

Instituto Tecnológico de Costa Rica
Vicerrectoría de Investigación y Extensión
Dirección de Proyectos

Informe final del proyecto de investigación
Caracterización de la geometría de la zona de subducción, deformación de la corteza y transferencia de esfuerzos en el sur de Costa Rica mediante análisis de datos masivos y simulaciones computacionales

2021

Tabla de contenido

Código y título del proyecto	3
Autores y direcciones	3
Resumen	3
Palabras clave	4
Introducción	4
Marco Teórico	5
Metodología	7
Resultados	10
Discusión y conclusiones	12
Recomendaciones	13
Referencias	13
Anexo	16

1. Código y título del proyecto

1.1. Código del Proyecto: **5402-1370-0009**

1.2. Nombre del proyecto: **Caracterización de la geometría de la zona de subducción, deformación de la corteza y transferencia de esfuerzos en el sur de Costa Rica mediante análisis de datos masivos y simulaciones computacionales**

2. Autores y direcciones

Investigador	Grado Académico	Afiliación
Esteban Meneses Rojas (coordinador)	Doctorado	Escuela de Computación, TEC
Marino Protti Quesada	Doctorado	Observatorio Vulcanológico y Sismológico de Costa Rica (OVSICORI-UNA)
Esteban Chaves Sibaja	Doctorado	Observatorio Vulcanológico y Sismológico de Costa Rica (OVSICORI-UNA)
Yuen Law Wan	Doctorado	RWTH Aachen University (Alemania)

3. Resumen

Dada la complejidad tectónica del continente (Protti et al., 1991; DeMets et al., 2010), Costa Rica es un país con un altos niveles de actividad sísmica y volcánica. Caracterizar las diferentes fuentes generadoras de riesgo en términos de su distribución espacial y temporal es crucial para establecer programas de alerta temprana y gestión del riesgo, así como para el mejoramiento de las normativas de construcción de edificaciones en todo el país. Gracias a una continua densificación instrumental, tanto sísmica como de GPS, se ha podido identificar que en la zona sur del país existe el potencial para generar terremotos con magnitud momento sísmico mayor a 7. Sin embargo, hasta el día de hoy no se conoce con exactitud la geometría de la zona de subducción y tampoco se entiende claramente la dinámica de subducción en términos de las propiedades mecánicas de la interfase y la distribución espacial del momento sísmico acumulado (zona sismogénica). Este proyecto tuvo como objetivo colaborar con el esfuerzo científico para caracterizar la zona de subducción en la región sur de Costa Rica. Para ello se utilizaron métodos de detección de eventos sísmicos, como la técnica de procesamiento de señales conocida como “template matching”. Esta técnica consiste en utilizar la señal de sismos previamente registrados (plantillas) por todas las estaciones sísmicas disponibles en una determinada ventana de tiempo para buscar en el registro continuo eventos con forma de onda similares que no hayan sido detectados por los sistemas automáticos debido a una baja relación señal/ruido o bien posibles errores en el procesamiento de las señales. Esta técnica tiene el potencial no sólo de incrementar el número de eventos en el catálogo sísmico actual, sino también de evidenciar procesos físicos a lo largo de la interfase, como por ejemplo: deslizamiento asísmico, enjambres sísmicos, tremores tectónicos, migración sísmica, entre otros. Posteriormente, se realizará una localización hipocentral de los eventos para estimar la geometría de la zona de

subducción, además de analizar la deformación de la corteza y la transferencia de carga utilizando un simulador de deformación elástica. Para todas las actividades mencionadas anteriormente la computación de alto rendimiento (HPC, por sus siglas en inglés) es un elemento catalizador para generar hallazgos transformadores. Este proyecto está basado en el uso de supercomputadoras y arquitecturas de computación masivamente paralela.

4. Palabras clave

Sismología computacional, computación de alto rendimiento.

5. Introducción

En los últimos 10 años Costa Rica, mediante la labor llevada a cabo en los principales observatorios sismológicos del país, ha experimentado un incremento en su capacidad para detectar, catalogar e interpretar la actividad sismotectónica que ocurre en la región, lo cual ha permitido un mejoramiento de los modelos actuales de prevención del riesgo sísmico y volcánico. Sin embargo, debido al creciente número de estaciones sismológicas utilizadas en el monitoreo continuo de la actividad tanto local como regional, resulta difícil el poder mantener un catálogo sísmico completo (que incluya un amplio rango de magnitudes) mediante los métodos tradicionales de monitoreo, los cuales están sujetos al error humano. Es fundamental para la sismología moderna implementar algoritmos que puedan procesar grandes cantidades de datos de forma consistente y le permitan a los investigadores hacer un análisis más expedito de los resultados.

Los terremotos más grandes y potencialmente destructivos del mundo ocurren en las zonas de subducción. Un ejemplo de ello es el pasado terremoto de Nicoya, Costa Rica, ocurrido el 5 de setiembre del 2012. La caracterización de los sismos (propiedades mecánicas, geometría, composición, etc) es fundamental para establecer programas de riesgo, desarrollo urbano y segmentación territorial para los habitantes de las áreas más cercanas a estas zonas sísmicamente activas. Observaciones recientes (Edwards et al., 2018) han demostrado que la interfase (el área de contacto entre la placa del Coco y la microplaca de Panamá) en la zona de subducción por debajo de la Península de Osa es bastante heterogénea, lo cual tiene implicaciones en el proceso de generación de terremotos, contenido espectral radiado, sacudida del suelo e intensidades percibidas en superficie. Contrario a los múltiples estudios realizados en la Península de Nicoya y en donde se conoce con muy buen detalle la geometría y dinámica del proceso de subducción de la placa del Coco por debajo de la placa Caribe, la geometría y dinámica que define la zona de subducción en el sur del país resulta menos clara. Aunque en la última década se han destinado esfuerzos para tratar de caracterizar esta región con mayor detalle, no ha sido suficiente debido al poco tiempo de registro de la nueva instrumentación en esa zona.

