

# A Pearson Correlation Analysis of the Software Engineering Practice in Micro and Small-Sized Software Industry of Sinaloa, Mexico

J. Aguilar-Calderón, A. Zaldívar-Colado, C. Tripp-Barba, R. Espinosa-Oliva, and C. Zurita-Cruz

**Abstract**—Is well-known that software development process for desktop, web or mobile applications in industry needs the adoption of best practices of Software Engineering. Nevertheless, not much contemporary information with regard to current practices in micro and small-sized software industry exists especially in Sinaloa, Mexico. This work presents an exploratory study which provides insight into industrial practices in the software industry of Sinaloa. A combination of both qualitative and quantitative data is collected, using semi-structured interviews and a detailed questionnaire from sixteen software factories. A Pearson correlation analysis was performed independently between the variables Company location, Scope of coverage, Number of workers, Time to live in the market, Projects completed, Time dedicated to activities related to the project, Outdated projects completed in order to determine the degree of relationship between each of the variables mentioned, with all. A correlation analysis and an analysis of variance were performed. The quantitative results will serve to obtain opportunities for further interpretation and comparison.

**Index Terms**—ANOVA, Micro and small-sized software factories, Pearson correlation analysis, Software engineering.

## I. INTRODUCCIÓN

LA Ingeniería de Software (SE, por sus siglas en inglés) es la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento del software [2]. Con la aplicación de la SE, es posible reducir los riesgos de fallas en el desarrollo e incrementar la posibilidad de entrega del producto en el tiempo estimado, con calidad y dentro de los costos presupuestados. En la literatura sobre la SE es posible encontrar diferentes paradigmas a utilizar para desarrollar software que, si bien es cierto que incluyen las fases genéricas de requisitos, diseño, construcción, pruebas, mantenimiento y gestión de la configuración [2], su uso para un equipo de desarrollo puede representar un problema por la falta de comprensión en su aplicación y en la poca relevancia

This work was supported by the PROFAPI/2016/002 project from Universidad Autónoma de Sinaloa (México) through the members of the Cuerpo Académico en Consolidación Señales y Sistemas (CASESIS-UAS-169) and Universidad Tecnológica de Chile INACAP.

J.A. Aguilar-Calderón, Universidad Autónoma de Sinaloa, Mazatlán, México (e-mail: ja.aguilar@uas.edu.mx).

A. Zaldívar-Colado, Universidad Autónoma de Sinaloa, Mazatlán, México (e-mail: azaldivar@uas.edu.mx).

C. Tripp-Barba, Universidad Autónoma de Sinaloa, Mazatlán, México (e-mail: ctrip@uas.edu.mx).

R. Espinosa-Oliva, Universidad Tecnológica de Chile INACAP, Chile (e-mail: respinosa@inacap.cl).

C.E. Zurita-Cruz, Universidad Autónoma de Sinaloa (UAS), Mazatlán, Sinaloa, México (e-mail: czurita@uas.edu.mx).

de su práctica cotidiana en los proyectos de desarrollo de software.

Por otra parte, la implementación y aplicación de la SE en la industria ha mejorado debido al surgimiento de modelos de madurez tales como MoPROSOFT (Modelo de Procesos para la Industria del Software) [3] y CMMI (*Capability Maturity Model Integration*) [4]. De acuerdo con el *Process Maturity Profile* [5], publicado por el *Software Engineering Institute* (SEI) cada seis meses, ambos modelos tienen como objetivo la mejora del proceso. El informe se basa en los resultados enviados al SEI para la adopción de CMMI, entre los resultados se muestra un incremento en el número de evaluaciones (llamadas *appraisals*, una actividad que ayuda a identificar las fortalezas y debilidades de los procesos en una organización). Además, destaca el aumento de nivel de madurez en un 16% en empresas que tienen 100 o menos empleados, principalmente en China, Estados Unidos, India y México. A nivel continental, Asia ocupa el primer lugar con 9,629 evaluaciones, Norteamérica es segunda con 3,932 y Europa es el tercer lugar con 1,435 todo lo anterior en el período de 2008 a 2017. Los datos mencionados son un indicador de que la industria del software evoluciona constantemente.

Aún cuando México aparece en el *Process Maturity Profile* como uno de los países en los que el desarrollo de software basado en modelos de madurez ha incrementado y mejorado (debido a la obtención de niveles de madurez en la escala *CMMI for Development*), no implica que no se deba impulsar aún más la aplicación de la práctica de la SE en empresas nacionales para continuar con esa mejora, principalmente en aquellas de reciente creación. Para lograr esto, es esencial comprender el estado actual de su ejercicio profesional [6] y compararlo con el campo académico [7] lo que permitirá impulsar esa colaboración y a su vez promover el mejoramiento del proceso de desarrollo de software. En este sentido, dada la importancia de la colaboración entre la industria y la academia, en [8] se proporciona una síntesis de los desafíos y las mejores prácticas que los investigadores y profesionales pueden utilizar para tomar decisiones informadas sobre cómo estructurar sus colaboraciones, considerando realizar talleres y seminarios regulares, asegurar el aprendizaje continuo, los fundamentos académicos y la investigación básica sobre problemas del mundo real.

No obstante los esfuerzos derivados de la investigación en universidades y en la industria del software para colocar la SE en práctica en micro, pequeñas y medianas empresas (MiPyMES), todavía existen problemas importantes estableci-

dos en la literatura científica relacionada con esta área que evitan mejorar la calidad del producto (software). Algunos de estos aspectos se mencionan en el estudio presentado en [9], sin embargo, cabe destacar que estos no son aspectos exclusivos que perjudican a MiPyMES, si no a la industria del *software* en general. De acuerdo con el *Chaos Manifesto* del *Standish Group* solo el 36% de los proyectos se concluyen satisfactoriamente mientras que el 19% fracasan y el 45% sufre cambios. Entre las causas de fallas más frecuentes se revela la inexistencia de un equipo de trabajo correctamente capacitado, no se incentiva al equipo y no se provee un programa continuo de capacitación [10]. También se encuentran cambios en los requisitos y en el alcance del proyecto, poca o nula participación del usuario y la falta de habilidades en la gestión de proyectos [11].

