico

Los datos para el cálculo de los índices agroclimáticos requeridos para los índices de amenaza por excesos y deficiencias hídricas en el contexto de pronóstico provienen del modelo WRF, suministrados por IDEAM diariamente a través de su página web en escala espacial de 10x10Km. A continuación, se detalla el alistamiento de datos:

#### A. Alistamiento de datos

##### i. Obtención de datos WRF:

El proceso de descarga de archivos se realiza de manera manual y con ayuda de un script de Python que se corre localmente para desagregar los archivos para cada tipo de variable. Este script también realiza una estructura de archivos que permite el cargue de los datos al sistema de almacenamiento.

#### B. Controles de calidad

Para las series de pronóstico de precipitación, temperatura, velocidad del viento y evapotranspiración del WRF no fue necesario realizar procesos de tratamiento de la información adicionales, controles de calidad o procesos de homogenización, ya que la fuente oficial de estos datos es el Modelo Global GFS FV3 de la GFDL (Laboratorio de Dinámica de Fluidos Geofísicos por sus siglas en inglés) el cual ya aplica robustos procesos de calidad previos a su publicación. Estos controles se encuentran descritos en el enlace <https://www.gfdl.noaa.gov/fv3/>.

### 3.1.4 Contexto de predicción climática

Como se indicó antes, la información priorizada para el contexto de pronóstico del tiempo atmosférico y de predicción climática del SIGRA se resume en productos provenientes de modelos dinámicos y estadísticos, de los productos de tiempo y clima que en la actualidad el IDEAM pone a disposición de los usuarios. En el marco de interés del SIGRA se priorizaron los siguientes:

#### **Predicción Climática:**

**I.**  
*Predicción estacional de la Temperatura Máxima, Media y Mínima determinista en °C y hasta para 6 meses, usando modelo CPT clásico.*

**II.**  
*Predicción estacional de la Precipitación determinista hasta para 6 meses, usando modelo CPT clásico.*

#### 3.1.4.1 Datos del modelo CPT

La información de predicción climática se encuentra disponible en: [http://bart.ideam.gov.co/wrfideam/new\\_modelo/CPT/txt/](http://bart.ideam.gov.co/wrfideam/new_modelo/CPT/txt/) y en formato TXT para las variables precipitación, temperatura máxima, temperatura media y temperatura mínima, proveniente del modelo estadístico CPT clásico y para un horizonte de 6 meses, basado en la habilidad de predicción usando el índice de bondad. Esta información se descargó y se validó la completitud de los datos.

Para las series de predicción de precipitación y temperatura del CPT no fue necesario realizar procesos de tratamiento de la información adicionales, controles de calidad o procesos de homogenización, ya que la fuente oficial de estos datos es el IRI (Instituto Internacional de Investigación para el Clima y la Sociedad por sus siglas en inglés) el cual ya aplica robustos procesos de calidad previos a su publicación. Estos controles se encuentran descritos en el enlace: [http://bart.ideam.gov.co/wrfideam/new\\_modelo/DOCUMENTOS/2020/NT\\_001\\_2020.pdf](http://bart.ideam.gov.co/wrfideam/new_modelo/DOCUMENTOS/2020/NT_001_2020.pdf)

## 3.2 Vulnerabilidad

Este capítulo presenta el alistamiento de datos utilizados para el cálculo del componente del riesgo agroclimático: vulnerabilidad. Los datos climáticos requeridos para la modelación de cultivos, herramienta utilizada para estimaciones de indicadores de vulnerabilidad como los rendimientos agrícolas potenciales, tienen procesos de control de calidad y validaciones ya descritos en el capítulo anterior. Por otra parte, algunos datos relacionados con indicadores de vulnerabilidad utilizados en el SIGRA y no incluidos de modo particular en este capítulo son productos publicados por las diferentes instituciones o grupos de trabajo que los generan. Por tal motivo a tales datos no se les realizaron controles de calidad ni procesos de normalización, salvo los relacionados con la validación de la integridad informática de los archivos o datos tomados de las fuentes originales.

### 3.2.1 Parámetros conservativos utilizados en las modelaciones de cultivos involucradas en análisis del componente vulnerabilidad.

Los conjuntos de “parámetros conservativos” corresponden a información que no se clasifica dentro de los contextos: histórico, monitoreo, pronóstico o predicción, al ser información transversal, necesaria e independiente de un contexto temporal y geográfico para el cálculo de la vulnerabilidad.

Como se mencionó anteriormente, diversos datos empleados para las modelaciones de cultivos involucradas en el cálculo de la vulnerabilidad, además de los climáticos, son datos provenientes de productos oficiales de diferentes fuentes y ya cuentan con controles de calidad durante el proceso de su producción.

#### 3.2.1.1 Parámetros de calibración de cultivos en las modelaciones de rendimientos

El conjunto de datos de calibración se basa en la documentación del modelo AquaCrop para Colombia tomado de los anexos de las publicaciones de la FAO en las siguientes rutas:

- ▶ <http://www.fao.org/3/y7231s/y7231s.pdf>
- <http://www.fao.org/3/i3430s/i3430s.pdf>
- <http://www.fao.org/3/i3429s/i3429s.pdf>
- <http://www.fao.org/3/i3428s/i3428s.pdf>





Estos datos comprenden:

**Parámetros conservativos:** son aquellos parámetros agrícolas que suelen ser constantes a nivel temporal y espacial, parámetros tales como:



Coefficiente único de cultivo – KC (dato que expresa la proporción de evapotranspiración de referencia que asume el cultivo para cada etapa fenológica).



Cobertura del dosel (es un valor estimado u observado sobre el % de superficie que cubren las hojas del cultivo en promedio).



Productividad hídrica – WP (medida relativa de la eficiencia del proceso transpirativo del cultivo para formar biomasa, este se puede emplear como constante teórica o ajustarlo mediante normalización a la concentración de CO<sub>2</sub> atmosférico)



Índice de Cosecha – HI (medida relativa de la proporción de biomasa que forma las partes de interés agronómico de la planta que son cosechadas y cuyo peso compone el rendimiento).

**Parámetros no conservativos:** varían de región a región y entre variedades y especies); estos parámetros son aplicables en las zonas para las que fueron generados y

aquellas otras que sean similares desde el punto de vista de la oferta edafoclimática. Algunos de los parámetros no conservativos de interés son:



Coefficiente basal de cultivo – Kcb. Expresa el máximo potencial proporcional de evapotranspiración de referencia de un cultivo en condiciones ideales y por tanto relacionado con la producción potencial de biomasa.



Longitud del ciclo de cultivo. Duración en días o tiempo térmico (Grados Día Calor –GDC) del ciclo total de vida del cultivo o de cada una de sus etapas fenológicas; varía principalmente en función de la altitud sobre el nivel del mar y por características genéticas del cultivar (precocidad).



Curva de crecimiento de la rizósfera  
Cuantifica experimental o teóricamente la profundidad a la que la raíz del cultivo explora el suelo y por lo tanto, define el grosor del perfil de suelo en que se calcula el Agua Total Disponible (ATD) para el cultivo a lo largo de su ciclo. En el modelo interno de UPRA se asume constante, pero es recomendable considerar la evolución temporal de este parámetro para la cuantificación adecuada de la ATD.



Escenarios de fechas de siembra (que definen el punto temporal del inicio de la simulación y que dependen fuertemente de la estacionalidad de las lluvias en sistemas de secano).



Condiciones iniciales de humedad del suelo ADTi-1, el cual es un parámetro altamente sensible del modelo, y no es fácilmente cuantificable.



Es importante resaltar que con base en resultados alcanzados por UPRA en el contexto de un modelo interno de cultivos desarrollado en el marco de sus trabajos en el SIGRA, modelo que se basa el motor de balance hídrico de Aquacrop, para las condiciones iniciales de humedad del suelo, se pueden usar guías incluidas en las publicaciones de FAO antes anotadas, para así mejorar su escalabilidad. . El modelo desarrollado en UPRA requiere sólo algunos parámetros de calibración para su funcionamiento, tales como Kcb y Tr.

### 3.2.1.2 Fechas de siembra o calendarios de cultivo

Las fechas de siembra y los calendarios de cultivos dependen de condiciones climáticas, de suelos, agronómicas y regulatorias de cada año, zona y cultivo. Para efectos del proyecto SIGRA la información utilizada corresponde con las siguientes fuentes:



**Resoluciones ICA:** La información disponible se encuentra ubicada en la página web <https://www.ica.gov.co/normatividad/normas-ica/resoluciones-seccionales> en formato PDF. La cual, corresponde a las fechas de siembra desde el primer semestre del 2017 al primer semestre del 2021. El alistamiento de datos consistió en la descarga de esta documentación para luego ser digitada manualmente en un documento .xlsx.