Utilizando los catálogos sísmicos actuales, la densa distribución de estaciones sismológicas y finalmente el creciente flujo de datos en tiempo real registrado localmente en el sur del país por parte del Observatorio Vulcanológico y Sismológico de Costa Rica (OVSICORI-UNA) en la Universidad Nacional, es teóricamente posible caracterizar la geometría del proceso de subducción por debajo de la Península de Osa al hacer una proyección hipocentral de los eventos localizados hasta la fecha. Esta metodología fue empleada en la zona del pacífico central, con muy buenos resultados. El visualizador Plinius, desarrollado en el TEC, es un ejemplo de las herramientas que se ajustan a dicha metodología.

No obstante, a diferencia del pacífico central (una de las zonas con mayor sismicidad de magnitud intermedia en el país), la zona sur tiene el potencial para generar sismos con magnitudes iguales o mayores a 7.0 (Protti et al., 2001). Los sismos de pequeña amplitud (M1-2.5) son parcialmente

enmascarados por el ruido sísmico de fondo, con un periodo fundamental de ~ 6 s. Sin embargo, es posible utilizar eventos sísmicos que presentan buena relación señal / ruido en múltiples estaciones como modelos o plantillas para hacer una búsqueda de formas de onda similares a lo largo de las series de tiempo registradas hasta la fecha y que han sido oscurecidos por el ruido sísmico de fondo o bien no han sido incluidos en los catálogos sísmicos actuales. Este procedimiento permite incrementar el tamaño los catálogos sísmicos hasta en un orden de magnitud. Como se señaló, crear una visualización de los hipocentros de dichos sismos permitiría definir con mayor certeza la geometría de la zona de subducción en el sur del país.

Aún cuando los observatorios sismológicos nacionales, el OVSICORI y la Red Sismológica Nacional (RSN), realizan una tarea sobresaliente de monitoreo, auscultación e investigación, en general, existe una enorme oportunidad en vincular esos esfuerzos con científicos computacionales del ITCR para acelerar nuevos descubrimientos en el área de ciencias de la tierra por medio de la computación de alto rendimiento. La plataforma de sistemas y programas desarrollados en este proyecto proveen una base sólida en la que se pueden construir desarrollos futuros para resolver otros problemas de geofísica, vulcanología y sismología.

6. Marco Teórico

Los procedimientos de análisis de datos y simulación en sismología requieren de gran cantidad de cálculos. Para obtener resultados en un tiempo aceptable, es necesario utilizar una plataforma de computación de alto rendimiento (ó HPC, por sus siglas en inglés) que permita procesar altos volúmenes de información en un tiempo corto. Afortunadamente, la Ley de Moore ha propiciado la construcción de procesadores paralelos con un creciente número de núcleos y por ende una mayor capacidad de cálculo (Hager, 2010). Además, la comunidad de HPC ha desarrollado un amplio ecosistema de herramientas de software, sistemas, arquitecturas y modelos de programación para la construcción de plataformas de computación de alto rendimiento. En particular, las tarjetas de procesamiento gráfico (ó GPU, por sus siglas en inglés) representan la punta de lanza en procesadores masivamente paralelos. Los GPUs han impulsado el reciente avance en machine learning y deep learning y que se manifiesta en bibliotecas como TensorFlow (Abadi, 2015), Caffe2 (Caffe2, 2018) y PyTorch (PyTorch, 2018). Estas bibliotecas subyacen la construcción de programas que efectúan reconocimiento de señales y otros patrones sobre GPUs.

La literatura de métodos de detección de eventos sísmicos es generosa, dado que la comunidad de sismología lleva décadas desarrollando estos métodos. Uno de los primeros algoritmos para la detección de sismos presentados en la literatura es STA/LTA (Allen, 1982). Esta técnica ha sido tradicionalmente utilizada por su capacidad de detección de sismos que presentan una alta razón señal a ruido. Sin embargo, no logra detectar todos los sismos y además genera falsos positivos en los casos en que se presenta una baja razón señal a ruido. Esta última condición aparece cuando ocurren eventos superpuestos o en el caso que no haya una densa distribución de estaciones sismológicas, entre otros casos. Lo anterior hace que esta técnica se considere de baja sensibilidad en cuanto a detección de eventos sísmicos (Yoon et. al, 2015).

Posteriormente, se han utilizado otras técnicas del procesamiento de señales como autocorrelación y *template matching*. La autocorrelación destaca por su mejora en la sensibilidad de detección respecto a STA/LTA y su capacidad para encontrar eventos no conocidos con forma de onda similar. Sin embargo, su alta complejidad computacional hace que este método no sea práctico para el análisis de bases de datos masivas (Yoon et. al, 2015). Template Matching se basa en el mismo principio de correlación cruzada, sin embargo utiliza un catálogo de plantillas previamente identificadas por expertos para identificar eventos similares en la señal. Este método presenta una buena sensibilidad en la detección de eventos con una menor complejidad computacional que la

autocorrelación, debido a que solo compara un catálogo definido de plantillas para la totalidad de la señal. Su principal dificultad es la dependencia en la extracción previa de plantillas, sumado a la baja generalidad del algoritmo para detectar eventos no conocidos (Yoon et. al, 2015).

Se han desarrollado *frameworks* y métodos más complejos que emplean las técnicas descritas anteriormente. Además, estos frameworks utilizan los métodos tradicionales en combinación con técnicas de reducción de dimensionalidad, extracción de características y clustering para reducir la complejidad computacional y mantener la sensibilidad en la detección de eventos. Algunos de estos frameworks son Repeating Signal Detector (RSD) propuesto en (Skoumal et. al, 2016), EQcorrscan propuesto en (Chamberlain et. al, 2018), Fingerprint And Similarity Thresh-olding (FAST) propuesta en (Yoon et. al, 2015) y PyMPA propuesto en (Vuan et. al, 2018).

RDS utiliza técnicas de procesamiento en el dominio temporal y en el dominio de la frecuencia incorporando el uso de la Transformada Rápida de Fourier (FFT) para la clasificación de los eventos en familias según su similitud, posteriormente realiza una clasificación por subfamilias por medio de una técnica de ordenamiento temporal (Skoumal et. al, 2016). Esto permite clasificar los eventos en familias de sismos y familias de ruido, que en este caso, se descartan. FAST es una técnica no supervisada que crea huellas de las formas de onda para extraer sus características y agruparlas por su similitud, alcanzando una aceleración de 140 veces respecto a la duración de técnica de autocorrelación paralela. Además, esta técnica permite la comparación a gran escala de las huellas extraídas de la forma de onda en una base de datos masiva y la detección de eventos no periódicos (Yoon et. al, 2015).

