Alvaro Rocha · Hojjat Adeli ·
Gintautas Dzemyda · Fernando Moreira ·
Valentina Colla
Editors

Information Systems and Technologies

WorldCIST 2023, Volume 1

Springer

Editors
Alvaro Rocha
ISEG
Universidade de Lisboa
Lisbon, Cávado, Portugal

Hojjat Adeli
College of Engineering
The Ohio State University
Columbus, OH, USA

Gintautas Dzemyda
Institute of Data Science and Digital Technologies
Vilnius University
Vilnius, Lithuania

Fernando Moreira
DCT
Universidade Portucalense
Porto, Portugal

Valentina Colla
TeCIP Institute
Scuola Superiore Sant'Anna
Pisa, Italy

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-45641-1 ISBN 978-3-031-45642-8 (eBook)
https://doi.org/10.1007/978-3-031-45642-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

This book contains a selection of papers accepted for presentation and discussion at the 2023 World Conference on Information Systems and Technologies (WorldCIST'23). This conference had the scientific support of the Sant'Anna School of Advanced Studies, Pisa, University of Calabria, Information and Technology Management Association (ITMA), IEEE Systems, Man, and Cybernetics Society (IEEE SMC), Iberian Association for Information Systems and Technologies (AISTI), and Global Institute for IT Management (GIIM). It took place in Pisa city, Italy, 4–6 April 2023.

The World Conference on Information Systems and Technologies (WorldCIST) is a global forum for researchers and practitioners to present and discuss recent results and innovations, current trends, professional experiences and challenges of modern Information Systems and Technologies research, technological development, and applications. One of its main aims is to strengthen the drive toward a holistic symbiosis between academy, society, and industry. WorldCIST'23 was built on the successes of: WorldCIST'13 held at Olhão, Algarve, Portugal; WorldCIST'14 held at Funchal, Madeira, Portugal; WorldCIST'15 held at São Miguel, Azores, Portugal; WorldCIST'16 held at Recife, Pernambuco, Brazil; WorldCIST'17 held at Porto Santo, Madeira, Portugal; WorldCIST'18 held at Naples, Italy; WorldCIST'19 held at La Toja, Spain; WorldCIST'20 held at Budva, Montenegro; WorldCIST'21 held at Terceira Island, Portugal; and WorldCIST'22, which took place online at Budva, Montenegro.

The Program Committee of WorldCIST'23 was composed of a multidisciplinary group of 339 experts and those who are intimately concerned with Information Systems and Technologies. They have had the responsibility for evaluating, in a 'blind review' process, and the papers received for each of the main themes proposed for the Conference were: A) Information and Knowledge Management; B) Organizational Models and Information Systems; C) Software and Systems Modeling; D) Software Systems, Architectures, Applications, and Tools; E) Multimedia Systems and Applications; F) Computer Networks, Mobility, and Pervasive Systems; G) Intelligent and Decision Support Systems; H) Big Data Analytics and Applications; I) Human-Computer Interaction; J) Ethics, Computers & Security; K) Health Informatics; L) Information Technologies in Education; M) Information Technologies in Radiocommunications; and N) Technologies for Biomedical Applications.

The conference also included workshop sessions taking place in parallel with the conference ones. Workshop sessions covered themes such as: Novel Computational Paradigms, Methods, and Approaches in Bioinformatics; Artificial Intelligence for Technology Transfer; Blockchain and Distributed Ledger Technology (DLT) in Business; Enabling Software Engineering Practices Via Latest Development's Trends; Information Systems and Technologies for the Steel Sector; Information Systems and Technologies for Digital Cultural Heritage and Tourism; Recent Advances in Deep Learning Methods and Evolutionary Computing for Health Care; Data Mining and Machine Learning in Smart Cities; Digital Marketing and Communication, Technologies, and Applications;

Digital Transformation and Artificial Intelligence; and Open Learning and Inclusive Education Through Information and Communication Technology.

WorldCIST'23 and its workshops received about 400 contributions from 53 countries around the world. The papers accepted for oral presentation and discussion at the conference are published by Springer (this book) in four volumes and will be submitted for indexing by WoS, Scopus, EI-Compendex, DBLP, and/or Google Scholar, among others. Extended versions of selected best papers will be published in special or regular issues of leading and relevant journals, mainly JCR/SCI/SSCI and Scopus/EI-Compendex indexed journals.