En Sinaloa (México) existen un número reducido de empresas dedicadas al desarrollo de software a la medida con capacidad para competir internacionalmente, esto se debe en parte a la falta de cumplimiento de estándares de calidad y certificaciones internacionales [12]. En el Índice de Potencial de Innovación Estatal 2018 [13], ocupa la posición número diez de treinta y dos estados debido a tres indicadores: disminución de patentes solicitadas por cada millón de habitantes, reducción del número de empresas certificadas y crecimiento del PIB (Producto Interno Bruto) del sector servicios en menor proporción que el resto de los estados. A pesar de que el estado ocupa el segundo lugar en emprendimiento y negocios en el Índice de Potencial de Innovación Estatal 2018, la industria no está explotando la ventaja competitiva con la que cuenta frente a otros estados, como lo es la ubicación geográfica con respecto a un consumidor de servicios y software como los Estados Unidos de Norteamérica. Incluso, al tener un mismo huso horario se facilita la comunicación directa con los clientes, factor crítico para la eficiencia de ese sector [12]. Aunque es cierto que la tecnología hoy en día permite tener buena comunicación con el cliente independientemente de su ubicación geográfica, el no contar con el mismo huso horario obliga a las empresas a desarrollar estrategias con sus empleados para mantener la comunicación con el cliente en óptimas condiciones, por ejemplo, solicitarle a un miembro del equipo de desarrollo permanecer en su puesto de trabajo fuera del horario de oficina, entre otras.

El gobierno del estado de Sinaloa, la Universidad Autónoma de Sinaloa (UAS) y las empresas locales que desarrollan software han constituido recientemente un tridente gobierno-universidad-empresas, con la idea de promover la industria. Este tridente funciona de la siguiente manera: las universidades producen recurso humano capacitado; el gobierno traerá soporte financiero a las empresas para certificaciones profesionales; y a su vez éstas contratarán egresados certificados en tecnologías que demanda la iniciativa privada (por ejemplo, REACT, biblioteca Javascript de código abierto diseñada para crear interfaces de usuario con el objetivo de facilitar el desarrollo de aplicaciones en una sola página). La idea en particular es muy atractiva, existen experiencias similares con éxito, como en la ciudad de Monterrey, Nuevo León, lugar en donde la industria creció significativamente [14]. Para lograr estos resultados, es necesario llevar a cabo

proyectos de investigación que permitan obtener el estado actual de las empresas que desarrollan software en términos de su proceso de desarrollo y su nivel de madurez para que estén disponibles para desplegar soluciones de calidad, pues no podrán competir internacionalmente si no tienen un método bien definido y un proceso de desarrollo que permita obtener un producto de calidad. Para lograr esto, sus actores necesitan estar convencidos de que las mejores prácticas de la SE deben ser aplicadas.

Teniendo en cuenta estas consideraciones, el objetivo de este documento es presentar el estado actual del proceso de desarrollo de software de las MiPyMES del estado de Sinaloa, a través del nivel de afectación en el producto derivado de la ubicación geográfica de la empresa, su ámbito de cobertura (si es estatal, nacional o internacional), el número de trabajadores con el que cuenta, el tiempo de vida en el mercado (antigüedad), los proyectos concluidos, el tiempo que han dedicado a las actividades de ingeniería de requisitos del proyecto y los proyectos desfasados que han sido concluidos. Para lograrlo, se diseñó un instrumento con el fin de recopilar datos conformado de veintitrés reactivos como técnica de recuperación de información. Un total de dieciséis empresas locales dedicadas al desarrollo de software de Los Mochis, Mazatlán y Culiacán (ciudades principales del estado), clasificadas como MiPyMES por la Secretaría de Economía del gobierno federal mexicano, fueron consideradas para este estudio. Se aplicó una entrevista a un empleado por cada empresa, el puesto correspondía con Líder de Proyectos, Coordinador de Proyectos o Gerente de Proyectos. Posteriormente, se realizó un análisis de correlación simple y múltiple con el objetivo de conocer su grado de asociación y determinar cuánto explican las variables independientes a la variable Proyectos Desfasados Concluidos, la variable dependiente.

Este documento está organizado de la siguiente manera: la Sección 2 presenta algunos trabajos relacionados con respecto a la investigación de prácticas de SE en MiPyMES. El método aplicado es detallado en la Sección 3. Los resultados se describen en la Sección 4. Finalmente, las conclusiones son presentadas en la Sección 5.

## II. TRABAJOS RELACIONADOS

Existen varios autores que plantean que aún existen grandes brechas de comunicación entre la industria y la academia, sobre todo respecto a la colaboración en SE, por ejemplo [8] y [15]. A pesar de ello, en los últimos años se han presentado propuestas enfocadas en mejorar esa percepción, realizando estudios en temas claves del desarrollo de software como el mantenimiento, implementación y pruebas.

En [16] a partir de la aplicación en un caso de estudio, enfocado en el uso de UML para el mantenimiento de software en la industria, se plantea la necesidad de integrar prácticas de modelado en el enfoque general de SE y presentan una lista de prácticas recomendadas que contribuyen a la mayor efectividad del modelado. En el mapeo sistemático propuesto en [8], se analizan propuestas existentes respecto a colaboraciones entre industria y empresa relacionadas con la SE. Fueron identificados diez temas desafíos y diecisiete temas de mejores

prácticas en aras de tomar decisiones informadas sobre cómo estructurar dichas colaboraciones. En [17], se resalta la falta de aplicación de modelos ágiles dentro de los procesos empíricos de SE que se aplican en la industria. Los autores presentan un protocolo para ser utilizado como guía práctica generalizable a la mayoría de las colaboraciones de la industria y la academia dentro del dominio de la SE, proponiendo puntos de acción y métodos que los investigadores académicos pueden utilizar al realizar investigaciones en un entorno industrial. Para facilitar la colaboración entre la industria y la academia, en [18], se propuso una taxonomía de pruebas. Su propósito es poder ayudar en la mejora de la comunicación entre profesionales e investigadores en el área de pruebas de software.

Por otra parte, los métodos empíricos han centrado la atención de la comunidad científica de la SE durante los últimos años [19]. Existen propuestas que han sido validadas a través de métodos estadísticos, utilizados para asegurar fundamentalmente la obtención de resultados creíbles [7], [20], [21], [22], [23]. En este sentido, el análisis de datos cuantitativos es frecuente en investigaciones de casos de estudio y se pueden utilizar los siguientes análisis: estadísticas descriptivas, desarrollo de modelos predictivos y prueba de hipótesis. En el caso específico del análisis de regresión, estas pruebas permiten investigar el grado de relación existente entre dos o más variables, por ejemplo, en [24] se utilizaron técnicas de regresión lineal para la estimación de esfuerzo y costos de proyectos de software. El primer estudio del proceso de diseño de software para el desarrollo de grandes proyectos fue publicado en 1988 [25]. Este comenzó desde el punto de vista de la SE y se enfocó en cómo los requisitos y las decisiones se tomaron, representaron, comunicaron, cambiaron y cómo estas decisiones afectaron el proceso de desarrollo posterior. Los autores afirmaron que en el equipo de desarrollo el intercambio de conocimiento, su integración, facilitar los cambios en el producto (cuando se negocian nuevos requisitos con el cliente) y una amplia comunicación y coordinación fueron los tres temas clave que impactaron en el éxito de los proyectos.