EQcorrscan está disponible en un paquete de código libre listo para su uso y se encuentra paralelizado a nivel de hilos. Se está trabajando en su adaptación a arquitecturas HPC con alto paralelismo y GPUs para mejorar el tiempo de procesamiento y superar barreras como las limitaciones de ancho de banda de memoria (Chamberlain et. al, 2018). El programa pyMPA emplea template matching y clustering para la clasificación y detección de eventos sísmicos e incorpora procesamiento multihilo para escalar el proceso de comparación de templates (Vuan et. al, 2018). Sumado a los frameworks mencionados, se han desarrollado técnicas supervisadas de aprendizaje de máquina como la presentada en (Reynen et. al, 2017) para la clasificación y detección de eventos sísmicos. En este trabajo, se alcanza un 98% en detección de eventos y la aproximación de su ubicación para eventos en el lapso de una semana utilizando métodos como regresión logística.

Gracias a la creciente popularidad de los algoritmos de aprendizaje de máquina basados en redes neuronales, se han presentado novedosos métodos de detección como ConvNetQuake (Perol et. al, 2018) en el que se reporta una precisión del 94.8% en la detección de sismos, cercana a la obtenida con el método de autocorrelación que alcanza el 100% de detección y 88% para el caso del método FAST. En cuanto al tiempo de ejecución, ConvNetQuake reduce considerablemente el tiempo de análisis de una señal con un tamaño de una semana a cerca de un minuto, en comparación con FAST que tarda unos 48 minutos, y el método de autocorrelación que tarda más de 9 días para procesar la misma cantidad de datos (Perol et. al, 2018).

Para detectar nuevos terremotos, seleccionar fases y comparar las capacidades de detección de terremotos de los algoritmos de la inteligencia artificial moderna con los de los métodos sismológicos tradicionales, nuestro flujo de trabajo incorporó el uso de EQTransformer (en adelante EQT) un nuevo modelo de aprendizaje profundo de código abierto diseñado para reconocimiento no supervisado y selección de fases de terremotos (Mousavi et al, 2020) a partir de formas de onda sísmicas diarias de tres componentes. El modelo fue entrenado originalmente usando el *Stanford Earthquake Dataset* (STEAD) (Mousavi et al, 2019), un conjunto de datos de entrenamiento distribuido que consta de 1,3 millones de observaciones de terremotos locales. La mayoría de estos eventos son más pequeños que M2.5 y han sido registrados dentro de distancias epicentrales de 100

km. EQT fue primero probado con la secuencia de réplicas del 6 de octubre de 2000, M6.7 Terremoto de Tottori, Japón (Mousavi et al, 2020). Describiremos brevemente el arquitectura de red de EQT a continuación. Durante la primera etapa, la estructura multitarea de EQT consiste en un codificador muy profundo que convierte las señales sísmicas del dominio del tiempo en representaciones de alto nivel de sus dependencias temporales mediante el uso convolucional y LSTM (long short-term memory) previos a la capas de auto atención. Los LSTM son un tipo de red neuronal recurrente (RNN) diseñado para manejar grandes secuencias. Su elemento principal es una celda de memoria que, dada una entrada, genera un estado oculto y actualiza la memoria a través de un mecanismo de puerta. Haciendo eso evita el problema de no aprender las dependencias a largo plazo (eso sucede en la práctica en los RNN regulares). Los bloques convolucionales explotan la estructura local y proporcionan mejores invariancia, mientras que el uso de capas LSTM ha demostrado ser importante para incorporar información posicional.

De las bibliotecas mencionadas anteriormente, únicamente EQT, FAST, EQCorrscan y ConvNetQuake utilizan computación de alto rendimiento o GPUs para la ejecución del algoritmo. En este trabajo es de interés la incorporación de HPC en los algoritmos de detección de eventos sísmicos con el fin de utilizar la plataforma de computación de alto rendimiento Kabré del Centro Nacional de Alta Tecnología (CeNAT), disponible para investigadores de CONARE.

A partir de la detección de nuevos eventos sísmicos y el consiguiente aumento del catálogo, se pueden utilizar programas de simulación de deformación elástica y visco-elástica, por ejemplo Gale 2.0.1 (Moresi et al., 2003; Moresi et al., 2012) y Plasti 1 (Fuller et al., 2006; Willett and Fuller, 2006), para estimar la deformación de la corteza y la transferencia de carga. Estas simulaciones permiten reconstruir la historia geológica de la zona y estimar la magnitud de futuros sismos. Más recientemente, la comunidad de geodinámica ha establecido una plataforma para crear y refinar programas computacionales en diferentes aristas de geofísica. Esta plataforma se denomina Computational Infrastructure for Geodynamics (CIG, 2018) y ofrece un índice con múltiples programas para cada tipo de problema en el área. En el caso de simuladores para la deformación de la corteza, la mayor referencia es PyLith (Aagaard, 2013). PyLith es un software portátil y escalable para la simulación de la deformación de la corteza en escalas espaciales que van desde metros hasta cientos de kilómetros y escalas temporales que van desde milisegundos a miles de años. Sus aplicaciones principales son modelado casi estático y dinámico de fallas sísmicas.

El visualizador Plinius, elaborado en el ITCR y entregado al OVSICORI (Guzmán, 2017), permite hacer una visualización tridimensional del catálogo de sismos y sobreponer diferentes capas de información. La versatilidad del visualizador permite analizar perfiles y cortes en tres dimensiones para estudiar fenómenos de origen volcánico y tectónico. Los datos que se obtengan en el actual proyecto podrán ser desplegados en Plinius. Si fuera necesario crear otro tipo de visualizaciones se podría utilizar ParaView (Ahrens, 2005) que representa una plataforma para el despliegue y análisis visual de datos científicos. ParaView es un visualizador general, pero al mismo tiempo poderoso para cualquier dominio científico. En particular, ParaView se utiliza para desplegar visualmente los resultados del simulador sismológico SPECFEM3D_Cartesian (Komatitsch, 2002).