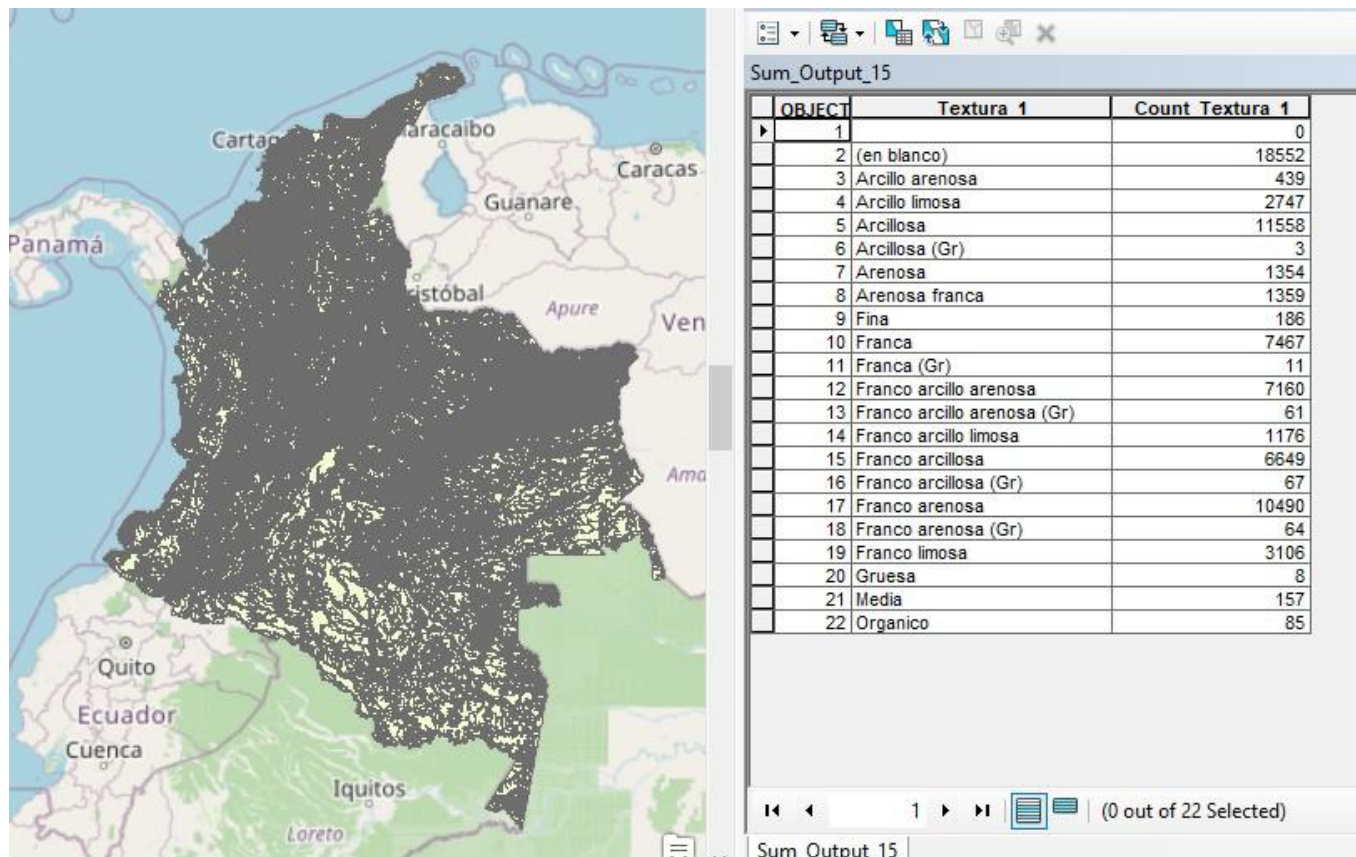
**Mapas de aptitud agroclimática decadales:** La información disponible se encuentra en la página web del IDEAM desde 1981 al 2010: [http://www.ideam.gov.co/inicio?p\\_p\\_id=101&p\\_p\\_lifecycle=0&p\\_p\\_state=maximized&p\\_p\\_mode=view&\\_101\\_struts\\_action=%2Fasset\\_publisher%2Fview\\_content&\\_101\\_assetEntryId=91482640&\\_101\\_type=content&\\_101\\_urlTitle=capas-geo](http://www.ideam.gov.co/inicio?p_p_id=101&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&_101_struts_action=%2Fasset_publisher%2Fview_content&_101_assetEntryId=91482640&_101_type=content&_101_urlTitle=capas-geo). Esta es una información climatológica generada por IDEAM en formato espacial, donde se indican los momentos y zonas más adecuadas para realizar labores agrícolas (incluyendo labores de siembra). El alistamiento de datos consistió en la descarga de esta información en formato PDF y SHP, y cargada en el CloudStorage de Google.

### 3.2.1.3 Atributo de textura del suelo - Capa de correlación

Es una capa que se encuentra en la geodatabase del IGAC para el modelo de cálculo de vulnerabilidad por déficit hídrico. La capa general es geométrica

de tipo polígono en escala 1:100.000 con cobertura nacional y atributo de textura del suelo de tipo texto como se muestra en la Ilustración 2.

**Ilustración 2:** solo para fines ilustrativos, Atributo de textura a nivel Nacional



Fuente: Elaboración propia

- El alistamiento de datos consiste en abrir la capa en ArcGis desktop y exportar la capa a formato shapefile. Finalmente, esta información se sube a la base de datos de CloudStorage de Google para su disponibilidad.

### 3.2.1.4 Datos de coeficientes de cultivos *k<sub>c</sub>*

Los requerimientos hídricos de los cultivos en cada etapa de desarrollo se evalúan a través de la evapotranspiración y de un coeficiente *K<sub>c</sub>* de cultivo que varía a lo largo del ciclo agrícola según cambia la demanda hídrica del cultivo.

Para el desarrollo de la metodología utilizada para el cálculo de la vulnerabilidad, en lo relacionado con la estimación parámetros conservativos y de valores de los coeficientes *K<sub>c</sub>* en las modelaciones

de cultivos, se consultaron diversas fuentes como los procedimientos propuestos en la Tesis Doctoral de César Terán "Determinación de la huella hídrica y modelación de la producción de biomasa de cultivos forrajeros a partir del agua en la Sabana de Bogotá (Colombia)"<sup>10</sup>.

De los trabajos de UPRA sobre el cálculo de la vulnerabilidad frente a déficit y exceso hídrico, se tomaron los coeficientes de cultivo propuestos por la FAO en el estudio de riego y drenaje No. 56 (Allen et. el, 2006) o alternativamente los propuestos en la

10 Disponible en: <https://dialnet.unirioja.es/servlet/tesis?codigo=104032>

parametrización del modelo Aquacrop que según indicó UPRA en los trabajos desarrollados para SIGRA han permitido ajustar mejor los modelos. Esto es recomendado para poder generalizar sobre los diferentes cultivos que puedan estar presentes en las zonas de cultivo. Estos coeficientes Kc son asignados dependiendo de la duración de cada fase fenológica principal y las duraciones en referencia se toman de los trabajos de calibración por defecto, como los realizados por FAO en Colombia y referenciados al inicio del capítulo.

### 3.2.1.5 Parámetros de calibración forrajes, como referencia para estimación de vulnerabilidad en sistemas productivos ganaderos (caso pasturas)

Para el cálculo de la vulnerabilidad en sistemas productivos de ganadería de carne y de leche, se realizaron modelaciones de reducciones en biomasa por impactos de amenazas de deficiencias o excesos hídricos en pasturas, para lo cual se tomaron los coeficientes del cultivo para la avena forrajera y raigrás, propuestos por la Tesis Doctoral de César Terán "Determinación de la huella hídrica y modelación de la producción de biomasa de cultivos forrajeros a partir del agua en la Sabana de Bogotá (Colombia)". Esta información contiene las temperaturas críticas y los parámetros determinados experimentalmente en el estudio, como base para las entradas requerida por el modelo AquaCrop. Los parámetros son los siguientes:

## 1. Fenología del cultivo

### 1.1

Umbral de temperaturas del aire

### 1.2

Desarrollo de la cobertura del dosel verde

### 1.3

Floración

### 1.4

Desarrollo de la zona de raíces

## 2. Transpiración

## 3. Producción de biomasa y formación de cosecha

### 3.1

Productividad del cultivo por el agua

### 3.2

Índice de Cosecha

## 4. Estrés

### 4.1

Estrés hídrico en el suelo

El alistamiento de datos de esta información consistió en la digitación manual de los parámetros en un archivo Excel para ponerlos a disposición en la base de datos del CloudStorage de Google.



### 3.2.2 Contexto histórico

#### 3.2.2.1 Concentración de dióxido de carbono (CO<sub>2</sub>)

Los datos disponibles para el contexto histórico de CO<sub>2</sub> fueron descargados desde el sitio web del Laboratorio de Monitoreo Global (GML) de la NOAA, en escalas temporales, diaria, mensual y anual (en este caso es de gran interés la escala anual y mensual). En los modelos de simulación usualmente esta información se emplea para normalizar el dato de Productividad hídrica "WP" del cultivo, dado que la misma varía en correlación directa a los cambios en contenido de gas carbónico atmosférico, es decir que se presentan aumentos en WP, asociados a aumentos en el CO<sub>2</sub>, y viceversa. Esto es fundamental para dinamizar los motores de simulación de acumulación de biomasa (Ochoa, 2020).

Los datos generados por el GML reflejan las concentraciones promedio de gas carbónico en la atmósfera. Para el contexto histórico se toman los promedios anuales que se presentan en la página web <https://gml.noaa.gov/ccgg/trends/data.html>, como constante para cada año.

Para los cálculos de vulnerabilidad en contexto histórico, se descargaron los datos anuales promedios desde 1959 a 2020 en formato CSV y se verificó la completitud de la información.

#### 3.2.2.2 Evaluaciones Agropecuarias Municipales (EVA) - Estadísticas de productividad agrícola y áreas sembradas

Contienen datos sobre el comportamiento histórico de la producción agropecuaria que son importantes para los análisis de vulnerabilidad en lo relacionado con los rendimientos reales obtenidos durante la ocurrencia de amenazas como sequía, exceso hídrico, y además se utilizan como referente de validación general para los modelos de estimación de rendimientos potenciales u óptimos, sin ocurrencia de amenazas. Esta información se encuentra disponible en la página web de la UPR: <https://www.upra.gov.co/web/guest/consulta-de-informacion>.



En una primera fase se realizó el alistamiento de los datos pertenecientes a las EVAs, para lo cual se verificó la completitud de los datos, desde el año 2007 hasta el año más reciente utilizado, el 2019. Se realizaron dos procesos de validación o control de calidad de la información de las EVAs, mediante la comparación de rendimientos y áreas sembradas EVAs, año 2013 vs datos similares del Censo Nacional Agropecuario (CNA) 2014.