En [26] una propuesta es presentada para determinar por qué falla el software. Al aplicar análisis estadístico se determinó que las entidades de software propensas a fallar están correlacionadas estadísticamente con las medidas de complejidad del código. Específicamente, se crearon modelos de regresión que predijeron con precisión la probabilidad de defectos posteriores a la liberación de cada software. En [27] los autores presentaron un estudio utilizando el método de correlación de Pearson para determinar la incidencia de los factores que afectan la calidad y el rendimiento de equipos de desarrollo colaborativos ubicados en la India y Canadá. Otro estudio de las prácticas actuales de SE, las necesidades de desarrollo y las formas preferidas de transferencia de tecnología de doce pequeñas y medianas empresas en Finlandia se publicó en [28]. Los autores determinaron que la clave de las necesidades de crecimiento en la industria son el desarrollo de las propias adaptaciones del proceso de SE, la mejora de procesos relacionados con la ingeniería de requisitos (RE por sus siglas en inglés) y la automatización de sus prácticas. Dirigir los esfuerzos a estas áreas podrían mejorar sustancialmente las posibilidades de una transferencia de tecnología exitosa y el

mejoramiento de los procesos en la industria.

Resumiendo, se han llevado a cabo varios estudios con respecto a la aplicación y práctica de la SE a través de análisis estadísticos, los resultados obtenidos presentan problemas como la necesidad de mejora de procesos de software, capacitación del equipo de desarrollo y el desarrollo de las propias adaptaciones del proceso de SE. En Sinaloa un estudio con características similares que se considere como un punto de partida para la mejora de la industria del software no se ha realizado particularmente con ese enfoque. Esto ha motivado el interés en conducir esta investigación para obtener el grado de relación de factores que influyen en el desarrollo no satisfactorio de proyectos de software en las MiPyMES del estado y proponer, en un futuro, soluciones *ad-hoc* que sirvan como base para el fortalecimiento de la industria.

### III. MÉTODO DE INVESTIGACIÓN

En este trabajo se entrevistó a dieciséis empresas (de treinta y dos contactadas) que se dedican al desarrollo de software clasificadas como MiPyMES, ubicadas en las ciudades más importantes del estado de Sinaloa, México (Los Mochis, Culiacán y Mazatlán). La entrevista se aplicó a un miembro del equipo de trabajo de cada una de las empresas. A través de la aplicación de la técnica de la entrevista [29], se diseñó un instrumento para la recopilación de datos con veintitrés preguntas, divididas en dos categorías para determinar la práctica de la SE. Las preguntas incluyeron elementos relacionados con la cantidad total de empleados, el tipo de proyectos que desarrollan, el tiempo invertido en las actividades de ingeniería de requisitos del proyecto, las herramientas que utilizan y las estrategias que utilizan para mejorar estos escenarios.

#### A. Diseño de Preguntas

Se elaboró un conjunto de preguntas para realizar una entrevista considerando que el trabajador de la empresa podría ser flexible en las respuestas de acuerdo con el flujo de conversación. La entrevista consideraba preguntas sobre los antecedentes de la empresa y del empleado que la contesta, así como también preguntas relacionadas con la práctica de la SE en proyectos de desarrollo de software. Se seleccionó la entrevista como medio para la recopilación de la información debido a que es un instrumento técnico que adopta la forma de un diálogo coloquial lo que nos permitió conocer la terminología utilizada por la empresa en un proceso de desarrollo de software, particularmente en el nombramiento del puesto de los integrantes del equipo y las etapas de su proceso.

#### B. Recopilación de Datos

Para la obtención de los datos se creó un instrumento compuesto de veintitrés reactivos a través de la aplicación web *Google Forms*. Las preguntas que lo integran pueden consultarse en el siguiente enlace: <https://bit.ly/2TIkLRf>. Posteriormente, se envió una invitación formal (vía correo electrónico) solicitando a los contactos de las empresas que accedieran y contestaran los reactivos, incluyendo un recordatorio posterior. Las empresas y sus contactos fueron extraídos de la base de

datos del gobierno y de estudiantes y egresados de la Facultad de Informática Mazatlán, Facultad de Informática Culiacán y Escuela de Ingeniería Mochis de la Universidad Autónoma de Sinaloa (UAS). Esta invitación cerrada aseguró la inclusión solo de profesionales de la industria en la población del instrumento.

En total se entrevistaron dieciséis empresas de las cuales la mitad cuenta con cincuenta o menos empleados. El resto tienen de trescientos hasta dos mil quinientos empleados y operan a nivel nacional e internacional, mientras que las empresas con cincuenta o menos trabajadores operan a nivel estatal (solo Sinaloa). Cabe destacar que únicamente catorce empresas aportaron datos completos a la encuesta. El software que desarrollan lleva un enfoque *business-to-business*, es decir, transacciones de comercio entre empresas, entre un fabricante y un mayorista o minorista, y *business-to-consumer* para la venta de productos directamente a los consumidores. El modelo de negocio más utilizado en las empresas entrevistadas es el *business-to-business*. Lo anterior indica que venden sus productos y servicios a otras empresas, fundamentalmente software y/o intercambio de datos.

Los entrevistados fueron seleccionados en función de su cargo (puesto) y por disposición de la empresa contactada. Participaron seis administradores de proyectos, cuatro directores generales, tres analistas (responsable para la preparación del documento de especificación de requisitos funcionales) y tres desarrolladores categoría *senior*.