We acknowledge all of those that contributed to the staging of WorldCIST'23 (authors, committees, workshop organizers, and sponsors). We deeply appreciate their involvement and support that was crucial for the success of WorldCIST'23.

April 2023

Alvaro Rocha
Hojjat Adeli
Gintautas Dzemyda
Fernando Moreira
Valentina Colla

Organization

Honorary Chair

Hojjat Adeli	The Ohio State University, USA

General Chair

Álvaro Rocha	ISEG, University of Lisbon, Portugal

Co-chairs

Gintautas Dzemyda	Vilnius University, Lithuania
Sandra Costanzo	University of Calabria, Italy

Workshops Chair

Fernando Moreira	Portucalense University, Portugal

Local Organizing Committee

Valentina Cola (Chair)	Scuola Superiore Sant'Anna—TeCIP Institute, Italy
Marco Vannucci	Scuola Superiore Sant'Anna—TeCIP Institute, Italy
Vincenzo Iannino	Scuola Superiore Sant'Anna—TeCIP Institute, Italy
Stefano Dettori	Scuola Superiore Sant'Anna—TeCIP Institute, Italy

Advisory Committee

Ana Maria Correia (Chair)	University of Sheffield, UK
Brandon Randolph-Seng	Texas A&M University, USA

Chris Kimble	KEDGE Business School & MRM, UM2, Montpellier, France
Damian Niwiński	University of Warsaw, Poland
Florin Gheorghe Filip	Romanian Academy, Romania
Janusz Kacprzyk	Polish Academy of Sciences, Poland
João Tavares	University of Porto, Portugal
Jon Hall	The Open University, UK
John MacIntyre	University of Sunderland, UK
Karl Stroetmann	Empirica Communication & Technology Research, Germany
Majed Al-Mashari	King Saud University, Saudi Arabia
Miguel-Angel Sicilia	University of Alcalá, Spain
Mirjana Ivanovic	University of Novi Sad, Serbia
Paulo Novais	University of Minho, Portugal
Wim Van Grembergen	University of Antwerp, Belgium
Mirjana Ivanovic	University of Novi Sad, Serbia
Reza Langari	Texas A&M University, USA
Wim Van Grembergen	University of Antwerp, Belgium

Program Committee

Abderrahmane Ez-Zahout	Mohammed V University, Morocco
Adriana Gradim	University of Aveiro, Portugal
Adriana Peña Pérez Negrón	Universidad de Guadalajara, Mexico
Adriani Besimi	South East European University, Macedonia
Agostinho Sousa Pinto	Polythecnic of Porto, Portugal
Ahmed El Oualkadi	Abdelmalek Essaadi University, Morocco
Akex Rabasa	University Miguel Hernandez, Spain
Alba Córdoba-Cabús	University of Malaga, Spain
Alberto Freitas	FMUP, University of Porto, Portugal
Aleksandra Labus	University of Belgrade, Serbia
Alessio De Santo	HE-ARC, Switzerland
Alexandru Vulpe	University Politehnica of Bucharest, Romania
Ali Idri	ENSIAS, University Mohamed V, Morocco
Alicia García-Holgado	University of Salamanca, Spain
Amélia Badica	Universti of Craiova, Romania
Amélia Cristina Ferreira Silva	Polytechnic of Porto, Portugal
Amit Shelef	Sapir Academic College, Israel
Alanio de Lima	UFC, Brazil
Almir Souza Silva Neto	IFMA, Brazil
Álvaro López-Martín	University of Malaga, Spain