#### ***1. Datos utilizados:***

En primer lugar, se eligieron y descendieron 3 conjuntos de datos, previamente almacenados en el repositorio de BigQuery en Google Cloud. Se eligieron las siguientes tablas:

- censo\_cultivos\_caracterizacion\_UPAS\_T: Contiene la información del CNA para el periodo 2013.

- eva\_2007\_2019: Contiene la información de las EVA en el periodo 2007-2019.

- SIPRA\_Zonas\_APTAS<sup>11</sup>: Contiene la información de las zonas aptas para diferentes tipos de cultivo en el país.

Mediante herramientas de Python, se realizó una exploración de datos para ver las columnas y estadísticas que traía cada dataset.

Aptitud\_alta, Aptitud\_media y Aptitud\_baja. En estos datos se incluye el promedio, la desviación estándar y sus percentiles, entre otros.

Estas tablas fueron unificadas para consolidar la información de las zonas aptas para los cultivos MAIZ, PAPA y ARROZ. Adicionalmente se realizó una exploración de datos estadísticos de las columnas

#### ***2. Transformaciones realizadas:***

La transformación realizada a cada conjunto de datos fue la siguiente.

#### **• Zonas Aptas:**

Estos datos tienen el objetivo de ser usados como filtro para que solo se usen datos de municipios que efectivamente sean aptos para un cultivo en específico.

#### **• EVA:**

Para la preparación de las EVA se realizaron filtros para obtener los cultivos de PAPA, ARROZ SECANO MECANIZADO Y MAIZ TECNIFICADO únicamente.

#### **• CNA:**

Por último, para el CNA, se tuvo que realizar un proceso más exhaustivo de limpieza y transformación de datos. Se reemplazan los "NA" o errores por ceros en las columnas numéricas, se excluye la papa china del análisis, se hace una verificación de si el cultivo tiene riego y se crea una nueva columna con los nombres de los cultivos renombrados.

Luego, se mantuvieron las filas de datos que tuvieran las siguientes características, para que los tipos de cultivo a comparar en EVA y CNA fueran los mismos:

- Cultivos de tipo arroz sin riego y con maquinaria.
- Cultivos de tipo maíz y con altitud menor a 1000 msnm.
- Cultivos de tipo papa.

A continuación, se calculó el rendimiento para cada fila de datos del CNA, usando la fórmula **“producción / área cosechada”**. Este conjunto de datos, a diferencia del proveniente de las EVA (datos municipales), viene por UPA (unidad de producción agrícola), por lo que después se hizo una agrupación por municipio y cultivo y se calculó el promedio de rendimiento, su desviación estándar, y las sumas de las áreas sembradas y cosechadas.

### iii. Cruce y análisis de datos

Para el cruce de los datos se realizó, en primer lugar, un filtro de aquellos municipios que sí estuvieran en zonas aptas para el cultivo específico. Luego, se hizo el cruce de la información del CNA y EVA que tuvieran el mismo cultivo y municipio.

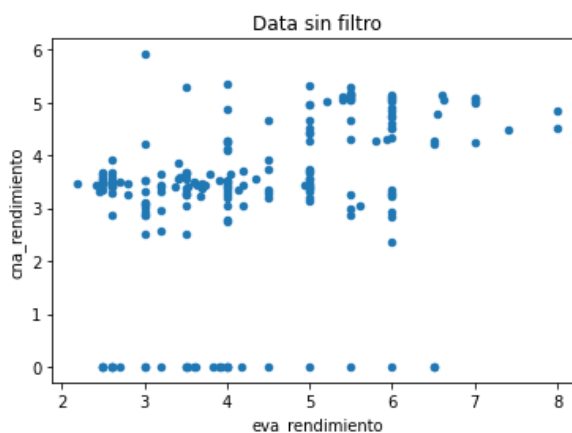
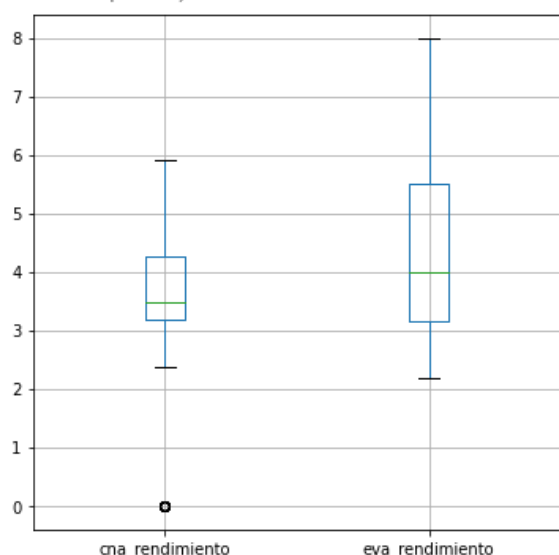
Ya teniendo el cruce de datos, se hizo el cálculo, por municipio y cultivo, de estadísticos de interés. Entre ellos, están:

- Diferencia absoluta entre rendimiento del EVA y CNA para el período común en un grupo de municipios productores.
- Desviación estándar del rendimiento dividida la media del rendimiento.
- La diferencia porcentual de los rendimientos.
- Análisis de correlación de cultivos.

## Maíz

**Ilustración 3: boxplot datos de rendimiento (t/ha) y gráfico de correlación de Maíz entre CNA y EVAs**

```
(
  cna_rendimiento  eva_rendimiento
cna_rendimiento  1.000000      0.359572
eva_rendimiento  0.359572      1.000000,
<AxesSubplot:>)
```



Fuente: Elaboración propia

► Para el maíz, el rendimiento según las EVAs presenta valores intercuartílicos menores que el rendimiento según el CNA para el período común analizado, en los diversos municipios considerados, como se aprecia en los *boxplots* de la Ilustración 4 y en el gráfico de correlación.

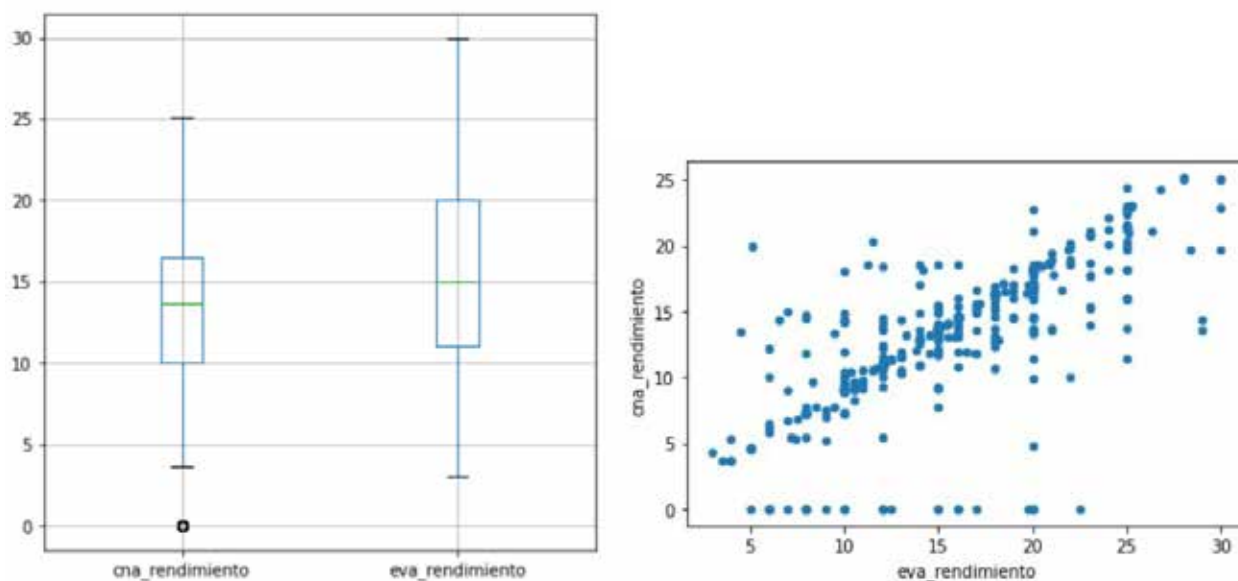
**Papa**

**Ilustración 4: matriz de correlación Papa**

	cna_rendimiento	eva_rendimiento
cna_rendimiento	1.000000	0.730742
eva_rendimiento	0.730742	1.000000

Fuente: Elaboración propia

**Ilustración 5: boxplot de rendimientos (t/ha) y gráfico de correlación para Papa en CNA y en EVA**



Fuente: Elaboración propia

► A diferencia del maíz la distancia intercuartilica de la papa entre el rendimiento de CNA y el de EVA es similar; tanto el primer como el tercer cuartil y la media de los EVA sobrepasan en valor a los valores de CNA. La correlación promedio entre el rendimiento del CNA y del EVA es de 0.73. Esto se puede ver en el gráfico de correlación donde el rendimiento sigue una línea recta que indica una correlación alta entre los valores.