Para una correcta captura de los datos fueron necesarios los siguientes pasos:

- 1) Los datos se recopilaron en dos formas: enviando un correo electrónico con el formulario de *Google* y la segunda a través de la visita *in situ* de los colaboradores del proyecto. La duración de la entrevista fue de aproximadamente media hora a una hora.
- 2) Se capturaron los datos recopilados en la entrevista *in situ* y vía formulario web. Es importante mencionar que algunas empresas nombran el cargo de Coordinador de Proyectos y en otras su equivalente es Líder de Proyectos o Gerente de Proyectos, por lo que fue necesario estandarizar los nombres de los cargos.
- 3) Los datos fueron analizados por medio del software *Microsoft Excel* para descartar las variables (atributos) de poca significancia debido a su baja incidencia en la variable dependiente Proyectos Desfasados Concluidos (PC). Se consideró esta variable dependiente porque las empresas consultadas afirman que todos sus proyectos se desfasan. El método estadístico aplicado para descartar las variables fue el de significancia estadística. En el proceso de depuración de variables poco significativas se eliminaron dieciséis variables, quedando finalmente siete, seis independientes y una dependiente, de tal forma que se seleccionaron las variables Ubicación de la Empresa (UE), Ámbito de Cobertura (AC), Número de Trabajadores (NT), Tiempo de Vida en el Mercado (TV), Proyectos Concluidos (PY), Tiempo Dedicado a Actividades de Ingeniería de Requisitos del Proyecto (TA), y Proyectos Desfasados Concluidos (PC). La Ubicación de la Empresa se codificó de acuerdo a su localización

geográfica en el estado de Sinaloa, de la siguiente manera, 1: Culiacán, 2: Mazatlán, 3: Los Mochis. Para el Ámbito de Cobertura se asignó una nomenclatura de acuerdo a su nivel de exposición, 1: Municipal, 2: Estatal, 3: Nacional, 4: Internacional.

- 4) Se realizó un análisis de correlación de Pearson ( $r$ ) de forma independiente entre las variables, con la finalidad de determinar el grado de relación entre cada una de las variables mencionadas, con todas las demás. A pesar de que la variable independiente es Proyectos Desfasados Concluidos, se considera interesante mostrar la matriz de correlaciones entre las siete variables, debido a que es posible conocer el grado de las interrelaciones entre ellas. Adicionalmente, se calculó el coeficiente de correlación múltiple ( $r$ ) con las mismas variables y datos, así como el coeficiente de determinación ( $R^2$ ) y el coeficiente de determinación ajustado ( $\bar{R}^2$ ) para encontrar el grado de asociación lineal entre la variable dependiente (respuesta) y variables independientes (explicativas); concretamente entre la variable Proyectos Desfasados Concluidos y la recta de regresión estimada con el resto de los factores. Se decidió utilizar PC como variable dependiente porque es un indicador de efectividad en las empresas que desarrollan software, al medir la cantidad de trabajo llevados a buen término a pesar de haber tenido retraso durante la elaboración.

### C. Limitación del Método de Estudio

En este trabajo se presentan datos detallados recopilados de dieciséis empresas de software en Sinaloa, México. La muestra es relativamente pequeña para generalizar, aunque deben proporcionar un punto de partida interesante para estudiar la práctica de la SE a razón de que no existe tal estudio de esta región de México. Esta afirmación se sustenta debido a que las empresas entrevistadas mencionaron su interés en este estudio debido a que no recordaban la realización de alguno de este tipo. Además, se realizó una búsqueda en repositorios de las universidades del estado que ofrecen estudios de licenciatura, ingeniería y posgrado en el área de informática y sistemas computacionales. Asimismo, se buscó en bases de datos bibliográficas como *Scopus* y en *DBLP Computer Science Bibliography*.

## IV. RESULTADOS

Para calcular la correlación entre las distintas variables que intervienen en este estudio en el desarrollo de software por empresas localizadas en el estado de Sinaloa, México, se realizó un análisis de correlación de forma independiente entre las siete variables. Además, se determinó la correlación múltiple, utilizándose para este proceso seis variables independientes o explicativas y la variable dependiente Proyectos Desfasados Concluidos (PC).

En la Tabla I, se muestran las variables involucradas en el modelo de correlación. Estas variables intervienen en el cálculo de correlación realizado para determinar la asociación entre ellas y de las seis independientes con el factor de respuesta Proyectos Desfasados Concluidos.

TABLA I  
RELACIÓN DE VARIABLES UTILIZADAS EN EL ANÁLISIS DE  
CORRELACIÓN. FUENTE: PROPIA

VARIABLE	UNIDAD DE MEDIDA	RANGO DE VALORES
UBICACIÓN DE LA EMPRESA (UE)	1: Culiacán, 2: Mazatlán, 3: Los Mochis	[1-3]
ÁMBITO DE COBERTURA (AC)	1: Municipal, 2: Estatal, 3: Nacional, 4: Internacional	[1-4]
NÚMERO DE TRABAJADORES (NT)	Número entero	[1-2500]
TIEMPO DE VIDA EN EL MERCADO (TV)	Número entero	[1-75]
PROYECTOS CONCLUIDOS (PV)	Número entero	[1-500]
TIEMPO DEDICADO A ACTIVIDADES DE INGENIERÍA DE REQUISITOS DEL PROYECTO (TA)	Porcentaje	[1-50]
PROYECTOS DESFASADOS CONCLUIDOS (PC)	Porcentaje	[50-100]

En el cálculo de correlación de Pearson, de cada una de las variables con las otras seis, se encontró un bajo índice de vinculación entre ellas (ver Tabla II), siendo los más sobresalientes cinco indicadores, de veintidós posibles.

La Tabla II presenta todas las correlaciones posibles entre las siete variables analizadas, se destacan cinco indicadores con valores mayores a 0.4 que, a pesar de no ser muy altos, son los que representan mayor vinculación entre ellas, de forma independiente (correlación más alta). La mayor asociación está dada por PC-AC con  $r = 0.528$ , seguido por  $r = -0.509$  de TA-NT; los siguientes valores son menores a 0.5:  $r = -0.459$ ,  $r = 0.429$  y  $r = 0.441$  para PC-UE, TV-NT y TV-AC, respectivamente. En este sentido, los p-valor calculados para cada correlación son altos, por lo que la probabilidad de error es también elevada. Con excepción de tres casos, las correlaciones de Ámbito de Cobertura (AC) y Proyectos Desfasados Concluidos (PC) con un p-valor de 0.052, donde únicamente existe un 5% de posibilidad de error. En la correlación de las variables Número de Trabajadores (NT) y Tiempo Dedicado a Actividades de Ingeniería de Requisitos del Proyecto (TA) el p-valor es de 0.066; al medir el grado de relación entre Ubicación de la Empresa (UE) y Proyectos Desfasados Concluidos (PC) se encontró un p-valor = 0.099; aunque estos dos últimos valores sobrepasan el 0.05 permitido ( $\alpha = 0.05$ ), no son tan altos como los demás encontrados. A continuación se presentan cinco gráficas (Fig. 1 a Fig. 5) para un mejor entendimiento del análisis de correlación de Pearson ( $r$ ) detallado en la Tabla II.