7. Metodología

El objetivo específico de comparar las técnicas disponibles siguió una metodología muy general, utilizada en el proyecto *Framework para simulación en paralelo de eventos sísmicos y volcánicos*. Dicha metodología consiste en analizar el software disponible para realizar alguna tarea ya consolidada en el área e implementar un flujo de trabajo en la plataforma de computación de alto rendimiento Kabré del Centro Nacional de Alta Tecnología (CeNAT). Debido a las particularidades de la comunidad de sismología y vulcanología, esto no es un trabajo trivial. Mucho del software

generado no se publica ni se coloca en repositorios, sino que se comparte de manera social, a través de conexiones personales. además, es común que no se encuentre bien documentado y que su uso dependa del know how de los investigadores.

Dicha metodología consiste en los siguientes pasos:

1. Levantar una lista exhaustiva del software disponible para ejecutar la tarea.
2. Evaluar la disponibilidad, mantenibilidad y documentación de cada posible software.
3. Discutir con los especialistas del área los detalles técnicos del método numérico implementado y su conveniencia para el proyecto.
4. Seleccionar un software, considerando su idoneidad para la tarea y dificultad de uso.

Los objetivos específicos de establecimiento de flujos de trabajo requirieron el desarrollo o implantación de software, para lo cual se utilizará una metodología ágil. Las diferencias entre el desarrollo de software tradicional y el desarrollo de software científico han incrementado con el tiempo (Kelly, 2007). Los científicos usan primordialmente el software y los resultados de ejecuciones del software para hacer simulaciones o análisis complejos de datos. Inclusive, en algunas ocasiones el software es utilizado para corroborar una teoría científica. Estas características del software científico implican que, en contraste con el desarrollo de software administrativo o empresarial, los programadores de software científico no pueden determinar cuál debe ser el resultado correcto de una aplicación en el sentido tradicional. Esto plantea dos desafíos particulares desde el punto de vista de ingeniería de software. Primero, la especificación de requisitos y la especificación funcional son altamente dinámicos. Debido a la naturaleza exploratoria de muchos proyectos científicos, la obtención y especificación de los requisitos son problemáticos porque pueden ser inciertos o incluso desconocidos por adelantado. De hecho, en la medida en que se considere necesaria cualquier especificación, los requisitos se escriben cerca de la finalización del software. La actividad requerida es un problema recurrente en proyectos de software, especialmente cuando los equipos son grandes o cuando los científicos dedican mucho tiempo para desarrollar software. Segundo, la definición de casos de prueba para validación y la verificación del software es extremadamente desafiante. A menudo no es obvio estipular si un error se encuentra dentro de la teoría científica o en la implementación de esa misma teoría. Aunque estos aspectos del software científico impiden el manejo y las pruebas de los requisitos desde el principio, la falta de conocimiento sobre los requisitos y principios de prueba y la falta de las actividades organizadas se han identificado como un área problemática en varios estudios (Sletholt, 2010). En los últimos años, el desarrollo de software científico ha incorporado metodologías no tradicionales, particularmente metodologías ágiles (Storer, 2017).

El término metodologías ágiles para el desarrollo de software consiste en una sombrilla que abarca varios métodos y prácticas que satisfacen el manifiesto ágil. En particular, las soluciones se forman en un esfuerzo colaborativo de equipos auto organizados, multifuncionales y en contacto con sus clientes. El manifiesto ágil sostiene que todo desarrollo de software (científico en nuestro caso) debe valorar: a) individuos e interacciones sobre procesos y herramientas, b) software funcional sobre documentación exhaustiva, c) colaboración con el cliente sobre negociación del contrato y d) respuesta al cambio sobre un plan riguroso. Una de las metodologías ágiles más populares es Scrum (Schwaber, 2018), donde se adopta una estrategia de desarrollo incremental, la calidad se basa en el conocimiento de las personas involucradas y diferentes fases del proyecto se traslapan. Esto implica que otros objetivos tradicionales pierden protagonismo, como planear y ejecutar completamente el producto, asumir que los procesos generan calidad y que las fases siguen un ciclo secuencial o en cascada. En la metodología scrum, el scrum master facilita la aplicación de la metodología, mientras que el product owner representa a los interesados en el producto y el team es el grupo de desarrolladores. Los avances están impulsados por sprints, que pueden durar entre una y cuatro semanas y cuyo objetivo es incrementar el software para agregar funcionalidad y

obtener un entregable. Un product backlog es una lista priorizada de necesidades que se deben incorporar al software. Reuniones periódicas determinan cuáles objetivos se alcanzarán en los diferentes sprints. Scrum es una metodología flexible que reconoce que los clientes pueden cambiar de necesidades durante la ejecución del proyecto. Scrum provee una gestión regular de las expectativas del cliente, resultados anticipados, adaptividad, mitigación de riesgos, productividad y calidad, alineamiento entre el team y el cliente, entre otros.

El objetivo específico de generar la geometría de la zona de subducción está relacionado con la creación de una visualización de los nuevos sismos encontrados. A partir de la tecnología de visualización ya disponible, se incluyeron nuevas características a sugerencia de los especialistas del área y se evaluó la utilidad científica y divulgativa del resultado. Este procedimiento se realizó iterativamente hasta conseguir una visualización que contenga todos los elementos significativos para la interpretación física de la geometría de la zona de subducción y la transferencia de carga.

En el caso específico de los componentes para la visualización científica, aplican también los factores mencionados anteriormente con respecto al desarrollo de software. Sin embargo, existen algunos puntos de partida y lineamientos que se pueden seguir dependiendo de ciertas variables, para mencionar las más importantes:

1. El tipo de datos: aquí es importante observar qué información está contenida en los datos a visualizar. Por ejemplo, diferentes técnicas de interacción se necesitarán para visualizar datos dinámicos (que cambian con el tiempo), como en el caso de deformaciones.
2. Formato: la técnica usada para la creación de los datos también influirá en la decisión de qué técnicas de visualización se pueden utilizar. Para imágenes obtenidas de tomografías por ejemplo, se usa generalmente visualización de volúmenes (volume rendering), ya que permite la presentación de estructuras internas. En otros casos, técnicas de visualización de superficies son requeridas, por ejemplo en aplicaciones arquitectónicas.
3. El volumen de datos: la cantidad de datos a presentar influirá también en la elección de técnicas de visualización. Por lo general, entre más datos se muestren simultáneamente, incrementará el tiempo que se requiere para generar cada imagen. Es por esta razón que en el caso de visualizaciones en tiempo real de grandes volúmenes de datos, se opta por técnicas LOD (Level of Detail o Nivel de detalle), en las que primero se muestra la imagen con poco detalle y esta se va mejorando poco a poco, o cuando el usuario así lo requiere.
4. La función: uno de los factores más importantes que determina la técnica de visualización a utilizar es qué tipo de información es la que se quiere obtener a partir las visualizaciones. En el caso de visualizaciones científicas, este también es uno de los factores más complejos, ya que en muchas ocasiones no es posible determinarlo con antelación. Es por esta razón que el proceso de creación de visualizaciones científicas es un proceso iterativo, en el que, primero se desarrollan visualizaciones simples que con la participación activa de los científicos, se van mejorando hasta obtener el producto esperado.