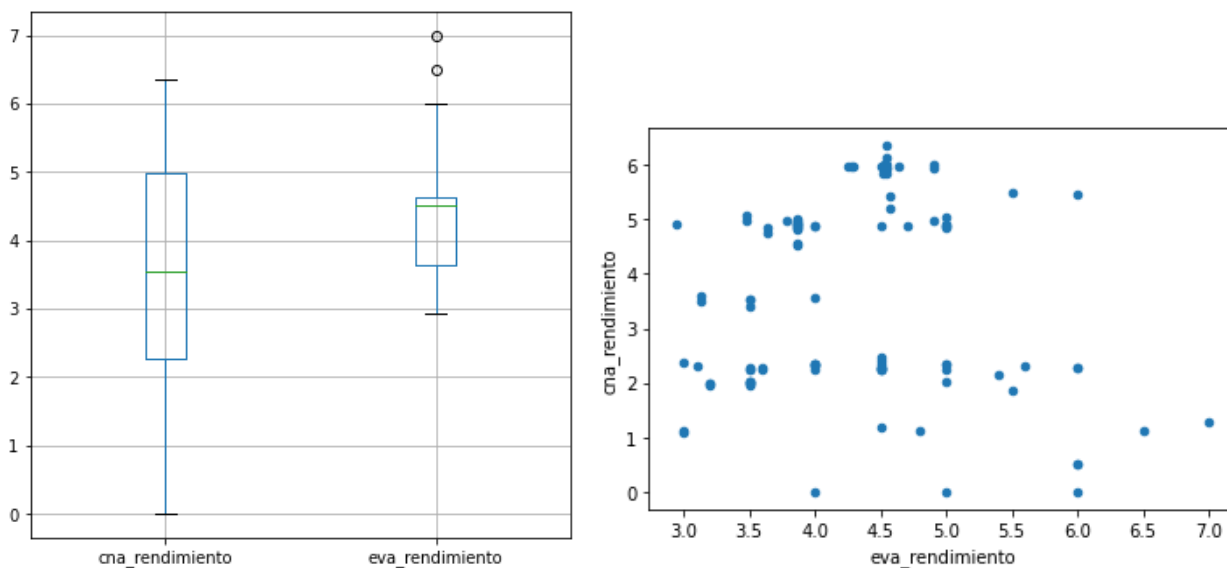
Arroz

**Ilustración 6: Matriz de correlación Arroz**

	cna_rendimiento	eva_rendimiento
cna_rendimiento	1.000000	-0.044224
eva_rendimiento	-0.044224	1.000000

Fuente: Elaboración propia

**Ilustración 7: boxplot de rendimientos (t/ha) y gráfico de correlación para Arroz en CNA y en EVA**



Fuente: Elaboración propia

Finalmente, en el arroz seco se identifica una diferencia significativa entre la distancia intercuartílica del EVA y el CNA. Por un lado, el rendimiento del CNA muestra un tercer cuartil más bajo que en el EVA, pero una mediana más alta. Esto podría explicar el comportamiento de la gráfica de correlación en donde los datos muestran una disparidad entre el rendimiento del EVA y del CNA.

### 1. Análisis de los datos históricos de las EVA

#### i. Datos utilizados:

En primer lugar, se eligió y descargó 2 fuentes de datos. La primera fue extraída del repositorio de BigQuery en Google Cloud, en que se encuentra almacenada, la siguiente tabla de la carpeta interna "censo\_2014": eva\_2007\_2019: Contiene la información de las EVA en el periodo 2007-2019. Para fines de evaluar posibles datos atípicos de rendimientos asociados con anomalías climáticas, una segunda fuente fue la información de episodios fríos y cálidos históricos (eventos ENSO Niño y Niña) representados a través del índice ONI, los cuales son reportados por el USA National Weather Service.

#### ii. Transformaciones realizadas

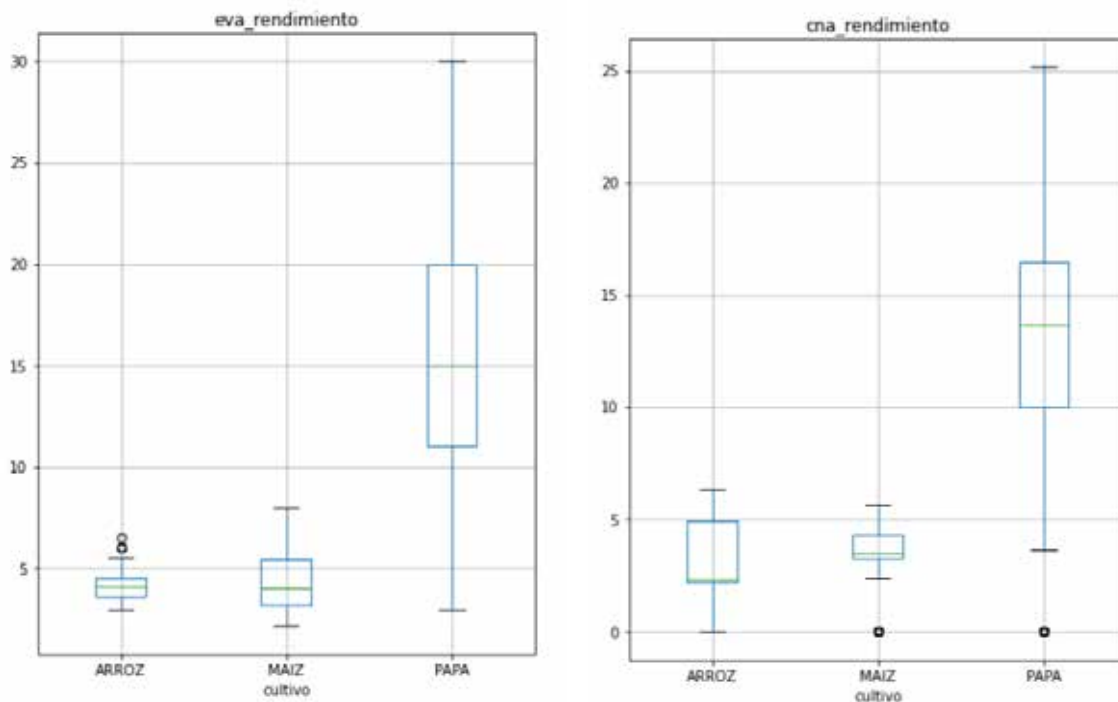
La transformación realizada a los conjuntos de datos fue la siguiente:

- **EVA:** Para la preparación de las EVA se realizaron filtros para obtener los cultivos de papa, arroz seco mecanizado y maíz tecnificado únicamente.
- **Índice ONI:** No se realizó ningún tipo de transformación.

#### iii. Análisis de datos

En primer lugar se determinaron los intervalos en los cuales debería estar oscilando los rendimientos estadísticamente. Para esto, se utilizó la metodología que brinda los boxplot, que permite determinar como dato atípico aquellos que se encuentran por fuera de los bigotes tal y como se observa en la figura.

**Ilustración 8: boxplots de distribución de rendimientos (t/ha)**



Fuente: Elaboración propia

Sobre esta base, se diseñó el algoritmo para extraer estos límites con base en los datos reportados en cada municipio del EVA, obteniendo como resultado la tabla que se observa a continuación, en la que se presentan datos de algunos municipios representativos:

**Tabla 43: intervalos de confianza para el rendimiento (t/ha)**

Dpto	Mpio	Cultivo	Periodo	N_datos_ rendimientos	Promedio_ de rendimiento	Intervalo 2,5%	Intervalo 97,5%
CÓRDOBA	AYAPEL	ARROZ SECANO MECANIZADO	A	12	4,4	3,7	5,2
CÓRDOBA	AYAPEL	ARROZ SECANO MECANIZADO	B	13	4,2	3,1	5,1
HUILA	AIPE	MAÍZ TECNIFICADO	A	13	3,93	3,47	4,50
HUILA	AIPE	MAÍZ TECNIFICADO	B	14	3,86	3,41	4,49
TOLIMA	SANTA ISABEL	PAPA	A	13	16,03	7,00	25,00
TOLIMA	SANTA ISABEL	PAPA	B	9	12,52	5,33	18,00

Fuente: Elaboración propia

Luego de determinar los datos atípicos, se cruzaron estos con la tabla de datos del EVA. Para ello, se creó una columna llamada Dato Atípico cuyo resultado es cero o uno. El primero representa que no es un dato atípico, mientras que el segundo, representa un dato atípico.

Posteriormente, se cruzó esta tabla de datos con su nueva columna con el índice ONI. Para lo descrito, se filtró la tabla de datos con aquellos que marcaron 1 en la columna de Dato Atípico. Luego se observó el periodo del dato atípico y se cruzó con el índice ONI.

El resultado obtenido, fue la creación de dos columnas llamadas: Atípico Por Fenómeno Climático y Fenómeno. La primera columna, muestra si el dato atípico se debe a un fenómeno climático; mientras que la segunda columna, describe que fenómeno o evento climático se presentó en ese periodo.

Finalmente, se realizó este proceso para todos los datos atípicos obtenidos en el primer análisis lo cual dio como resultado la tabla que se observa a continuación.

**Tabla 44: resultado final análisis EVA cruzado con el índice ONI**

Dpto	Mpio	Cultivo	Año	Año_Periodo	Rendimiento	Dato Atípico	Atípico por Fenómeno Climático	Fenómeno
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2009	2009 <sup>a</sup>	1.3	1	SI	Niña
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2009	2009B	1.3	1	SI	Niño
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2010	2010 <sup>a</sup>	1.5	1	SI	Niña
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2010	2010B	1.2	1	SI	Niño
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2011	2011 <sup>a</sup>	2	1	SI	Niña
ANTIOQUIA	ARBOLETES	ARROZ SECANO MANUAL	2011	2011B	2.5	1	SI	Niña
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2011	2011B	2	1	SI	Niña
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2012	2012 <sup>a</sup>	2	1	SI	Niña
ANTIOQUIA	APARTADO	ARROZ SECANO MANUAL	2012	2012B	2	1	NO	NA
ANTIOQUIA	BELMIRA	PAPA	2006	2006B	25	1	SI	Niño
ANTIOQUIA	ABEJORRAL	PAPA	2007	2007 <sup>a</sup>	16	1	NO	NA
ANTIOQUIA	BELMIRA	PAPA	2007	2007 <sup>a</sup>	25	1	NO	NA
ANTIOQUIA	BARBOSA	PAPA	2008	2008 <sup>a</sup>	17	1	SI	Niña
ANTIOQUIA	ABEJORRAL	PAPA	2009	2009 <sup>a</sup>	16.7	1	SI	Niña
ANTIOQUIA	BARBOSA	PAPA	2009	2009 <sup>a</sup>	17	1	SI	Niña
ANTIOQUIA	BARBOSA	PAPA	2010	2010 <sup>a</sup>	17	1	SI	Niño
ANTIOQUIA	ANGOSTURA	PAPA	2011	2011 <sup>a</sup>	6	1	SI	Niña
ANTIOQUIA	ANGOSTURA	PAPA	2014	2014 <sup>a</sup>	26	1	NO	NA
ANTIOQUIA	ANGOSTURA	PAPA	2015	2015 <sup>a</sup>	26	1	SI	Niño

Fuente: Elaboración propia

No obstante lo anterior, en reuniones posteriores con UPRA, entidad beneficiaria, y luego de analizar los procesos antes descritos, que eran requeridos en los Términos de Referencia del Proyecto SIGRA, aquella entidad resaltó que dado que los datos de las EVA se construyen por medio de la metodología de estadísticas por consenso, era recomendable reemplazar los datos de rendimientos agrícolas de las EVA, originalmente procesados, por microdatos de los muestreos geográficos realizados en las Encuestas Nacionales Agropecuarias (ENA), debidamente anonimizados. Si bien los datos que se publican de las ENA se encuentran a nivel departamental o regional, el microdato de base, de carácter interno en el Departamento Nacional de Estadística (DANE), corresponde a un muestreo estadístico a productores en campo, por lo que estos datos puntuales de rendimientos son más precisos y llegan a nivel de predio.