Ana Carla Amaro	Universidade de Aveiro, Portugal
Ana Isabel Martins	University of Aveiro, Portugal
Anabela Tereso	University of Minho, Portugal
Anabela Gomes	University of Coimbra, Portugal
Anacleto Correia	CINAV, Portugal
Andrew Brosnan	University College Cork, Ireland
Andjela Draganic	University of Montenegro, Montenegro
Aneta Polewko-Klim	University of Białystok, Institute of Informatics, Poland
Aneta Poniszewska-Maranda	Lodz University of Technology, Poland
Angeles Quezada	Instituto Tecnologico de Tijuana, Mexico
Anis Tissaoui	University of Jendouba, Tunisia
Ankur Singh Bist	KIET, India
Ann Svensson	University West, Sweden
Anna Gawrońska	Poznański Instytut Technologiczny, Poland
Antoni Oliver	University of the Balearic Islands, Spain
Antonio Jiménez-Martín	Universidad Politécnica de Madrid, Spain
Aroon Abbu	Bell and Howell, USA
Arslan Enikeev	Kazan Federal University, Russia
Beatriz Berrios Aguayo	University of Jaen, Spain
Benedita Malheiro	Polytechnic of Porto, ISEP, Portugal
Bertil Marques	Polytechnic of Porto, ISEP, Portugal
Boris Shishkov	ULSIT/IMI-BAS/IICREST, Bulgaria
Borja Bordel	Universidad Politécnica de Madrid, Spain
Branko Perisic	Faculty of Technical Sciences, Serbia
Carla Pinto	Polytechnic of Porto, ISEP, Portugal
Carlos Balsa	Polythecnic of Bragança, Portugal
Carlos Rompante Cunha	Polytechnic of Bragança, Portugal
Catarina Reis	Polytechnic of Leiria, Portugal
Célio Gonçalo Marques	Polythenic of Tomar, Portugal
Cengiz Acarturk	Middle East Technical University, Turkey
Cesar Collazos	Universidad del Cauca, Colombia
Christine Gruber	K1-MET, Austria
Christophe Guyeux	Universite de Bourgogne Franche Comté, France
Christophe Soares	University Fernando Pessoa, Portugal
Christos Bouras	University of Patras, Greece
Christos Chrysoulas	London South Bank University, UK
Christos Chrysoulas	Edinburgh Napier University, UK
Ciro Martins	University of Aveiro, Portugal
Claudio Sapateiro	Polytechnic of Setúbal, Portugal
Cosmin Striletchi	Technical University of Cluj-Napoca, Romania
Costin Badica	University of Craiova, Romania

Cristian García Bauza	PLADEMA-UNICEN-CONICET, Argentina
Cristina Caridade	Polytechnic of Coimbra, Portugal
David Cortés-Polo	University of Extremadura, Spain
David Kelly	University College London, UK
Daria Bylieva	Peter the Great St.Petersburg Polytechnic University, Russia
Dayana Spagnuelo	Vrije Universiteit Amsterdam, Netherlands
Dhouha Jaziri	University of Sousse, Tunisia
Dmitry Frolov	HSE University, Russia
Dulce Mourato	ISTEC - Higher Advanced Technologies Institute Lisbon, Portugal
Edita Butrime	Lithuanian University of Health Sciences, Lithuania
Edna Dias Canedo	University of Brasilia, Brazil
Egils Ginters	Riga Technical University, Latvia
Ekaterina Isaeva	Perm State University, Russia
Eliana Leite	University of Minho, Portugal
Enrique Pelaez	ESPOL University, Ecuador
Eriks Sneiders	Stockholm University, Sweden
Esperança Amengual	Universitat de les Illes Balears, Spain
Esteban Castellanos	ESPE, Ecuador
Fatima Azzahra Amazal	Ibn Zohr University, Morocco
Fernando Bobillo	University of Zaragoza, Spain
Fernando Molina-Granja	National University of Chimborazo, Ecuador
Fernando Moreira	Portucalense University, Portugal
Fernando Ribeiro	Polytechnic Castelo Branco, Portugal
Filipe Caldeira	Polythecnic of Viseu, Portugal
Filipe Portela	University of Minho, Portugal
Filippo Neri	University of Naples, Italy
Firat Bestepe	Republic of Turkey Ministry of Development, Turkey
Francesco Bianconi	Università degli Studi di Perugia, Italy
Francisco García-Peñalvo	University of Salamanca, Spain
Francisco Valverde	Universidad Central del Ecuador, Ecuador
Frederico Branco	University of Trás-os-Montes e Alto Douro, Portugal
Galim Vakhitov	Kazan Federal University, Russia
Gayo Diallo	University of Bordeaux, France
Gema Bello-Orgaz	Universidad Politecnica de Madrid, Spain
George Suciu	BEIA Consult International, Romania
Ghani Albaali	Princess Sumaya University for Technology, Jordan