Una vez que se definieron las características tanto de los datos de entrada, como de las herramientas de visualización esperadas, fue posible determinar las técnicas de visualización requeridas. Se partió entonces de soluciones conocidas en el campo para extender, si es necesario, las herramientas de visualización creadas en el proyecto *Framework para simulación en paralelo de eventos sísmicos y volcánicos*, y se adaptaron a las necesidades del actual proyecto de manera iterativa.

8. Resultados

Se implementa un mecanismo de análisis de forma jerárquica, donde un transformador realiza un análisis a lo largo de la secuencia completa, que es la entrada directa para el decodificador del predictor de detección. El uso de transformadores en redes convolucionales junto con un mecanismo de auto-atención interna para impulsar la transducción de secuencias y el aprendizaje de dependencias a largo plazo es más rápido que las recurrentes regulares o arquitecturas convolucionales. La salida del mecanismo global de atención es también el insumo para la los bloques de atención local en los submodelos de *picking*. Estos bloques de atención local examinar pequeñas porciones de la forma de onda e incorporar el información obtenida por la etapa de atención global. La información resultante es luego utilizada por tres decodificadores para mapear tres secuencias de probabilidades asociadas con: 1) la existencia de una señal de terremoto (discriminando terremotos del ruido de fondo, antropogénico o fuentes atmosféricas), 2) la probabilidad de predicción de la fase P (la probabilidad de que la fase elegida corresponda a la llegada tiempo de la onda P) y 3) la probabilidad de predicción de la fase S (la probabilidad de que la fase elegida corresponda a la hora de llegada de la onda S).

Para evaluar la calidad de las probabilidades de predicción, los decodificadores también proporcionan como salida variaciones estadísticas basadas en inferencia bayesiana. La estructura de la red de nuestro flujo de trabajo se muestra en la Fig.1. Como entrada, utilizamos formas de onda de tres componentes de banda ancha de 24 horas de 10 estaciones sísmicas ubicadas en la región de estudio y operado por OVSICORI-UNA. La etapa de transferencia crea una base de datos de formas de onda para el grupo seleccionado de estaciones y período de análisis. Luego, en DB2EQT, las grabaciones sísmicas en la base de datos se convierten al formato de archivo aceptado por EQT. Una vez que la base de datos está en el formato adecuado, EQT DETECT realiza la detección de eventos y selección de fase haciendo un pase hacia adelante sobre un modelo entrenado con EQT. Para este propósito, EQT proporciona 2 modelos que fueron entrenados con diferentes conjuntos de datos. Seleccionamos el modelo de formación más actualizado que incorpora la mayor cantidad de observaciones de terremotos en todo el mundo con magnitudes por debajo de M2.5 para minimizar el tasa de falsos positivos de las predicciones. El principal resultado de esta etapa consiste en una predicción base de datos para cada estación utilizada en el análisis. La base de datos obtenida incluye las horas de origen y la duración del evento para todos los terremotos detectados por el modelo, junto con su correspondiente probabilidad de detección y variación estadística. Los tiempos de llegada de las ondas P y S, y su predicción asociada probabilidades, también se incluye. En esta etapa desarrollamos y aplicamos el control de calidad (QC) con rutinas que incluyen el monitoreo del dominio del tiempo formas de onda para todos los eventos detectados en todas las estaciones utilizadas, y diagramas de tiempo de viaje diferencial de Wadati para evaluar P y evitar las detecciones falsas. Finalmente, EQT ASSOCIATE usa los resultados de la predicción y red sísmica para realizar la asociación de eventos, eliminando valores atípicos y la construcción del catálogo sísmico final para el periodo de estudio. La visualización y control de calidad de la sísmica obtenida en el catálogo se realizan en la etapa de POST-PROCESAMIENTO.

Se implementó todo el producto de detección sísmica en la Supercomputadora Kabré en el Centro Nacional de Alta Tecnología (CeNAT). Kabré es un cluster heterogéneo de computación con diferentes secciones para cada aplicación. La sección de inteligencia artificial contiene nodos *nukwa*, cada uno equipado con una unidad de procesamiento gráfico (GPU), con dos configuraciones diferentes. Las primera configuración tiene como características de configuración un procesador Intel Xeon E3-1225 v5 con 4 núcleos que funcionan a 3,30 GHz, 16 GB de memoria principal y un GPU NVIDIA Tesla K40c con 12 GB de memoria. En la segunda configuración se presentan nodos cada uno con un Intel Xeon Silver Procesador 4214R con 24 núcleos funcionando a

2,40 GHz, 32 GB de memoria principal y una GPU NVIDIA Tesla V100 con 32 GB de memoria. Se utilizaron varias herramientas computacionales para construir la infraestructura de detección. Los scripts de Python se escribieron para la transformación de datos durante las etapas de pre y / o post-procesamiento.