La Fig. 1 muestra una gráfica con el análisis correlacional de las variables PC contra UE, como se aprecia la relación es negativa y existe poca correspondencia entre ellas, como lo demuestra  $R^2 = 0.2109$ . Los puntos que menos se ajustan al patrón, son dos (1, 70) y (2, 50), que reflejan 70 proyectos desfasados concluidos en la ciudad de Culiacán y 50 en la ubicación 2, Mazatlán. El patrón muestra que en la ubicación 1, Culiacán, los proyectos oscilan entre 90 y 100; en la ubicación 2 entre 60 y 100. Por otro lado, la única empresa de la ubicación 3 Los Mochis, tiene 60 proyectos desfasados concluidos.

La Fig. 2 se observa que el coeficiente de correlación es  $r = 0.413$  y  $R^2 = 0.0894$ , por lo que TV y AC están poco relacionadas, es decir, el tiempo de vida de la empresa tiene muy poco que ver con su ámbito de cobertura. Las empresas encuestadas, con un área de cobertura local, estatal, nacional e internacional llevan en el mercado de 1 a 20 años, excepto dos organizaciones de nivel internacional una con 44 años y otra con 75, por lo que su dispersión es notoria comparada con el resto de los datos.

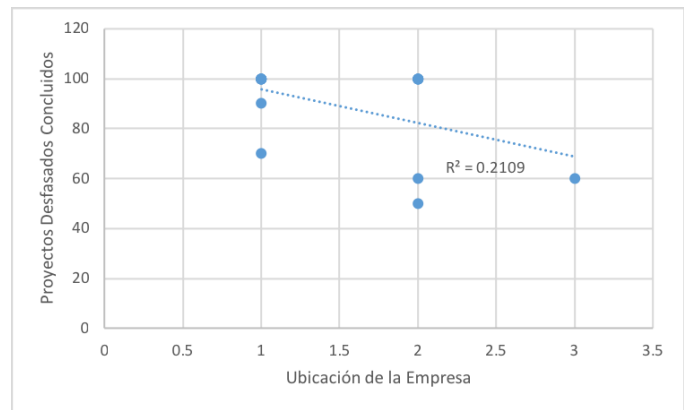


Fig. 1. Proyectos Desfasados Concluidos (PC) - Ubicación de la Empresa (UE).

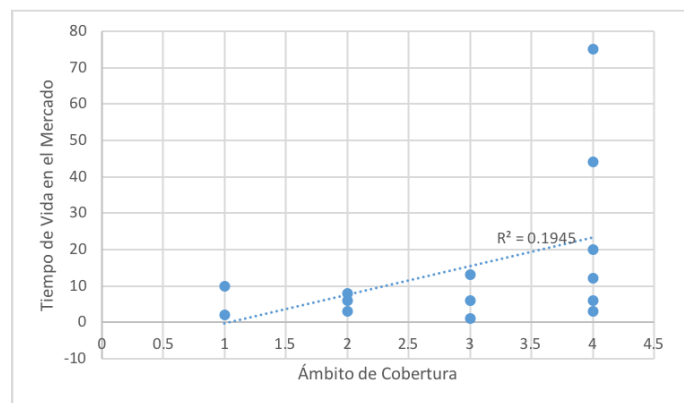


Fig. 2. Tiempo de Vida en el Mercado (TV) - Ámbito de Cobertura (AC).

Como se observa en la gráfica de la Fig. 3, en nuestro estudio afecta que el proyecto se desfase pero se concluya, los proyectos desfasados concluidos en pocas ocasiones se debe al ámbito de cobertura de la empresa, es decir, si la empresa es local, estatal, nacional o internacional, lo demuestra  $r = -0.509$ , que es uno de los valores más altos obtenidos. Podrían considerarse como valores atípicos una empresa de ámbito local que únicamente tiene 50 proyectos desfasados concluidos, una empresa de nivel estatal con 60 proyectos desfasados concluidos y una empresa de cobertura internacional con 60 proyectos desfasados concluidos; el resto de las empresas tiene entre 70 y 100 proyectos desfasados concluidos.

Inusualmente, una empresa cuenta con dos mil quinientos trabajadores y tiene cuarenta y cuatro años de vida, es por eso que la correlación es muy baja. Esto se sustenta en que la industria del software en Sinaloa es relativamente joven, pues las empresas oscilan entre los dos y diez años de antigüedad y solo una minoría cuenta con doce y trece años, con la excepción mencionada (Fig. 4). En cuanto al tiempo de vida de la empresa en el mercado, como ya se mencionó, doce de las catorce oscilan entre uno y veinte años, con dos empresas que tienen cuarenta y cuatro y setenta y cinco años. Respecto al número de trabajadores, once de las empresas mantienen entre cuatro y setenta y cinco empleados, restando tres organizaciones con noventa, trescientos y dos

TABLA II  
ANÁLISIS DE CORRELACIÓN DE PEARSON (R). CON UNA PROBABILIDAD DE ERROR (P-VALOR) DE: 0.159

	UE		AC		NT		TV		PY		TA		PC	
	r	p-valor	r	p-valor	r	p-valor	r	p-valor	r	p-valor	r	p-valor	r	p-valor
UE	1													
AC	0.164	0.575	1											
NT	0.140	0.633	0.227	0.435	1									
TV	-0.183	0.532	<b>0.441</b>	0.115	<b>0.429</b>	0.126	1							
PY	0.380	0.180	-0.172	0.557	0.129	0.660	0.239	0.411	1					
TA	0.241	0.406	-0.057	0.846	<b>-0.509</b>	0.063	0.066	0.823	0.168	0.566	1			
PC	<b>-0.459</b>	0.099	<b>0.528</b>	0.052	0.165	0.574	0.294	0.308	-0.061	0.837	0.021	0.943	1	

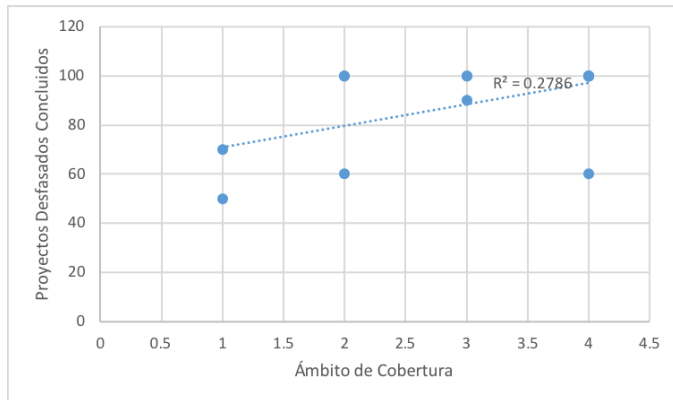


Fig. 3. Proyectos Desfasados Concluidos (PC) - Ámbito de Cobertura (AC).