Gian Piero Zarri	University Paris-Sorbonne, France
Giovanni Buonanno	University of Calabria, Italy
Gonçalo Paiva Dias	University of Aveiro, Portugal
Goreti Marreiros	ISEP/GECAD, Portugal
Graciela Lara López	University of Guadalajara, Mexico
Habiba Drias	University of Science and Technology Houari Boumediene, Algeria
Hafed Zarzour	University of Souk Ahras, Algeria
Haji Gul	City University of Science and Information Technology, Pakistan
Hakima Benali Mellah	Cerist, Algeria
Hamid Alasadi	Basra University, Iraq
Hatem Ben Sta	University of Tunis at El Manar, Tunisia
Hector Fernando Gomez Alvarado	Universidad Tecnica de Ambato, Ecuador
Hector Menendez	King's College London, UK
Hélder Gomes	University of Aveiro, Portugal
Helia Guerra	University of the Azores, Portugal
Henrique da Mota Silveira	University of Campinas (UNICAMP), Brazil
Henrique S. Mamede	University Aberta, Portugal
Henrique Vicente	University of Évora, Portugal
Hicham Gueddah	University Mohammed V in Rabat, Morocco
Hing Kai Chan	University of Nottingham Ningbo China, China
Igor Aguilar Alonso	Universidad Nacional Tecnológica de Lima Sur, Peru
Inês Domingues	University of Coimbra, Portugal
Isabel Lopes	Polytechnic of Bragança, Portugal
Isabel Pedrosa	Coimbra Business School - ISCAC, Portugal
Isaías Martins	University of Leon, Spain
Issam Moghrabi	Gulf University for Science and Technology, Kuwait
Ivan Armuelles Voinov	University of Panama, Panama
Ivan Dunđer	University of Zabreb, Croatia
Ivone Amorim	University of Porto, Portugal
Jaime Diaz	University of La Frontera, Chile
Jan Egger	IKIM, Germany
Jan Kubicek	Technical University of Ostrava, Czech Republic
Jeimi Cano	Universidad de los Andes, Colombia
Jesús Gallardo Casero	University of Zaragoza, Spain
Jezreel Mejia	CIMAT, Unidad Zacatecas, Mexico
Jikai Li	The College of New Jersey, USA
Jinzhi Lu	KTH-Royal Institute of Technology, Sweden
Joao Carlos Silva	IPCA, Portugal

João Manuel R. S. Tavares	University of Porto, FEUP, Portugal
João Paulo Pereira	Polytechnic of Bragança, Portugal
João Reis	University of Aveiro, Portugal
João Reis	University of Lisbon, Portugal
João Rodrigues	University of the Algarve, Portugal
João Vidal Carvalho	Polythecnic of Coimbra, Portugal
Joaquin Nicolas Ros	University of Murcia, Spain
John W. Castro	University de Atacama, Chile
Jorge Barbosa	Polythecnic of Coimbra, Portugal
Jorge Buele	Technical University of Ambato, Ecuador
Jorge Gomes	University of Lisbon, Portugal
Jorge Oliveira e Sá	University of Minho, Portugal
José Braga de Vasconcelos	Universidade Lusófona, Portugal
Jose M Parente de Oliveira	Aeronautics Institute of Technology, Brazil
José Machado	University of Minho, Portugal
José Paulo Lousado	Polythecnic of Viseu, Portugal
Jose Quiroga	University of Oviedo, Spain
Jose Silvestre Silva	Academia Militar, Portugal
Jose Torres	Universidty Fernando Pessoa, Portugal
Juan M. Santos	University of Vigo, Spain
Juan Manuel Carrillo de Gea	University of Murcia, Spain
Juan Pablo Damato	UNCPBA-CONICET, Argentina
Kalinka Kaloyanova	Sofia University, Bulgaria
Kamran Shaukat	The University of Newcastle, Australia
Karima Moumane	ENSIAS, Morocco
Katerina Zdravkova	University Ss. Cyril and Methodius, North Macedonia
Khawla Tadist	Marocco
Khalid Benali	LORIA—University of Lorraine, France
Khalid Nafil	Mohammed V University in Rabat, Morocco
Korhan Gunel	Adnan Menderes University, Turkey
Krzysztof Wolk	Polish-Japanese Academy of Information Technology, Poland
Kuan Yew Wong	Universiti Teknologi Malaysia (UTM), Malaysia
Kwanghoon Kim	Kyonggi University, South Korea
Laila Cheikhi	Mohammed V University in Rabat, Morocco
Laura Varela-Candamio	Universidade da Coruña, Spain
Laurentiu Boicescu	E.T.T.I. U.P.B., Romania
Lbtissam Abnane	ENSIAS, Morocco
Lia-Anca Hangan	Technical University of Cluj-Napoca, Romania
Ligia Martinez	CECAR, Colombia
Lila Rao-Graham	University of the West Indies, Jamaica