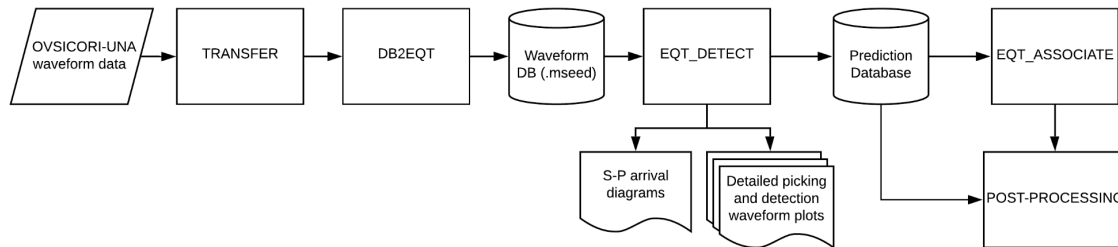


Fig 1. Flujo de trabajo de OKSP

Exploramos las capacidades de detección y selección de fase de EQT a diferentes umbrales de probabilidad durante nuestro período de estudio (Figura 2). Por debajo del intervalo de confianza del 80%, EQT está contaminado por un número creciente de detecciones de "falsos positivos", debido a la mejora del ruido ambiental y/o antropogénico (una reducción en la relación señal / ruido), como se muestra en la formas de onda verticales para todas las estaciones utilizadas en este trabajo (Fig. 2b-2f). Como consecuencia, no podemos separar tasas premonitadoras de sismicidad de fondo de la secuencia observada de réplicas generada por el terremoto de M6.5. Los umbrales de probabilidad más altos (80%) exhibieron una mejora significativa en la relación S / N y destacó un aumento acelerado en el número de terremotos inmediatamente después la ocurrencia del sismo principal, como predijo Gutenberg y Richter y se encuentran en estudios previos en California e Italia utilizando algoritmos de IA similares. Como estamos analizando la primera etapa de la secuencia del terremoto de Puerto Armuelles, nuestros resultados no muestran una desintegración de Omori en el número de réplicas. Basado en la frecuencia de tiempo comportamiento de la ocurrencia del terremoto (Fig. 2a) y relaciones S / N exhibido en las formas de onda con diferentes probabilidades de detección, seleccionamos el 80% como nuestro límite de detección más bajo para expandir el catálogo de terremotos de OVSICORI-UNA para la secuencia del terremoto de Puerto Armuelles. La diferencia por día en el número de terremotos entre EQT y OVSICORI-UNA, para diferentes probabilidades de detección es interesante. En general, encontramos que para los 6 días del análisis y las probabilidades de detección entre el 80% y el 90%, EQT superó a las capacidades tradicionales de detección desarrolladas por OVSICORI-UNA, y recuperó un total de 1100 terremotos más que no fueron detectado y / o localizado previamente por enfoques tradicionales.

Además, en el umbral del 90%, EQT se reprodujo con increíble detalle que generó el catálogo mainshock-aftershock por OVSICORI (mostrado como una línea negra gruesa en la Fig. 2a) y expandió en 28 el número de nuevas detecciones que anticiparon la aparición del sismo principal (posibles sismos previos). Se extrajo y revisó cuidadosamente las formas de onda en el dominio del tiempo de todas las detecciones previas al choque principal con una probabilidad del 90% umbral y se encontró que 23 de los 28 eventos detectados por EQT corresponden con los verdaderos temblores que se produjeron a partir de 25 de junio hasta horas antes del sismo principal. Estos terremotos no fueron detectados ni localizados por OVSICORI-UNA. Para este período, las detecciones perdidas de eventos ya presentes no se encontraron en los catálogos del OVSICORI (es decir, falsos negativos). Con base en esta información, las métricas de clasificación para el modelo se calcularon. Encontramos que para esta muestra el modelo era 100% exhaustivo y 82% preciso, lo que da una puntuación F1 de 0,90. Estos resultados son prometedores para una implementación

preliminar y evaluación de nuestro *pipeline*, ya que demuestran que es posible ampliar el catálogo sísmico OVSICORI-UNA de forma robusta, una vez que se establece un umbral de probabilidad adecuado para el modelo. Calculamos la magnitud de la secuencia premonitoria a través de medidas de amplitud relativa y encontramos una magnitud promedio de 0.8 ± 0.3 . Los terremotos con magnitudes y amplitudes similares son generalmente oscurecidos en la serie de tiempo por ruido de alta frecuencia de fuentes naturales (por ejemplo, viento y mareas oceánicas) y/o fuentes antropogénicas (por ejemplo, difusas, como carreteras, ferrocarriles, etc., y armónico de maquinaria industrial). Por lo tanto, estos terremotos de pequeña magnitud son extremadamente importantes para mejorar nuestra comprensión de la nucleación, ruptura de terremotos, evolución y física de fallas durante el ciclo del terremoto.