Es importante mencionar que las ENA son diseñadas y realizadas en campo por el DANE, por lo cual estos datos ya cuentan con los controles de calidad apropiados y se consideran de una fuente confiable.

Teniendo en cuenta que los rendimientos agrícolas históricos publicados en las ENA corresponden a escala departamental y que los microdatos se recolectan en sitios que no son los mismos entre uno y otro año, éstos están siendo tratados, en la fase actual del SIGRA, como una referencia localizada para la calibración de los modelos de cultivo con los que se obtienen rendimientos potenciales y anuales o semestrales, necesarios para las estimaciones de vulnerabilidad, según las amenazas priorizadas.

### 3.2.3 Contexto de monitoreo

#### 3.2.3.1 Datos de concentración de dióxido de carbono atmosférico (gas carbónico o CO<sub>2</sub>)

Estos datos son generados por el Laboratorio de Monitoreo Global (GML) de la NOAA, dichos datos reflejan la concentración de gas carbónico promedio de la atmósfera. Para el contexto de monitoreo se descargó el promedio mensual y semanal de la serie dispuesto en la página web <https://gml.noaa.gov/ccgg/trends/data.html>, en formato CSV y se verificó la completitud de los datos.

## 3.3 Exposición

Se presenta a continuación el procedimiento para el cálculo de otro componente del riesgo agroclimático: Exposición. Éste se asume en el SIGRA como el costo de producción expuesto ante diversas amenazas. Esta exposición se calcula como el producto de la matriz de coeficientes técnicos y el vector de los factores de producción,

En una primera fase se consideró apropiado alistar y procesar la información siguiente:

### I.

**Número de hectáreas sembradas en las zonas aptas:** Series históricas de hectáreas cultivadas y/o sembradas.

### II.

**Matrices de coeficientes técnicos:** las matrices de coeficientes técnicos muestran las cantidades necesarias para la producción agropecuarias. Las cuales, se dividen en: insumos y mano de obra.

### III.

**Precios de factores de producción:** es el valor económico de cada uno de los diferentes insumos y mano de obra, que se puedan utilizar en los sistemas productivos.

Una vez se avanzó dentro del proyecto en relacionado con los procesos necesarios para alistar la información sobre áreas históricas sembradas, se reconoció que los datos provenientes de las EVA y las ENA tenían diversos niveles de dificultades para su utilización, como la incertidumbre sobre la localización de los sectores realmente sembrados dentro de cada municipio en el caso de las EVA, y la inexistencia de datos municipales apropiados, en el caso de las ENA. Por ello finalmente y en consenso con UPRA, se optó por trabajar la

exposición en unidades de exposición/hectárea, de tal manera que luego los resultados se puedan extender a áreas totales apropiadas, una vez se disponga de números de hectáreas correspondientes a determinados sistemas productivos en municipios o regiones de interés.

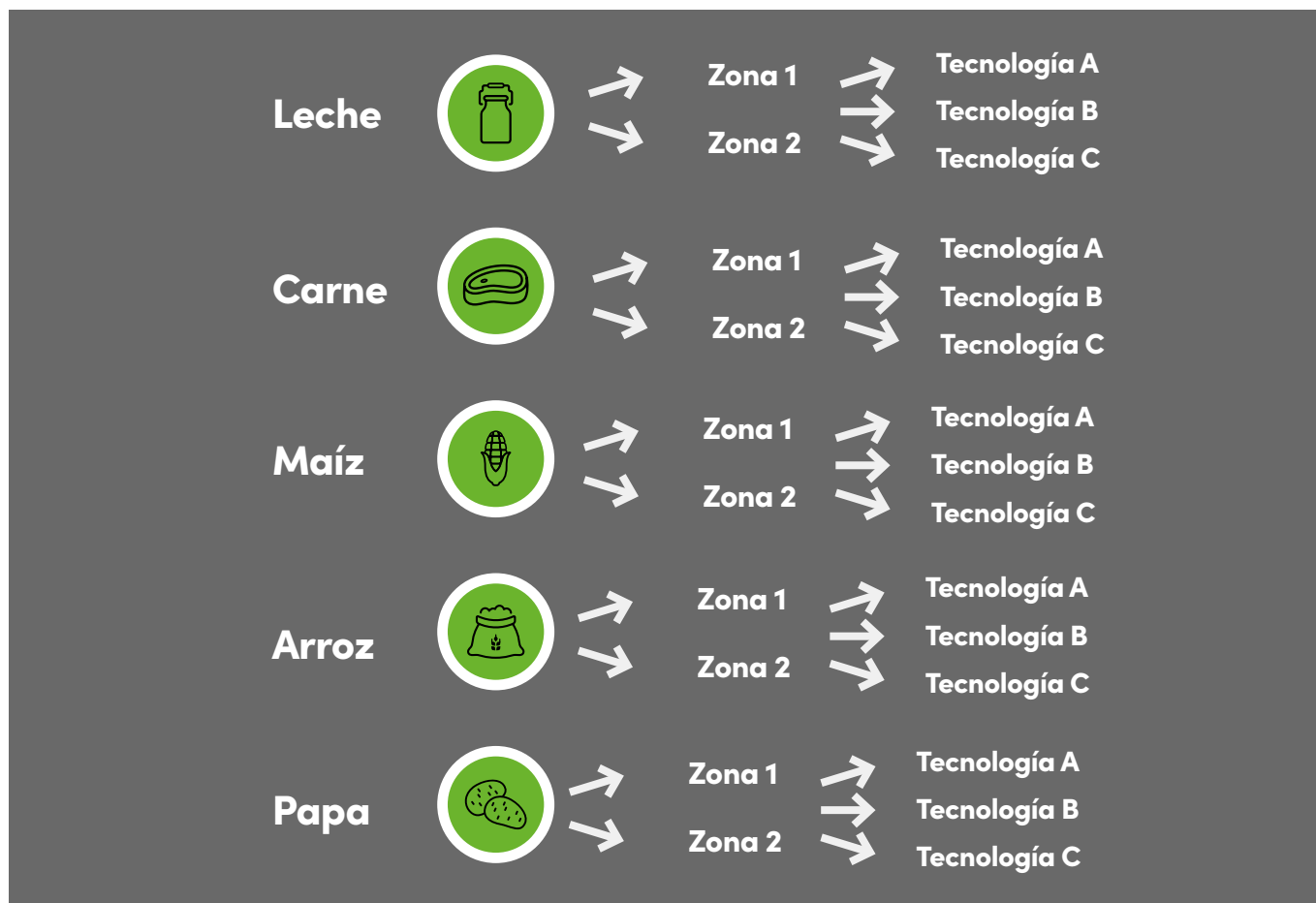
En lo relacionado con la información requerida para la elaboración de las matrices de coeficientes técnicos, según las evaluaciones iniciales realizadas, estos datos no se encuentran disponibles en fuentes públicas, por lo tanto, se optó por utilizar la información disponible en el repositorio de una de las empresas que conformaba la Unión Temporal que realizó el trabajo. La información utilizada sobre este aspecto, e ingresada en el SIGRA, es de consulta pública.

Por otra parte, debido al compromiso que tiene el desarrollo del SIGRA con la inclusión social, será

necesario que, en próximas etapas, las matrices de coeficientes técnicos sean elaboradas para los diferentes niveles de tecnología usados en cada sistema productivo, con énfasis en el nivel de tecnología de los pequeños productores para cada sistema. Adicionalmente, las condiciones climáticas, de suelos, topográficas, prácticas de producción, factores de producción empleados, entre otras, generan variaciones en las matrices de coeficientes técnicos que se deben considerar en próximos trabajos, para que los coeficientes técnicos se elaboren por zonas productivas.

De acuerdo con lo anterior, el número de matrices de coeficientes técnicos depende no solo de los sistemas productivos; también se debe considerar las zonas de producción y el nivel de tecnología como se muestra en la ilustración 10.

**Ilustración 9: Distribución conceptual de las matrices de coeficientes técnicos.**



Fuente: Elaboración propia

### 3.3.1 Matriz de coeficientes técnicos

Como se indicó previamente, dado que la información requerida para la elaboración de las matrices de coeficientes técnicos no se encuentra disponible en una fuente pública, se ha utilizado la información disponible en el repositorio de una de las empresas que conformaba la Unión Temporal que realizó el trabajo. Esta información, que se ha ingresado en el SIGRA, es de consulta pública.

Los controles de calidad realizados para las matrices de coeficientes técnicos se basaron en consultas realizadas con expertos agrónomos, quienes validaron que las cantidades de insumos y mano de obra son las adecuadas y se encuentran dentro de los rangos apropiados para los sistemas productivos relacionados en las matrices.