Łukasz Tomczyk	Pedagogical University of Cracow, Poland
Luis Alvarez Sabucedo	University of Vigo, Spain
Luís Filipe Barbosa	University of Trás-os-Montes e Alto Douro
Luis Mendes Gomes	University of the Azores, Portugal
Luis Pinto Ferreira	Polytechnic of Porto, Portugal
Luis Roseiro	Polytechnic of Coimbra, Portugal
Luis Silva Rodrigues	Polythencic of Porto, Portugal
Mahdieh Zakizadeh	MOP, Iran
Maksim Goman	JKU, Austria
Manal el Bajta	ENSIAS, Morocco
Manuel Antonio Fernández-Villacañas Marín	Technical University of Madrid, Spain
Manuel Ignacio Ayala Chauvin	University Indoamerica, Ecuador
Manuel Silva	Polytechnic of Porto and INESC TEC, Portugal
Manuel Tupia	Pontifical Catholic University of Peru, Peru
Manuel Au-Yong-Oliveira	University of Aveiro, Portugal
Marcelo Mendonça Teixeira	Universidade de Pernambuco, Brazil
Marciele Bernardes	University of Minho, Brazil
Marco Ronchetti	Universita' di Trento, Italy
Mareca María PIlar	Universidad Politécnica de Madrid, Spain
Marek Kvet	Zilinska Univerzita v Ziline, Slovakia
Maria João Ferreira	Universidade Portucalense, Portugal
Maria José Sousa	University of Coimbra, Portugal
María Teresa García-Álvarez	University of A Coruna, Spain
Maria Sokhn	University of Applied Sciences of Western Switzerland, Switzerland
Marijana Despotovic-Zrakic	Faculty Organizational Science, Serbia
Marilio Cardoso	Polythecnic of Porto, Portugal
Mário Antunes	Polythecnic of Leiria & CRACS INESC TEC, Portugal
Marisa Maximiano	Polytechnic Institute of Leiria, Portugal
Marisol Garcia-Valls	Polytechnic University of Valencia, Spain
Maristela Holanda	University of Brasilia, Brazil
Marius Vochin	E.T.T.I. U.P.B., Romania
Martin Henkel	Stockholm University, Sweden
Martín López Nores	University of Vigo, Spain
Martin Zelm	INTEROP-VLab, Belgium
Mazyar Zand	MOP, Iran
Mawloud Mosbah	University 20 Août 1955 of Skikda, Algeria
Michal Adamczak	Poznan School of Logistics, Poland
Michal Kvet	University of Zilina, Slovakia
Miguel Garcia	University of Oviedo, Spain