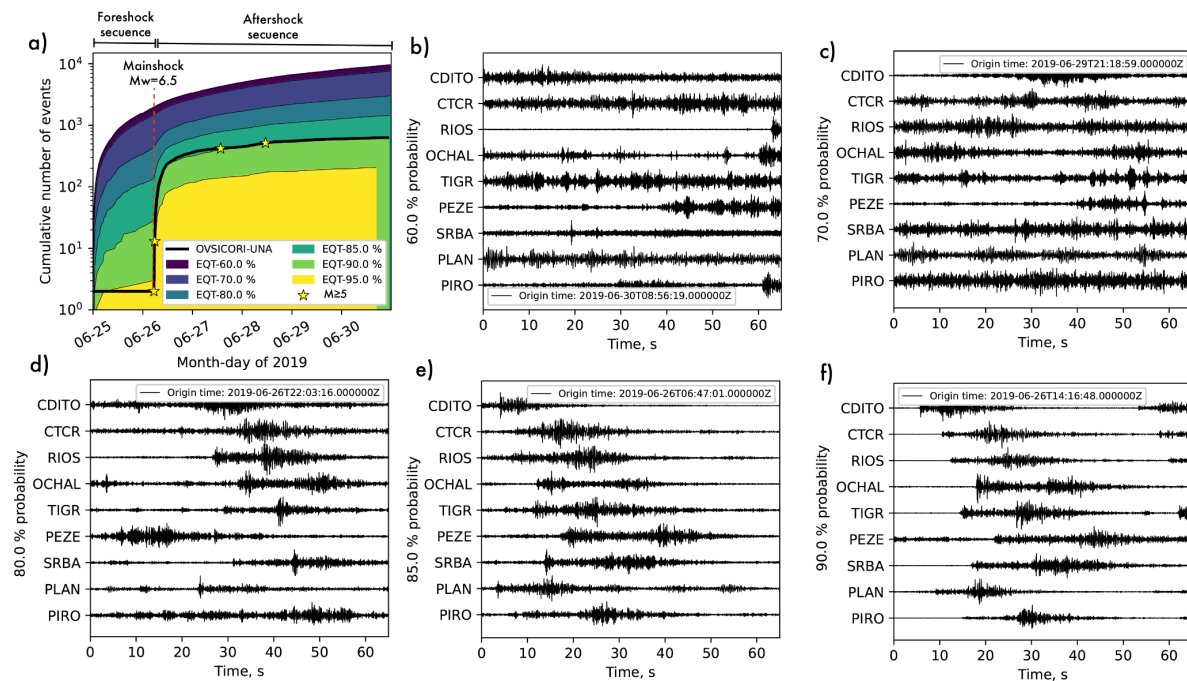


Fig 2. Figura que muestra el rendimiento de detección sísmica EQT frente al catálogo sísmico OVSICORI-UNA del 25 al 30 de junio de 2019. El panel a) muestra el número acumulado (en escala logarítmica) de terremotos detectados por EQT en función del tiempo para la secuencia de terremotos de Puerto Armuelles, codificado por colores por el umbral de probabilidad de detección. La línea sólida negra representa el número acumulado de eventos localizados por OVSICORI-UNA para el mismo período de estudio y región. Las estrellas amarillas en la parte superior de la línea negra resaltan la ocurrencia de eventos con magnitudes 5.0. Los paneles b) a f) muestran ejemplos de grabación del componente vertical para las estaciones que se muestran en la Fig.2, extraído 20 s antes y 70 s después del tiempo de origen detectado y el paso de banda filtrado entre 1 y 15 Hz. La figura muestra eventos seleccionados al azar para cada uno de los umbrales de probabilidad de detección

9. Discusión y conclusiones

Este documento presentó OKSP, un nuevo flujo de trabajo basado en algoritmos de aprendizaje profundo para la detección automática de terremotos en Costa Rica. Nuestros resultados muestran que nuestra herramienta es capaz de procesar eficientemente datos de forma de onda del OVSICORI-UNA y realizar con precisión la detección de eventos y la fase *picking* con un modelo previamente entrenado en EQTransformer (EQT). Para nuestro caso de estudio, el día anterior y los primeros 5 días siguientes el terremoto de Puerto Armuelles M6.5 ocurrido en junio el 26 de febrero

de 2019, OKSP encontró 1100 terremotos anteriormente no identificados por OVSICORI-UNA. Los nuevos datos constituyen una secuencia robusta que permite comprender mejor los procesos premonitorios dinámicos en la región.

Nuestros resultados experimentales demuestran que los algoritmos de IA, junto con infraestructuras informáticas de alta gama y una amplia bases de datos, proporcionan un marco de vanguardia que puede mejorar el seguimiento y la comprensión de los terremotos. OKSP es el primer producto de esta naturaleza que se ha propuesto para la detección de terremotos en Costa Rica. OKSP expande la corriente capacidades de detección, reduce los tiempos de procesamiento y acelera búsquedas científicas en sismología. Vemos un futuro prometedor para OKSP en varias dimensiones. Primero, el producto propuesto tiene el potencial de acelerar la respuesta científica en observación y monitoreo sísmico, ya que nuestro objetivo es que sea capaz de procesar formas de onda en tiempo real datos de todas las estaciones sísmicas disponibles en Costa Rica. En segundo lugar, también buscamos la implementación de capacidades para el procesamiento masivo de datos, lo que ayudará a examinar el extensa información sísmica registrada por OVSICORI-UNA durante las ultimas décadas. Los conocimientos que se obtendrán de la expansión del catálogo sísmico utilizando metodologías novedosas, como el que propusimos, contribuirá a la precisión en la comunicación y difusión de información relacionada con procesos geodinámicos de sistemas de fallas. En tercer lugar, reconocemos la oportunidad ilimitada de explorar diferentes profundidades modelos de aprendizaje para mejorar la detección y caracterización de una amplia variedad de fenómenos sísmicos.

10. Recomendaciones

El trabajo interdisciplinar conlleva automáticamente riesgos en los proyectos de investigación. Por una parte, hay que conciliar diferencias en las culturas de los grupos de investigación (organización, vocabulario, metodologías de trabajo). Por otra parte, usualmente hay una dependencia entre las actividades de uno y otro grupo. Sin embargo, todas esos riesgos son eclipsados por la enorme oportunidad de desarrollar proyectos con un alto impacto. El conjunto es mucho más que la suma de las partes.

La computación como herramienta se ha establecido con un mecanismo fundamental en dos pilares científicos. Primero, la simulación computacional ofrece una contraparte al desarrollo teórico y a la recolección de datos experimentales. Segundo, el análisis computacional de datos (grandes o pequeños) permite dilucidar patrones ocultos a simple vista. Esto último ha sido potenciado por la reciente ola de inteligencia artificial y el uso de redes neuronales profundas.

11. Referencias

Aagaard, B.T.; Knepley, M.G.; Williams, C.A. (2013), A domain decomposition approach to implementing fault slip in finite-element models of quasi-static and dynamic crustal deformation, *Journal of Geophysical Research: Solid Earth*, 118 (6) , 3059-3079.