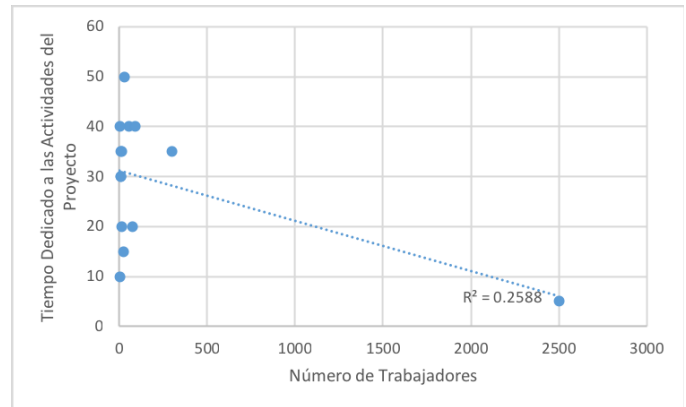


Fig. 5. Tiempo Dedicado a Actividades de Ingeniería de Requisitos del Proyecto (TA) - Número de Trabajadores (NT).

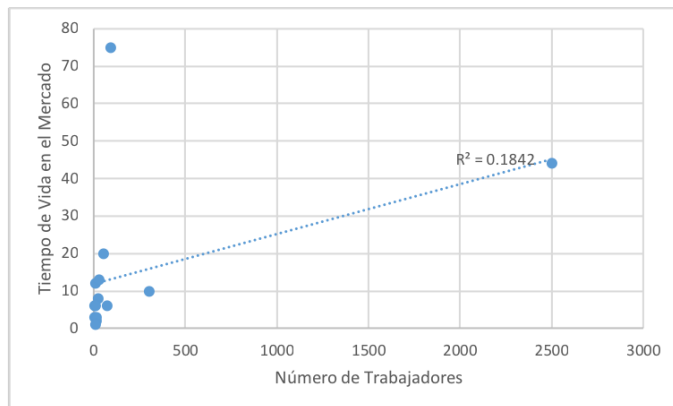


Fig. 4. Tiempo de Vida en el Mercado (TV) - Número de Trabajadores (NT).

mil quinientos trabajadores. En la Fig. 4 se muestran como valores con mayor dispersión (90, 75), lo que significa noventa trabajadores para una empresa de setenta y cinco años de antigüedad; y (2500, 44), para dos mil quinientos empleados en una empresa con cuarenta y cuatro años.

En la gráfica de la Fig. 5 se observa que el tiempo dedicado a las actividades de ingeniería de requisitos con el proyecto incide en su conclusión con desfase porque repercute en esto el número de trabajadores de la empresa, como lo muestra  $r = -0.509$ , que es uno de los valores más elevados obtenidos en el análisis de correlación y asimismo con el coeficiente de determinación  $R^2 = 0.2588$ . Además, se ilustra el tiempo, en porcentaje, dedicado a la ingeniería de requisitos de acuerdo

al número de trabajadores; la mayoría de las empresas, 13 de ellas, afirma dedicar del 10 al 50% del tiempo a esta actividad, con excepción de una empresa que únicamente le dedica el 5%, con 2500 trabajadores contratados.

En cuanto al análisis de correlación múltiple, éste obtuvo como resultado una R con valor de 0.804, el coeficiente de determinación  $R^2 = 0.647$  y un  $\bar{R}^2$  ajustado de 0.293. Se realizó también el análisis de varianza (ANOVA) considerando las siete variables mencionadas y se mantiene como variable dependiente PC. Los resultados obtenidos se muestran en la Tabla III, donde SUMA DE CUADRADOS corresponde a la suma de las desviaciones al cuadrado de cada valor con respecto a la media (este valor mide la dispersión máxima de los datos), GL es el número de grados de libertad del tamaño de la muestra (para este caso  $n=16$ ), la MEDIA CUADRÁTICA obtenida a partir de SUMA DE CUADRADOS/GL y F como la relación de dos cuadrados medios, este valor ayuda a medir la variabilidad de los datos. Finalmente, con la significancia (SIG) se obtiene una validación estadística del modelo.

La Tabla III presenta el análisis de varianza de las siete variables, UE, AC, NT, TV, PY, TA y la variable dependiente PC. Como se observa, el valor crítico o significancia (SIG) es mayor a 0.05, lo que indica poca relación entre las variables; pero el estadístico F es mayor que su valor crítico, lo que indica cierto grado de bondad en el modelo compuesto por las siete variables.

TABLA III  
ANÁLISIS DE VARIANZA DE LAS SIETE VARIABLES

MODELO	SUMA DE CUADRADOS	GL	MEDIA CUADRÁTICA	F	SIG
REGRESIÓN	2895.316	6	482.553	1.831	.240
RESIDUAL	1581.607	6	263.601		
TOTAL	4476.923	12			

## V. CONCLUSIONES

La práctica de un proceso de desarrollo bien definido basado en las mejores prácticas de la SE en empresas de desarrollo de software de micro y pequeño tamaño aún no se ha analizado a profundidad en México, especialmente en Sinaloa. Este trabajo presentó el estado actual del proceso de desarrollo de software de las MiPyMES del estado a través de una serie de variables. El estudio se basó en la información obtenida a través de una entrevista (compuesta de veintitrés preguntas) aplicado a dieciséis empresas de desarrollo de software.

En esta investigación, con respecto a la correlación de las variables medidas a las empresas de desarrollo de software en Sinaloa, México, puede concluirse que las siete variables analizadas, Ubicación de la Empresa, Ámbito de Cobertura, Número de Trabajadores, Tiempo de Vida en el Mercado, Proyectos Concluidos, Tiempo Dedicado a Actividades de Ingeniería de Requisitos del Proyecto y Proyectos Desfasados Concluidos, no tienen una fuerte asociación entre ellas, es decir, no hay correlación directa entre los pares de variables, caso contrario cuando se hace un análisis multivariable considerando las seis variables independientes y la variable dependiente, en donde  $r = 0.804$ , demostrando una mayor asociación entre ellas. En este sentido, los índices de correlación más altos van de 0.413 a -0.509, pero, en conjunto muestran una correlación múltiple considerablemente mayor si se tienen en cuenta los índices bajos resultantes de  $r$ ,  $r = 0.804$ ,  $R^2 = 0.647$  y  $\bar{R}^2 = 0.293$ . Lo anterior puede explicarse como la sinergia de los diferentes factores que influyen en el desarrollo de software, donde cada uno de ellos, aunque en apariencia insignificante, es clave para lograr llevar a buen término un proyecto de esta naturaleza.