Miguel Melo	INESC TEC, Portugal
Mihai Lungu	University of Craiova, Romania
Mircea Georgescu	Al. I. Cuza University of Iasi, Romania
Mirna Muñoz	Centro de Investigación en Matemáticas A.C., Mexico
Mohamed Hosni	ENSIAS, Morocco
Monica Leba	University of Petrosani, Romania
Nadesda Abbas	UBO, Chile
Narjes Benameur	Laboratory of Biophysics and Medical Technologies of Tunis, Tunisia
Natalia Grafeeva	Saint Petersburg University, Russia
Natalia Miloslavskaya	National Research Nuclear University MEPhI, Russia
Naveed Ahmed	University of Sharjah, United Arab Emirates
Neeraj Gupta	KIET group of institutions Ghaziabad, India
Nelson Rocha	University of Aveiro, Portugal
Nikola S. Nikolov	University of Limerick, Ireland
Nicolas de Araujo Moreira	Federal University of Ceara, Brazil
Nikolai Prokopyev	Kazan Federal University, Russia
Niranjan S. K.	JSS Science and Technology University, India
Noemi Emanuela Cazzaniga	Politecnico di Milano, Italy
Noureddine Kerzazi	Polytechnique Montréal, Canada
Nuno Melão	Polytechnic of Viseu, Portugal
Nuno Octávio Fernandes	Polytechnic of Castelo Branco, Portugal
Nuno Pombo	University of Beira Interior, Portugal
Olga Kurasova	Vilnius University, Lithuania
Olimpiu Stoicuta	University of Petrosani, Romania
Patricia Zachman	Universidad Nacional del Chaco Austral, Argentina
Paula Serdeira Azevedo	University of Algarve, Portugal
Paula Dias	Polytechnic of Guarda, Portugal
Paulo Alejandro Quezada Sarmiento	University of the Basque Country, Spain
Paulo Maio	Polytechnic of Porto, ISEP, Portugal
Paulvanna Nayaki Marimuthu	Kuwait University, Kuwait
Paweł Karczmarek	The John Paul II Catholic University of Lublin, Poland
Pedro Rangel Henriques	University of Minho, Portugal
Pedro Sobral	University Fernando Pessoa, Portugal
Pedro Sousa	University of Minho, Portugal
Philipp Jordan	University of Hawaii at Manoa, USA
Piotr Kulczycki	Systems Research Institute, Polish Academy of Sciences, Poland

Prabhat Mahanti	University of New Brunswick, Canada
Rabia Azzi	Bordeaux University, France
Radu-Emil Precup	Politehnica University of Timisoara, Romania
Rafael Caldeirinha	Polytechnic of Leiria, Portugal
Raghuraman Rangarajan	Sequoia AT, Portugal
Raiani Ali	Hamad Bin Khalifa University, Qatar
Ramadan Elaiess	University of Benghazi, Libya
Ramayah T.	Universiti Sains Malaysia, Malaysia
Ramazy Mahmoudi	University of Monastir, Tunisia
Ramiro Gonçalves	University of Trás-os-Montes e Alto Douro & INESC TEC, Portugal
Ramon Alcarria	Universidad Politécnica de Madrid, Spain
Ramon Fabregat Gesa	University of Girona, Spain
Ramy Rahimi	Chungnam National University, South Korea
Reiko Hishiyama	Waseda University, Japan
Renata Maria Maracho	Federal University of Minas Gerais, Brazil
Renato Toasa	Israel Technological University, Ecuador
Reyes Juárez Ramírez	Universidad Autonoma de Baja California, Mexico
Rocío González-Sánchez	Rey Juan Carlos University, Spain
Rodrigo Franklin Frogeri	University Center of Minas Gerais South, Brazil
Ruben Pereira	ISCTE, Portugal
Rui Alexandre Castanho	WSB University, Poland
Rui S. Moreira	UFP & INESC TEC & LIACC, Portugal
Rustam Burnashev	Kazan Federal University, Russia
Saeed Salah	Al-Quds University, Palestine
Said Achchab	Mohammed V University in Rabat, Morocco
Sajid Anwar	Institute of Management Sciences Peshawar, Pakistan
Sami Habib	Kuwait University, Kuwait
Samuel Sepulveda	University of La Frontera, Chile
Snadra Costanzo	University of Calabria, Italy
Sandra Patricia Cano Mazuera	University of San Buenaventura Cali, Colombia
Sassi Sassi	FSJEGJ, Tunisia
Seppo Sirkemaa	University of Turku, Finland
Shahnawaz Talpur	Mehran University of Engineering & Technology Jamshoro, Pakistan
Silviu Vert	Politehnica University of Timisoara, Romania
Simona Mirela Riurean	University of Petrosani, Romania
Slawomir Zolkiewski	Silesian University of Technology, Poland
Solange Rito Lima	University of Minho, Portugal
Sonia Morgado	ISCPSI, Portugal