### 3.3.2 Vectores de precios de factores de producción

Para el cruce de la matriz de coeficientes técnicos con los vectores de precios y el número de hectáreas sembradas, cuyo resultado es la exposición, se realizó la construcción de la matriz de costos a partir de los insumos disponibles en la información de Sistema de Información de Precios (SIPSA) reportada con el DANE en la página web: <https://www.dane.gov.co/index.php/estadisticas-por-tema/agropecuario/sistema-de-informacion-de-precios-sipsa/componente-insumos-1>.

Para el alistamiento de datos, inicialmente se descargaron los precios de los factores de producción que reporta el DANE. Luego de su descarga, se

procedió a listar los datos en la hoja “Matriz de precios” en un formato CSV con los siguientes campos: año, mes, cod\_departamento, nombre\_departamento, cod\_municipio, nombre\_municipio, ciudad, nombre\_producto, presentacion\_producto, precio\_promedio, tipo\_insumo y categoría. Posteriormente, se procedió a la creación de consultas a través de tablas dinámicas en la hoja “Vector de precios”, para poder realizar el cruce con la matriz de coeficientes técnicos; las cuales se construyen de la siguiente forma:

- **Filas:** nombre\_producto y presentacion\_producto
- **Valores:** promedio de la columna precio\_promedio.

Finalmente, al momento de realizar el cruce con la matriz de coeficientes técnicos, se convirtió el precio a la unidad descrita en esta. Para esto, se tuvo en cuenta que si la matriz tiene un insumo cuya unidad de medida es el kg/ha y el precio en SIPSA del producto esta para la presentación de 50 kg, se dividió el precio por la presentación para obtener la misma unidad descrita en la matriz.

Para el control de calidad de los datos de los vectores de precios no fue necesario realizar procesos de tratamiento de la información adicionales, controles de calidad o procesos de homogenización, ya que la fuente oficial de estos datos es el DANE el cual ya aplica robustos procesos de calidad previos a su publicación. Estos controles se encuentran descritos en el enlace: [https://www.dane.gov.co/files/investigaciones/fichas/agropecuario/insumos/DSO-SIPSA\\_I-FME-001.pdf](https://www.dane.gov.co/files/investigaciones/fichas/agropecuario/insumos/DSO-SIPSA_I-FME-001.pdf)



# 04 Conclusiones



- Como un aporte para usuarios interesados en estos temas, se describen los procesos de alistamiento de información, según las características de cada una de las fuentes propuestas, para consolidar una base de datos lo más completa posible, para la implementación de procesos de cálculo de los componentes del riesgo agroclimático, como son: las amenazas, la vulnerabilidad y la exposición.
- Los procesos de alistamiento y control de calidad de datos climáticos propuestos fueron probados y aplicados a la base de datos climáticos históricos proporcionada por el IDEAM a UPRA, en el marco de implementación del SIGRA. Esto permitió consolidar una base de datos históricos, que, debido a la limitada disponibilidad de series de datos de ciertas variables, deberá ser sometida a nuevos procesos de generación de datos faltantes, descritos en el documento. Sin embargo, este proceso de control de calidad ya implementado constituye un avance para próximas etapas de implementación del SIGRA.
- Las series climáticas históricas del IDEAM, disponibles en el portal de datos abiertos fueron descargadas, consolidadas y sometidas a los mismos controles de calidad aplicados a los datos suministrados directamente por el IDEAM. Por ello constituyen una fuente complementaria de información, que luego del control de calidad que sea necesario, se puede integrar con la base suministrada directamente por el IDEAM durante la fase de implementación.
- Los datos meteorológicos del IDEAM, actualizados diariamente en el portal de datos abiertos y los compartidos directamente desde la OSPA, constituyen el principal insumo para el monitoreo meteorológico y de amenazas agroclimáticas requerido en el SIGRA. Estos datos deben ser descargados y sometidos a controles de calidad automáticos, según los procesos expuestos en este documento, para posteriormente ser integrados en el SIGRA y utilizados para el monitoreo de amenazas y la vulnerabilidad.
- Se propone el uso de la base de datos satelitales CHIRPS y siguiendo la recomendación del IDEAM, los datos de lluvias del hidrostimador del satélite GOES 16, para complementar la información de precipitaciones en los submódulos histórico y de monitoreo. Para el efecto se deberá descargar la información en las escalas de tiempo: mensual (histórico), pentadal (5 días, monitoreo) y diario, que será procesada de tal forma que se extraigan los valores de precipitación para cada celda con cobertura en las áreas de cultivos priorizados. Esta información será combinada con la información de las estaciones IDEAM de las bases de datos históricas y de monitoreo (datos abiertos). La información combinada de lluvias resultante podrá ser utilizada en el cálculo de los índices agroclimáticos para las amenazas por déficit o exceso hídrico.
- Para estimar la distribución territorial de las variables climáticas en los módulos histórico y de monitoreo, necesarios para la generación de los productos de amenazas agroclimáticas, se utilizan modelos digitales de elevación, con el fin de explicar de manera más precisa el efecto del terreno, por ejemplo, en la temperatura y la precipitación, para lo cual se plantea el uso de métodos de interpolación que permitan relacionar los valores observados, su localización y parámetros del relieve como elevación, pendiente y aspecto.
- Se analizaron diversos modelos de pronóstico del tiempo y predicción climática estacional del IDEAM que podrían ser utilizados para el cálculo de las amenazas en los contextos de pronóstico y predicción, sin embargo, se requiere aclarar y acordar con IDEAM la interoperabilidad de

- dichos productos de información, pues no resultan operables para los cálculos que se requieren de amenaza y vulnerabilidad futuros, debido a los formatos tipo imagen en los que están publicados. Además, antes de integrarlos al SIGRA, se considera necesario aclarar detalles técnicos relevantes para la extracción y procesamiento de los datos necesarios para el cálculo de las amenazas.
- En lo relacionado con la información necesaria para calcular la vulnerabilidad frente al componente de riesgo: Amenaza,, que en el contexto del SIGRA se ha sintetizado a partir de modelaciones necesarias para estimar reducciones productivas cuando ocurre la amenaza, por ser información proveniente de productos o publicaciones oficiales, en general no requiere de implementación de controles de calidad sobre los datos, salvo los asociados a la verificación de integridad informática de los mismos. En el documento se presentan las fuentes de referencia utilizadas para estimar o tomar datos sobre parámetros “conservativos” y “no conservativos” necesarios en las modelaciones de rendimientos anuales y potenciales de los sistemas productivos involucrados en esta etapa inicial del SIGRA.
  - La Exposición, que en el SIGRA se asume como el costo de producción expuesto ante diversas amenazas, se calcula como el producto de la matriz de coeficientes técnicos y el vector de los factores de producción por unidad de área sembrada.
  - El control de calidad realizado para las matrices de coeficientes técnicos utilizadas se basó en consultas realizadas con expertos agrónomos en este tema, quienes validaron las cantidades de insumos y mano de obra incluidas en las matrices y verificaron que los valores se encuentran dentro de los rangos apropiados para los sistemas productivos relacionados en las matrices.
  - Para el cruce de la matriz de coeficientes técnicos con los vectores de precios, cuyo resultado es la exposición por unidad de área, se construyó una matriz de costos a partir de insumos disponibles en la información de Sistema de Información de Precios (SIPSA) reportada en el DANE.

## GLOSARIO

**Amenaza o Peligro:** cualquier elemento o evento capaz de producir un daño. Acaecimiento potencial de un suceso o tendencia físico de origen natural o humano, o un impacto físico, que puede causar pérdidas de vidas, lesiones u otros efectos negativos sobre la salud, así como daños y pérdidas en propiedades, infraestructuras, medios de subsistencia, prestaciones de servicios, ecosistemas y recursos ambientales (IPCC, 2018). En este documento La amenaza climática se define como el factor externo al sistema expuesto, representado por la ocurrencia potencial de un suceso de origen natural, que puede causar daños en un lugar específico, con determinada intensidad y duración (CEBAR , 2018).

**Cámara Induarroz:** Cámara sectorial de la ANDI creada con el objetivo de garantizar la sostenibilidad de la actividad arrocera y la competitividad del conjunto de los agentes vinculados a la cadena. Agrupa y representa desde 1965 a los industriales del arroz en Colombia en los diferentes escenarios públicos y privados a nivel nacional e internacional (ANDI, 2020).

**Ciclo de vida:** período compuesto por un conjunto de etapas sucesivas de un organismo, desde su emergencia o nacimiento hasta su reproducción y/o muerte.

**Desarrollo epidémico:** conjunto de eventos relacionados con la propagación en el tiempo, de una enfermedad o ataque de plaga en un cultivo.

**Diseminación o dispersión:** fenómeno mediante el cual un patógeno o insecto es transportado o se moviliza de un sitio a otro.

**Edáfico - Edáfica:** perteneciente o relativo al suelo; resultante de o influido por factores del suelo o de otros sustratos en lugar de por factores climáticos. Una necesidad edáfica del cultivo en lo que respecta a un estado o serie de estados del entorno del suelo. (FAO, 2009).

**Evapotranspiración de Referencia (ET<sub>0</sub>):** proceso biológico - físico que consiste en la salida de agua en forma de vapor, desde el suelo hacia la atmósfera. Este proceso se compone de dos subprocesos: 1. Evaporación, que es el fenómeno de pérdida de agua por cambio de fase líquido a gaseoso, desde la superficie desnuda del suelo hacia el aire, esto sucede por efecto de la radiación solar y la temperatura del suelo. 2. Transpiración, fenómeno biológico que ocurre mediado por la actividad biológica de la cobertura vegetal presente en el suelo y que ocurre como consecuencia de la toma de nutrientes en fase acuosa en la raíz de las plantas, por medio del fenómeno de capilaridad, el agua pasa del suelo hacia la planta y llega a las hojas, donde ocurre el intercambio gaseoso necesario para la fotosíntesis (toma de CO<sub>2</sub> y liberación de vapor de agua y O<sub>2</sub> hacia la atmósfera). La suma de los dos procesos es la evapotranspiración, se denomina de referencia cuando la cobertura que se utiliza para su estimación es una pastura con unas características específicas de altura, y resistencia al flujo del aire, a partir de ésta se puede calcular la evapotranspiración de cualquier otra cobertura vegetal (o cultivo específico), por medio de coeficientes de ajuste fenológicos (K<sub>c</sub>) (FAO 2006).