Los resultados presentan dos cuestiones, estas son: i)

- 1) si la empresa es local, estatal, nacional o internacional, en nuestro estudio, afecta que el proyecto se desfase pero se concluya, lo demuestra  $r = -0.509$ , que es uno de los dos valores más altos obtenidos.
- 2) el tiempo dedicado a las actividades de ingeniería de requisitos del proyecto afecta su conclusión con desfase porque repercute en esto el número de trabajadores de la empresa, como lo muestra  $r = -0.509$ .

Lamentablemente, el uso de prácticas de SE *ad-hoc* (dentro de un proceso de desarrollo que no está bien definido) como lo es la ingeniería de requisitos, es poco significativo para las organizaciones de micro y pequeño tamaño. Con base en estos resultados, se recomienda impulsar en las empresas de desarrollo de software la relevancia del uso de la SE en su práctica a través de estudios.

Cabe destacar que en esta investigación no partimos de un supuesto, debido que planteamos una investigación de nivel

exploratoria - descriptiva, debido a que se está en un primer acercamiento al objeto de estudio y se desea, 1) tener una visión más clara de dicho objeto y 2) establecer su estructura y comportamiento.

Como trabajo futuro se pretende considerar la inclusión de un mayor número de variables y determinar las significativas en el proceso que se estudia, así como la construcción de un aceptable modelo probabilístico de predicción. Particularmente se realizará un estudio descriptivo correlacional-explicativo, mediante pruebas de hipótesis correlacionales y relaciones causa-efecto. Además, se identificarán lagunas para establecer las bases para una actualización de los planes de estudio de las carreras de computación e informática con respecto a la práctica de la SE en las universidades locales. Esto ayudará a fortalecer la promoción de la relación entre universidades y la empresa para obtener recurso humano mejor capacitado.

## AGRADECIMIENTOS

Especial agradecimiento a las dieciséis empresas que participaron en el proyecto cuyos nombres se omiten a propósito por cuestiones de anonimato.

## REFERENCIAS

- [1] H. Kopka and P. W. Daly, *A Guide to L<sup>A</sup>T<sub>E</sub>X*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
- [2] P. Bourque and R. E. Fairley, "Guide to the software engineering body of knowledge (SWEBOK (R)): Version 3.0", IEEE Computer Society Press, 2014.
- [3] H. Oktaba and A. Vázquez, "MoProSoft®: A software process model for small enterprises", In *Software Process Improvement for Small and Medium Enterprises: techniques and case studies*, IGI Global, pp. 170-176, 2008.
- [4] M. B. Chrissis, M. Konrad and S. Shrum, "CMMI guidelines for process integration and product improvement", Addison-Wesley Longman Publishing Co., Inc., 2003.
- [5] CMMI Institute. [Online]. Available: <http://www.cmmiinstitute.com>, Accessed on: Feb. 6, 2019.
- [6] D. Ståhl, and J. Bosch, "Modeling continuous integration practice differences in industry software development", *Journal of Systems and Software*, vol. 87, pp. 48-59, ISSN 0164-1212, 2014.
- [7] C. Wohlin, "Empirical software engineering research with industry: top 10 challenges", In *Proceedings of the 1st International Workshop on Conducting Empirical Studies in Industry (CESI '14)*, IEEE Press, Piscataway, NJ, USA, 43-46, 2013.
- [8] V. Garousi, K. Petersen and B. Ozkan, "Challenges and best practices in industry-academia collaborations in software engineering: A systematic literature review". *Information and Software Technology*, vol. 79, pp. 106-127, 2016.
- [9] M. Munoz, A. Pena, J. Mejia and G. Lara, "Coverage of the university curricula for the Software Engineering industry in Mexico". *IEEE Latin America Transactions*, pp.2382-2388, 2016.
- [10] Standish Group, T., *The CHAOS Manifesto*. Think Big, Act Small, Boston: The Standish Group, 2016.
- [11] K. E. Emam and A. G. Koru, "A Replicated Survey of IT Software Project Failures", *IEEE Software*, pp. 84-90, 2008.
- [12] A. Miranda, L. Sandoval, R. Martínez and E. López, "Análisis de la industria del software. Caso Sinaloa", In *Proceedings of the Congreso Internacional de Investigación*, Academia Journals Celaya 2013, ISSN 1946-5351, Vol. 5, No. 3, 2013.
- [13] E. Díaz, C. Castro, and E. Santamaría, Índice Nacional de Ciencia, Tecnología e Innovación 2018, Centro de Análisis para la Investigación en Innovación, A.C. <https://www.caiinno.org/wp-content/uploads/2018/12/INDICE-2018.pdf>, Accessed on: Feb. 3, 2019.
- [14] Gobierno del Estado de Nuevo León, "50 empresas de software hacia Industria 4.0". [Online]. Available: <http://www.nuevoleon.gob.mx/noticias/impulsaranuevo-leon-50-empresas-de-software-hacia-industria-40>, Accessed on: Feb. 3, 2019.