Sonia Sobral	Portucalense University, Portugal
Sorin Zoican	Polytechnic University of Bucharest, Romania
Souraya Hamida	Batna 2 University, Algeria
Stalin Figueroa	University of Alcala, Spain
Sümeyya Ilkin	Kocaeli University, Turkey
Syed Asim Ali	University of Karachi, Pakistan
Syed Nasirin	Universiti Malaysia Sabah, Malaysia
Tatiana Antipova	Institute of Certified Specialists, Russia
Tatianna Rosal	Universtiy of Trás-os-Montes e Alto Douro, Portugal
Tero Kokkonen	JAMK University of Applied Sciences, Finland
The Thanh Van	HCMC University of Food Industry, Vietnam
Thomas Weber	EPFL, Switzerland
Timothy Asiedu	TIM Technology Services Ltd., Ghana
Tom Sander	New College of Humanities, Germany
Tomaž Klobučar	Jozef Stefan Institute, Slovenia
Toshihiko Kato	University of Electro-communications, Japan
Tuomo Sipola	Jamk University of Applied Sciences, Finland
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Valentim Realinho	Polythecnic of Portalegre, Portugal
Valentina Colla	Scuola Superiore Sant'Anna, Italy
Valerio Stallone	ZHAW, Switzerland
Vicenzo Iannino	Scuola Superiore Sant'Anna, Italy
Vitor Gonçalves	Polythecnic of Bragança, Portugal
Victor Alves	University of Minho, Portugal
Victor Georgiev	Kazan Federal University, Russia
Victor Hugo Medina Garcia	Universidad Distrital Francisco José de Caldas, Colombia
Victor Kaptelinin	Umeå University, Sweden
Viktor Medvedev	Vilnius University, Lithuania
Vincenza Carchiolo	University of Catania, Italy
Waqas Bangyal	University of Gujrat, Pakistan
Wolf Zimmermann	Martin Luther University Halle-Wittenberg, Germany
Yadira Quiñonez	Autonomous University of Sinaloa, Mexico
Yair Wiseman	Bar-Ilan University, Israel
Yassine Drias	University of Algiers, Algeria
Yuhua Li	Cardiff University, UK
Yuwei Lin	University of Roehampton, UK
Zbigniew Suraj	University of Rzeszow, Poland
Zorica Bogdanovic	University of Belgrade, Serbia

Contents

Reconstruction of Meteorological Records with PCA-Based Analog Ensemble Methods

Murilo M. Breve[1,2], Carlos Balsa[1,2], and José Rufino[1,2(✉)]

[1] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
{murilo.breve,balsa,rufino}@ipb.pt
[2] Laboratório para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal

Abstract. The Analog Ensemble (AnEn) method has been used to reconstruct missing data in time series with base on other correlated time series with full data. As the AnEn method benefits from the use of large volumes of data, there is a great interest in improving its efficiency. In this paper, the Principal Component Analysis (PCA) technique is combined with the classical AnEn method and a K-means cluster-based variant, within the context of reconstructing missing meteorological data at a particular station using information from neighboring stations. This combination allows to reduce the dimension of the number of predictor time series, while ensuring better accuracy and higher computational performance than the AnEn methods: it reduces prediction errors by up to 30% and achieves a computational speedup of up to 2x.

Keywords: Meteorological data reconstruction · Analogue ensemble · K-means clustering · Principal component analysis · MATLAB · R

1 Introduction

Information about past weather states is crucial to many scientific domains and practical applications. In the renewable energy field, for instance, it is vital to know the historical weather data and meteorological patterns, in order to estimate the productive potential of a given site, before making substantial financial investments [9]. However, full meteorological data may not always be available or may be absent altogether. In this scenario, data reconstruction techniques come into play. These should be numerically accurate and computationally efficient.