**Helada:** según FAO (2010) e IDEAM (2012), una "helada" es la ocurrencia de una temperatura del aire de 0°C o inferior, medida a una altura de entre 1,25 y 2,0 m por encima del nivel del suelo, dentro de una garita meteorológica adecuada. Un evento de helada se convierte en un evento de congelación cuando se forma hielo extracelular dentro de las plantas, lo cual puede o no dañar el tejido de la planta, según factores de tolerancia, como, por ejemplo, el contenido de solutos de las células.

**Modelo Digital de Elevación (MDE):** SITNA (2020) lo define como una representación visual y matemática de los valores de altura con respecto al nivel medio del mar, que permite caracterizar las formas del relieve y los elementos u objetos presentes en el mismo. Estos valores están contenidos en un archivo de tipo ráster con estructura regular, el cual se genera utilizando equipo de cómputo y software especializados. En los modelos digitales de elevación existen dos cualidades esenciales que son la exactitud y la resolución horizontal o grado de detalle digital de representación. Los modelos digitales de elevación pueden ser de dos tipos: modelos digitales de superficie (MDS) y modelos digitales de terreno (MDT).

**Modalidad productiva o situación de producción:** conjunto de prácticas aplicadas en el manejo de un cultivo, en unas condiciones específicas conocidas.

**Precocidad:** característica de una variedad o híbrido que alcanza la madurez reproductiva y fisiológica en tiempo más corto que otros genotipos de la especie.

**Predicción Climática:** las predicciones climáticas son declaraciones intrínsecamente probabilísticas sobre las condiciones climáticas futuras en escalas de tiempo desde estaciones hasta décadas o más, y en escalas espaciales que van desde lo local hasta lo regional y global. Una predicción climática es una estimación de la evolución real del clima futuro, por ejemplo, a escala estacional. Como la evolución futura del sistema climático puede ser muy sensible a las condiciones iniciales, estas predicciones suelen tener carácter probabilístico, y solemos hablar, más que de predicciones, de evolución probable del clima (AEMET 2021).

**Resistencia genética:** atributo de una variedad o híbrido consistente en su capacidad para impedir, retardar o soportar el ataque de una plaga o patógeno.

**Riesgo:** Probabilidad de que un evento indeseable se presente, así como sus consecuencias o impactos.

**Sistema de modelamiento de cultivos:** es un conjunto de modelos que pretenden simular de forma simplificada el comportamiento biológico de un sistema de producción agrícola. Un sistema de modelamiento consta de uno o varios modelos de cultivo específicos. El objetivo de un modelo es establecer relaciones e interacciones entre los elementos que componen el sistema de producción, para describir por medio de ecuaciones su dinámica de desarrollo y aproximar por medio de datos y parámetros el resultado de las variables de respuesta del sistema como métricas biológicas y consumo de recursos. Existen dos tipos de modelos: 1. Empíricos, que surgen del establecimiento estadístico de relaciones entre las variables de entrada y las de salida del sistema y que son el resultado de experimentos que miden las entradas y salidas de un sistema modelo en condiciones controladas. 2. Mecanicistas: que abordan la simulación mediante la descripción de los procesos fundamentales físicos y/o biológicos que gobiernan el comportamiento del sistema, generando una interrelación entre las ecuaciones que los componen, para modular el comportamiento entre las entradas y las salidas del sistema.

**Supervivencia:** Capacidad de un organismo o microorganismo para mantenerse viable después de estar expuesto a eventos o condiciones ambientales desfavorables.

**Vector:** organismo portador y transmisor de un patógeno.

**Vulnerabilidad:** propensión o predisposición a ser afectado negativamente. La vulnerabilidad comprende una variedad de conceptos que incluyen la sensibilidad o susceptibilidad al daño y la falta de capacidad de respuesta y adaptación (IPCC, 2018).

## BIBLIOGRAFÍA

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, 6, 661-675. DOI: <https://doi.org/10.1002/joc.3370060607>

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (2006). *Evapotranspiración del cultivo: guías para la determinación de los requerimientos de agua de los cultivos*. Roma: FAO, 298(0).

Asibi, A., Chai, Q. and Coulter, J. 2019. Rice blast, a disease with implications for global food security. *Agronomy* 9, 451; doi:10.3390/agronomy9080451.

ANDI. (2020). Inudarroz. Retrieved from Información general website: <http://www.andi.com.co/Home/Camara/14-induarr>

Barragán, A., Pollet, A., Prado, M., Lagnaoui, A., Onore, G., Aveiga, I., Lery, X. y Zeddani, J. 2004. La polilla guatemalteca *Tecia solanivora* (Povolny) (Lepidoptera: Gelechiidae) en Ecuador. Diagnóstico y perspectivas de manejo bajo un método de predicción. Páginas 5-23 En: Pollet, A., Onore, G., Chamorro, F. y Barragán, Eds. *Memorias de II Taller Internacional de Polilla Guatemalteca*. Quito, 2004.

Barreto, N., Espitia, E., Galindo, R., Gordo, E., Cely, L. y Lopez, A. 2004. Fluctuación de la población de *Tecia solanivora* (Povolny) (Lepidoptera: Gelechiidae) en tres intervalos de altitud en Cundinamarca y Boyacá, Colombia. Páginas 25 - 43 En: Pollet, A., Onore, G., Chamorro, F. y Barragán, Eds. *Memorias de II Taller Internacional de Polilla Guatemalteca*. Quito, 2004.

Bloomberg. (2020). Our business is built on change. Retrieved from <https://www.bloomberg.com/>

Bonilla Morales, N. 2017. La mancha de asfalto en maíz. Conferencia presentada en el Encuentro de los sectores frijol y maíz. Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria. INTA Costa Rica. (Manuscrito no publicado).

Bordin, L. C. et al. 2016. Critical-point models to relate yield and disease intensity of the multiple pathosystem in rice leaf spots disease. *Ciencia Rural* 46 (1): 7-12.

Brealey, R. A., Myers, S. C., & Allen, F. (2014). *Principles of corporate finance*. <https://doi.org/10.1093/acprof:oso/9780199671342.001.0001>

Brown, S., Culbreath, A., Todd, J., Gorbet, D., Baldwin, J. and Beasley, J. 2005. Development of a method of risk assessment to facilitate integrated management of spotted wilt of peanuts. *Plant Disease* 89 (4): 348-356.

Casas, M., & Cepeda, E. (2008). Modelos ARCH, GARCH y EGARCH: aplicaciones a series financieras. *Cuadernos de Economía*, 27(1), 287-319.

CME Group. (2020). About us.

Correa, F. Tarreau, D., Martínez, C., vales, M., Escobar, F. Prad, G. y Aricapa, G. 2002. Combinaciones de genes en arroz para el desarrollo de resistencia durable a *Pyricularia grisea* en Colombia. *Fitopatología Colombiana* 26 (2): 47-554.

Cryer, J., & Chan, K. (1985). Time series analysis. In *European Journal of Operational Research* (Vol. 20). [https://doi.org/10.1016/0377-2217\(85\)90052-9](https://doi.org/10.1016/0377-2217(85)90052-9)

FAO. 2009. Glosario de agricultura orgánica. Recuperado de: <http://www.fao.org/tempref/docrep/fao/012/k4987t/k4987t.pdf> PDF

FAO. 2010. Protección contra las heladas: fundamentos, práctica y economía. Volumen 2. Serie Sobre el Medio Ambiente y la Gestión de los Recursos Naturales. ISSN 2071-0992. Roma, Italia.

Funk, C., Peterson, P., Landsfeld, M. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data* 2, 150066 (2015). <https://doi.org/10.1038/sdata.2015.66>

García, A., Ferreira, C., Godoy, W. and Meagher, R. 2018. A computational model to predict the population dynamics of *Spodoptera frugiperda*. *Journal of Pest Science* 1-13. <https://doi.org/10.1007/s10340-018-1051-4>.

Greene, W. (2018). *Econometrics Analysis*. Retrieved from <http://library1.nida.ac.th/termpaper6/sd/2554/19755.PDF>

Greer, C. A., and Webster, R. K. 2001. Occurrence, distribution, epidemiology, cultivar reaction, and management of rice blast disease in California. *Plant Dis.* 85:1096-1102.

Grönberg, L., Andersson, B., and Yuen, J. 2012. ¿Can weed hosts increase aggressiveness of *Phytophthora infestans* on potato? *Phytopathology* 102:429-433.

Henderson, D., Williams, C. J., and Miller, J. S. 2007. Forecasting late blight in potato crops of southern Idaho using logistic regression analysis. *Plant Dis.* 91:951-956.

IDEAM (2012). Actualización Nota Técnica Heladas 2012. Subdirección de Meteorología. Contrato N° 201/2012. Bogotá, 79 p. Recuperado de: <http://www.ideam.gov.co/documents/21021/21147/Documento+FINAL+actualizacion+nota+tecnica+heladas.PDF/e10a0183-62e6-410a-8e96-7e0739f6f06b>

IPCC 2018. Glosario del Grupo de Trabajo II AR5. [https://www.ipcc.ch/site/assets/uploads/2018/02/AR5\\_WGII\\_glossary\\_ES.PDF](https://www.ipcc.ch/site/assets/uploads/2018/02/AR5_WGII_glossary_ES.PDF)

Jaramillo, C., Barragán, E. and Monje, B. 2019. Populations of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) cause significant damage to genetically modified corn crops. *Rev. Fac. Nac. Agron. Medellín* 72(3): 8953-8962.