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg,

- Martin Wicke, Yuan Yu y Xiaoqiang Zheng (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. White paper disponible en www.tensorflow.org.
- Ahrens, James, Geveci, Berk, Law, Charles (2005). *ParaView: An End-User Tool for Large Data Visualization*, Visualization Handbook, Elsevier.
- Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6), S225-242. <https://doi.org/10.1130/SPE262-p173>
- Arroyo, Ivonne and Guillermo, Alvarado and R. Flueh, E. (2003). *Local Seismicity at the Cocos Ridge - Osa Peninsula Subduction Zone, Costa Rica*. AGU Fall Meeting Abstracts
- Arroyo H., Ivonne G. (2005). *Sismicidad y neotectonica en la region de influencia del Proyecto Boruca: hacia una mejor definicion sismogenica del Sureste de Costa Rica*. *Revista Geológica de América Central*, 32, p. 71.
- Artavia, Silvia. (2018). *Científicos 'descenderán' a las profundidades para mapear los temblores en Costa Rica*. Periódico digital La Nación. Accedido 12 de febrero, 2018.
- Caffe2, *A New Lightweight, Modular, and Scalable Deep Learning Framework* (2018). <https://caffe2.ai/>. Accedido 8 de junio, 2018
- Centro Nacional de Alta Tecnología (2018). Supercomputadora Kabré. <http://cluster.cenat.ac.cr/>. Accedido 8 de junio, 2018.
- Chamberlain, C. J., Hopp, C. J., Boese, C. M., Warren-Smith, E., Chambers, D., Chu, S. X., ... Townend, J. (2017). EQcorrscan: Repeating and Near-Repeating Earthquake Detection and Analysis in Python. *Seismological Research Letters*. <https://doi.org/10.1785/0220170151>
- CIG: Computational Infrastructure for Geodynamics (2018). <https://geodynamics.org/cig/software/pylith/>. Accedido 8 de junio, 2018.
- Igel, Heiner. (2017). *Computational Seismology: A Practical Introduction*. Oxford: Oxford University Press.
- Edwards, J. et al (2018). *Corrugated megathrust revealed offshore from Costa Rica*. *Nature Geoscience*.
- Fuller, C.W.; Willett, S.D.; Brandon, M.T. (2006), *Formation of forearc basins and their influence on subduction zone earthquakes*, *Geology*, 34 (2) , 65-68
- Gooch, A., Gooch, B., Shirley, P., & Cohen, E. (1998, July). A non-photorealistic lighting model for automatic technical illustration. In Proceedings of the 25th annual conference on Computer graphics and interactive techniques (pp. 447-452). ACM.
- Guzmán, Marcela (2017). *Visualizador de datos sísmicos PLINIUS*. *Revista Investiga TEC*. <http://revistas.tec.ac.cr/investigacion/?q=node/294> Accedido 9 de junio, 2018.
- Hager, Georg, Gerhard Wellein (2010), *Introduction to High Performance Computing for Scientists and Engineers*. CRC Press.

- Kelly, Diane F. (2007). *A Software Chasm: Software Engineering and Scientific Computing*. IEEE Software 24, 6 (November 2007), 120-119.
- Komatitsch, D., Tromp, J. (2002), Spectral-element simulations of global seismic wave propagation-I. Validation, *Geophysical Journal International*, 149 (2) , 390-412.
- Lücke, Oscar and Arroyo, Ivonne. (2015). *Density structure and geometry of the Costa Rican subduction zone from 3-D gravity modeling and local earthquake data*. Solid Earth.
- Moresi, L.; Dufour, F.; Mäuhlhaus, H.B. (2003), *A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials*, *Journal of Computational Physics*, 184 (2) , 476-497
- S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza (2019), *Stanford earthquake dataset (stead): A global data set of seismic signals for ai*, *IEEE Access*, vol. 7, pp. 179 464–179 476, 2019.
- S. Mousavi, W. Ellsworth, W. Zhu, L. Chuang, and G. Beroza (2020), *Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking*. *Nature Communications*, vol. 11, 2020.
- Protti, M. Güendel, F. y Malavassi, E. (2001), *Evaluación del Potencial Sísmico de la Península de Nicoya*; Editorial Fundacion UNA, 1ra. edición, Heredia, Costa Rica, 2001, 144 p.
- PyTorch, *Tensors and Dynamic neural networks in Python with strong GPU acceleration* (2018). <https://pytorch.org/>. Accedido 8 de junio, 2018.
- Reynen, A., & Audet, P. (2017). Supervised machine learning on a network scale: Application to seismic event classification and detection. *Geophysical Journal International*, 210(3), 1394–1409. <https://doi.org/10.1093/gji/ggx238>
- Rivi, M., Calori, L., Muscianisi, G., & Slavnic, V. (2012). *In-situ visualization: State-of-the-art and some use cases*. PRACE White Paper, 1-18
- Shearer, P.M. (2009). *Introduction to Seismology*. New York: Cambridge University Press.
- Sletholt, Magnus, Jo Erskine Hannay, Hans Petter Langtangen, Dietmar Pfahl (2010). *Making Modern Scientific Software Development Explicitly Agile*. The Changing Dynamics of Scientific Collaborations.
- Schwaber, Ken y Sutherland, Jeff. *The Scrum Guide*: <https://www.scrum.org/resources/scrum-guide> Accedido 3 de junio, 2018
- Shi, S., & Hsu, C. H. (2015). A survey of interactive remote rendering systems. *ACM Computing Surveys (CSUR)*, 47(4), 57
- Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2016). An efficient repeating signal detector to investigate earthquake swarms. *Journal of Geophysical Research: Solid Earth*, 121(8), 5880–5897.
- Storer, Tim (2017). *Bridging the Chasm: A Survey of Software Engineering Practice in Scientific Programming*. *ACM Computing Surveys (CSUR)*, 50 (4).
- Vuan, A., Sukan, M., Amati, G., & Kato, A. (2018). Improving the detection of low-magnitude seismicity preceding the Mw6.3 L'Aquila earthquake: Development of a scalable code based on the

cross correlation of template earthquakes. *Bulletin of the Seismological Society of America*, 108(1), 471–480. <https://doi.org/10.1785/0120170106>

Yoon, Clara E. and O'Reilly, Ossian and Bergen, Karianne J. and Beroza, Gregory C. (2015). *Earthquake detection through computationally efficient similarity search*. *Science Advances*, 1(11).

12. Anexo

Se adjunta el correo de recepción el artículo enviado para una publicación en una conferencia internacional indexada en SCOPUS. El texto del artículo se encuentra disponible en el repositorio arXiv.org:

<http://arxiv.org/abs/2109.02723>

BIP 2021 submission 2

BIP 2021 <bip2021@easychair.org>

Mon 8/30/2021 3:53 PM

To: Esteban Meneses <esteban.meneses@acm.org>

Dear authors,

We received your submission to BIP 2021 (2021 International Conference on Bioinspired Processing):

Authors : Leonardo van der Laat, Ronald J.L. Baldares, Esteban J. Chaves and Esteban Meneses
Title : OKSP: A Novel Deep Learning Automatic Event Detection Pipeline for Seismic Monitoring in Costa Rica
Number : 2

The submission was uploaded by Ronald Loaiza Baldares <rloaizab@cenat.ac.cr>. You can access it via the BIP 2021 EasyChair Web page

<https://easychair.org/conferences/?conf=bip2021>

Thank you for submitting to BIP 2021.

Best regards,
EasyChair for BIP 2021.