- [15] N. M. Minhas, K. Petersen, N. B. Ali and K. Wnuk, "Regression testing goals-View of practitioners and researchers", In Proceedings of the 24th Asia-Pacific Software Engineering Conference Workshops (APSECW), pp. 25-31, 2017.
- [16] A. M. Fernández-Sáez, M. R.V. Chaudron and M. Genero, "An industrial case study on the use of UML in software maintenance and its perceived benefits and hurdles", *Empirical Software Engineering*, pp. 3281-3345, 2018.
- [17] K. Biesialska, X. Franch and V. Muntés-Mulero, "Protocol and tools for conducting agile software engineering research in an industrial-academic setting: a preliminary study", In Proceedings of the IEEE/ACM 6th International Workshop on Conducting Empirical Studies in Industry (CESI), pp. 29-32, 2018.
- [18] E. Engström, K. Petersen, N. Bin Ali and E. Bjarnason, "SERP-test: a taxonomy for supporting industry-academia communication", *Software Quality Journal*, pp. 1269-1305, 2017.
- [19] F. García et al., "La Experimentación en la docencia de Ingeniería del Software", In Jornadas de Enseñanza Universitaria de la Informática (JENUI'04), pp.237-245, 2004.
- [20] L. Briand et al., "Empirical studies of object-oriented artifacts, methods, and processes: state of the art and future directions", *Empirical Software Engineering*, pp.387-404, 1999.
- [21] N. Juristo and A. M. Moreno, "Basics of software engineering experimentation", Springer Science & Business Media, 2013.
- [22] B. A. Kitchenham, et al., "Preliminary guidelines for empirical research in software engineering", *IEEE Transactions on software engineering*, pp.721-734, 2002.
- [23] D. E. Perry, A. A. Porter and L. G. Votta, "Empirical studies of software engineering: a roadmap", In Proceedings of the conference on The future of Software engineering, pp. 345-355, 2000.
- [24] V. Anandhi and R. M. Chezian, "Regression techniques in software effort estimation using cocomo dataset", *International Conference on Intelligent Computing Applications*, pp. 353-357, 2014.
- [25] B. Curtis, H. Krasner and N. Iscoe, "A field study of the software design process for large systems", *Communications of the ACM*, pp. 1268-1288, 1988.
- [26] N. Nagappan, T. Ball and A. Zeller, "Mining metrics to predict component failures". In Proceedings of the 28th international conference on Software engineering, pp. 452-461, 2006.
- [27] H. K. Edwards and V. Sridhar, "Analysis of the effectiveness of global virtual teams in software engineering projects", In Proceedings of the 36th Annual Hawaii International Conference on System Sciences, pp. 9, 2003.
- [28] U. Nikula, J. Sajaniemi and H. Kälviäinen, "A State-of-the-practice Survey on Requirements Engineering in Small-and Medium-sized Enterprises". Lappeenranta University of Technology Lappeenranta, Finland, 2000.
- [29] T. Punter, M. Ciolkowski, B. Freimut and I. John, "Conducting on-line surveys in software engineering", In Proceedings of the International Symposium on Empirical Software Engineering, pp. 80-88, 2003.



**José Alfonso Aguilar-Calderón** was born in Los Mochis, Sinaloa, Mexico in 1983. He received the B.S. in computer systems and M.S. in applied informatics degrees from the Universidad Autónoma de Sinaloa, Mazatlán, in 2005 and 2007, and the Ph.D. degree in application of the informatics from Universidad de Alicante, Alicante, Spain, in 2011. From 2006-2007, he was software developer in Empresaria and DpSoft, and from 2007-2009 in computer systems department in Coppel S.A de C. V. in Quality Assurance area. Since 2012 he has been a Full-time Research-Professor at the Facultad de Informática Mazatlán from Universidad Autónoma de Sinaloa, at Mazatlán, Mexico. He is the author of one book, more than 20 articles, and 3 copyrights of software code. His research interests include Goal-oriented Requirements Engineering (GORE), Web Engineering (WE), Requirements Engineering (RE), Model-Driven Web Engineering (MDWE). He is an Associate Editor and co-founder of the journal *Revista de Investigación en Tecnologías de la Información (RITI)*. Dr. Aguilar was a recipient of the Best Demo and Poster Award in 2012 at the 12th International Conference on Web Engineering (ICWE) and the Best Paper Award in the 6th IEEE International Conference on Adaptive Science & Technology (ICAST) in 2014. He is member of the National System of Researchers (Sistema Nacional de Investigadores, SNI-1) at Mexico since 2013.



**Anibal Zaldivar-Colado** was born in Mazatlán, Sinaloa, Mexico in 1972. He received the B.S. in civil engineering (1997) and in computer systems (2000) from the Facultad de Informática of the Universidad Autónoma de Sinaloa (México); and his MSc. and Ph.D. degrees from the Universidad de Durango (UAD), Mexico, in 2000 and 2011 respectively. Currently he is an Professor in the Universidad Autónoma de Sinaloa (Mazatlán, México). He is Editor of two collective books, Author of more than 14 articles and 4 chapters of refereed books, in addition to multiple presentations at conferences. His research activity is focused on applied mathematical modelling, systems simulation and educational technology. Dr. Zaldivar was a recipient of the Paper Award in the 6th IEEE International Conference on Adaptive Science & Technology (ICAST) in 2014. He is member of the National System of Researchers (Sistema Nacional de Investigadores, SNI-1) at Mexico since 2014.



**Carolina Tripp Barba** was born in Mazatlán, Sinaloa, Mexico in 1984. She received the B.S. in computer systems from the Faculty of Informatics of the Universidad Autónoma de Sinaloa (México) in 2006; and her MSc. and Ph.D. degrees in Telematics Engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2009 and 2013 respectively. Since 2013 she has been a full-time research-professor at the Facultad de Informática Mazatlán from Universidad Autónoma de Sinaloa, Mexico. She is coordinator of a collective book, autor in two chapters book and of more than 30 articles. Her research activity is focused on VANETs (Vehicular Ad hoc Networks), the design of routing protocols for vehicular networks, QoS provision and Smart Cities. She is Editor and co-founder of the journal *Revista de Investigación en Tecnologías de la Información (RITI)*, invited editor in the Ad Hoc Networks and co-chair in IEEE Consumer Communications & Networking Conference. Dra. Tripp is member of the National System of Researchers (Sistema Nacional de Investigadores, SNI-1) at Mexico since 2014.





**Roberto Espinosa Oliva** was born in Matanzas, Cuba in 1981. He received the B.S. in computer systems and M.S. in applied informatics degrees from the Universidad de Matanzas, Cuba, in 2005 and 2008, and the Ph.D. degree in application of the informatics from Universidad de Alicante, Alicante, Spain, in 2014. He was a full-time professor in the Department of Informatics at the University of Matanzas, Cuba from 2005 to 2016. He completed postdoctoral studies at the University de la Frontera, Chile from 2016 to 2018. Currently, he is a full-time

research-professor in the Universidad Tecnológica de Chile INACAP, Chile. His research interests are focused on user requirements, friendly application of data mining techniques, development of model-driven techniques and Big Data.



**Carlos Eduardo Zurita Cruz** received his Ph.D. degree from the Centro de Investigación e Innovación Educativa del Noroeste (CIEN), México, in 2015. Currently he is an Associate Professor in the Universidad Autónoma de Sinaloa (Mazatlán, México). His research activity is focused on networks area, statistical analysis and educational technology.