Johnson, D. A., Cummings, T. F., Abi Ghanem, R., and Alldredge, J. R. 2009. Association of solar irradiance and days of precipitation with incidence of potato late blight in the semiarid environment of the Columbia Basin. *Plant Dis.* 93:272-280.

Kim, C. K. and Kim, C. H. 1993. The rice leaf blast simulation model EPIBLAST. En: F. W. T. Penning de Vries et al. (eds.), *Systems Approaches for Agricultural Development*, 309-321. © 1993 Kluwer Academic Publishers.

Márquez, M.L. (2018). Propuesta metodológica para el análisis de riesgo en el comercio internacional de animales y sus productos. Trabajo de grado para optar el título de Magister. Universidad de la Salle.

Martínez-Jaime, O., Salas, M. y Díaz, J. 2018. Curva de crecimiento poblacional de *Spodoptera frugiperda* en maíz en Irapuato, Guanajuato, México. *Agronomía Mesoamericana* 29 (2): 315-323.

Maziero, J. M. N., Maffia, L. A., and Mizubuti, E. S. G. 2009. Effects of temperature on events in the infection cycle of two clonal lineages of *Phytophthora infestans* causing late blight on tomato and potato in Brazil. *Plant Dis.* 93:459-466.

Montgomery, D., & Runger, G. (2008). *Probabilidad y Estadística aplicada a la Ingeniería*. Limusa Wiley.

Morales, F. and Jennings P. 2010. Rice hoja blanca: a complex plant-virus-vector pathosystem. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* 5 (043). doi: 10.1079/PAVSNNR20105043.

Mousanejad, S., Alizadeh, A. and Safaie, N. 2009. Effect of weather factors on spore population dynamics of rice blast fungus in the Guilan Province. *Journal of Plant Protection Research* 49 (3): 319-329.

Narayanasamy, M., Kennedy, J. and Geethalakshmi, V. 2017. Weather Based Pest Forewarning Model for Major Insect Pests of Rice – An Effective Way for Insect

Olatinwo, R. O., Paz, J. O., Brown, S. L., Kemerait, R. C., Jr., Culbreath, A. K., Beasley, J. P., Jr., and Hoogenboom, G. 2008. A predictive model for spotted wilt epidemics in peanut based on local weather conditions and the Tomato spotted wilt virus risk. *Phytopathology* 98:1066-1074.

Ochoa Cobeña, C. P. (2020). Calibración, validación y utilización del modelo AquaCrop para el cultivo de la chufa (*Cyperus esculentus* L. var. *sativus* Boeck.) en Valencia.

Otte, E. (Editor). (1992). Anaplasmosis y Babesiosis en Colombia. Informe Técnico. Proyecto Colombo-Aleman ICA-GTZ. Santafe de Bogotá.

Pest Prediction. *Annual Research & Review in Biology* 21 (4): 1-13.

Pollet, A. et al. 2003. Predicción de daños de la polilla guatemalteca *Tecia solanivora*.

Pring, M. J. (2014). *Study guide for Technical Analysis Explained*. Mc Graw Hill.

(Povolny) 1973 (Lepidoptera: Gelechiidae) en el Ecuador. *Bol. San. Veg. Plagas*, 29: 233-242.

Quandl. (2020). The world's most powerful data lives on Quandl.

Raveloson, H., Ramonta, I., Tharreau, D. and Sester, M. 2018. Long-term survival of blast pathogen in infected rice residues as major source of primary inoculum in high altitude upland ecology. *Plant Pathology* 67: 610-618.

Reuter H.I, A. Nelson, A. Jarvis, 2007, An evaluation of void filling interpolation methods for SRTM data, *International Journal of Geographic Information Science*, 21:9, 983-1008.

Salazar, C. y Betancourth, C. 2009. Evaluación de extractos de plantas para el control de la polilla guatemalteca (*Tecia solanivora*) en cultivos de papa en Nariño, Colombia. *Agronomía Colombiana* 27 (2): 219-226.

Sanchez, A., Ortiz, D., Tobón, J., Camargo, H., Gutiérrez, M. (Sin fecha) Proyecto piloto VECOL de excelencia sanitaria. Ganadería bovina de leche. Bogotá.

Savary, S., Teng, P., Willocquet, L. and Nuttr, F. 2006. Quantification and modeling of crop losses. A review of purposes. *Annu. Rev. Phytopathol.* 2006. 44:4.1-4.24. doi: 10.1146/annurev.phyto.44.070505.143342.

Savary, S., Willocquet, L., Elazegui, F., Teng, P., Du, P., Zhu, D., Tang, Q., Huang, S., Lin, X., Singh, H., and Srivastava, R. 2000. Rice pest constraints in tropical Asia: Characterization of injury profiles in relation to production situations. *Plant Dis.* 84:341-356.

Seebold, K. W., Datnoff, L. E., Correa-Victoria, F. J., Kucharek, T. A., and Snyder, G. H. 2000. Effect of silicon rate and host resistance on blast, scald, and yield of upland rice. *Plant Dis.* 84:871-876.

Shakya, S. K., Goss, E. M., Dufault, N. S., and van Bruggen, A. H. C. 2015. Potential effects of diurnal temperature oscillations on potato late blight with special reference to climate change. *Phytopathology* 105:230- 238.

Shakya, S. K., Goss, E. M., Dufault, N. S., and van Bruggen, A. H. C. 2015. Potential effects of diurnal temperature oscillations on potato late blight with special reference to climate change. *Phytopathology* 105:230-238.

SITNA. (2020). Modelos Digitales De Elevación – Estimación Del Potencial De Producción De Energía Eléctrica Fotovoltaica En Las Cubiertas De Edificaciones En Dos Municipios. Obtenido de: <https://pcsitna.navarra.es/Lists/ManualesSitna/Ver.aspx?ID=158#:~:text=Un%20MDE%20es%20una%20representaci%C3%B3n,objetos%20presentes%20en%20el%20mismo.&text=e%20Modelo%20Digital%20de%20Superficies,%2C%20infraestructuras%2C%20vegetaci%C3%B3n%2C%20etc>

Sorza, S. M. 2014. Informe final del contrato PS0482 de 2014. CORPOICA. Proyecto “Reducción del riesgo y adaptación al cambio climático”.

Sporleder, M., Kroschel, J., Gutierrez, M. and Lagnaoui, A. 2004. A Temperature-Based Simulation Model for the Potato Tuberworm, *Phthorimaea operculella* Zeller (Lepidoptera; Gelechiidae). *Environ. Entomol.* 33(3): 477-486.

Urrea, V., Ochoa, A., & Mesa, O. (2016). Validación de la base de datos de precipitación CHIRPS para Colombia a escala diaria, mensual y anual en el período 1981-2014. In XXVII Congreso Latinoamericano de Hidráulica. Retrieved from <http://ladhi2016.org/>

Valdez-Torres, J., Soto-Landeros, F., Osuna-Enciso, T. y Báez-Sañudo, M. 2012. Modelos de predicción fenológica para maíz blanco (*Zea mays* L.) y gusano cogollero (*Spodoptera frugiperda* J. E. Smith). *Agrociencia* 46: 399-410.

VARGAS, A., Santos, A. N. A., CÁRDENAS, E., & OBREGÓN, N. (2011). Análisis de la distribución e interpolación espacial de las lluvias en Bogotá, Colombia. *Dyna*, 78(167), 151-159.

Vivas-Carmona, L., Astudillo, D. y Monasterio, P. 2017. Fluctuación poblacional del insecto sogata, *Tagosodes orizicolus* empleando una trampa de luz y su relación con variables climáticas en Calabozo Estado Guárico, Venezuela. *J Selva Andina Biosph.*5(2):70-79.

Wang, C., Zhao, J., & Huang, M. (2010). Measurement of the fluctuation risk of the China fruit market price based on VaR. *Agriculture and Agricultural Science Procedia*, 1, 212-218. <https://doi.org/10.1016/j.aaspro.2010.09.026>

Willoquet, L., Elazegui, F., Castilla, N., Fernandez, L., Fischer, K., Peng, S., Teng, P., Srivastava, R., Singh, H., Zhu, D., and Savary, S. 2004. Research priorities for rice pest management in tropical Asia: A simulation analysis of yield losses and management efficiencies. *Phytopathology* 94:672-682.

Willoquet, L., et al. 2002. Structure and validation of RICEPEST, a production situation - driven, crop growth model simulating rice yield response to multiple pest injuries for tropical Asia. *Ecological Modelling* 153 (2002) 247-268.

Yáñez, R., Vásquez, A., Arreguín, J., Soria, J. y Quijano, J. 2019. Sistema de alerta contra el gusano cogollero *Spodoptera frugiperda* (J. E. Smith) (Insecta: Lepidoptera: Noctuidae). *Revista Mexicana de Ciencias Agrícolas* 10 (2): 405-416.

Zapata, C. J., Piñeros, L. C., & Castaño, D. A. (2004). El método de simulación de montecarlo en estudios de confiabilidad de sistemas de distribución de energía eléctrica. *Scientia et Technica*, (24), 55-60. Retrieved from <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/7297/4293>

Zou, H. F., Xia, G. P., Yang, F. T., & Wang, H. Y. (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing*, 70(16-18), 2913-2923. <https://doi.org/10.1016/j.neucom.2007.01.009>

Zuluaga, A., (2009) Factores de Riesgo asociados a Leptospirosis en Hatos Bovinos de Pereira 2002-2005. *INVESTIGACIONES ANDINAS* No. 19 Vol 11.







Embajada Británica  
Colombia

**CAF** BANCO DE DESARROLLO  
DE AMÉRICA LATINA