A well-known approach for meteorological data reconstruction is the Analog Ensemble (AnEn) method. Initially, it was used as a post-processing technique, to improve the accuracy of deterministic numerical forecast models [13]: past observations that are similar to the forecast are used to enhance the accuracy of

A. Rocha et al. (Eds.): WorldCIST 2023, LNNS 799, pp. 85–96, 2024.
https://doi.org/10.1007/978-3-031-45642-8_8

the forecast. The AnEn method can also be used directly for weather forecasting [6,18]. More recently [5], AnEn was used to reconstruct data of a meteorological variable by means of data from other variables at the same site, or based on data from the same or other variables from neighbor locations.

Compared to other machine learning methods, implementing the AnEn approach is considered relatively simple [1]. Concerning prediction assertiveness, a comparison [12] between AnEn and a Convolutional Neural Network (CNN), as post-processing methods of a Weather Research and Forecasting (WRF) model, showed that both methods improved equally the prediction accuracy. Similarly, in a homogeneous comparison [16] of the same methods, used as Empirical-statistical downscaling techniques, AnEn outperformed the CNN.

Large training datasets (historical observations from which missing values are derived) are advantageous for the AnEn methods: the more data is available, the easier it is to capture the variation tendency of the variable(s) to be reconstructed [7,16]. At the same time, more data entails more processing time. Hence, there is a lot of interest in improving the computational efficiency of these methods, while preserving (and ideally improving) their numerical accuracy. To this end, several variants of the AnEn methods have been investigated.

ClustAnEn (Cluster-based AnEn) is a variant of the AnEn method, based on K-means clustering, that is particularly efficient from a computational standpoint [3,4]. In this variant, there is a prior grouping of all feasible analogs, which allows selecting the analogs only by their group, instead of searching all possible analogs one by one. In addition to significantly reduce computational costs, this variant does not reduce the accuracy of the reconstructions.

Another approach that can be used to leverage the AnEn method is its combination with the Principal Components Analysis (PCA) technique. PCA is commonly used in multivariate statistics to minimize the datasets size while maintaining the most important information. Recently, PCA proved to be effective in the context of the reconstruction of meteorological data when combined with the AnEn method, originating the PCAnEn hybrid method [2].

This work consolidates previous investigations on the PCAnEn method and expands it further by applying the same rationale (integration with the PCA technique) to the ClustAnEn method (which results in the new PCClustAnEn variant). Thus, data coming from multiple predictor stations is reduced to one or two principal components (PCs) that are then used (instead of the original variables records) when applying the original AnEn method or its ClustAnEn variant, to reconstruct the missing data. The PCAnEn and PCClustAnEn techniques are also compared in numerical accuracy and computational efficiency.

The rest of the paper is organized as follows. Section 2 describes the dataset used and points out the correlations between meteorological variables and stations. Sections 3 and 4 introduce the PCA technique and combine it with the AnEn methods. Section 5 applies the new methods to reconstructs meteorological variables. Section 6 lays out final considerations and future work directions.

2 Meteorological Dataset

The data used in this paper for the reconstruction experiments originates from the US government National Data Buoy Center (NDBC) [14]. NBDC operates a network of data gathering buoys and coastal stations dispersed across various regions of the globe. This work uses data from various stations placed in the south of the Chesapeake Bay and surroundings (see Fig. 1). The predicted station is WDSV2 (name in red), and the predictor stations (name in white) are located within a 30 km radius from it (note: in the remaining of the paper, the suffix "V2" is omitted for simplicity). These stations are either at (buoys) or close to (coastal stations) sea level, and share similar climatological conditions.

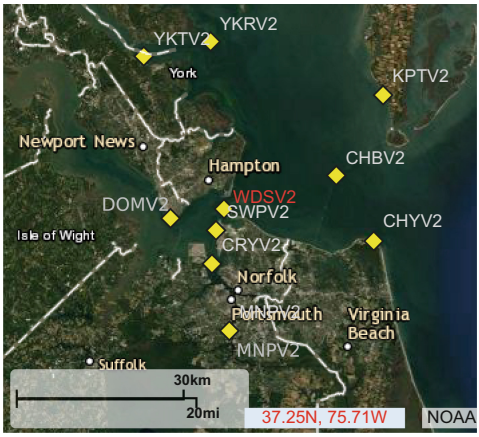


Fig. 1. Geolocation of the meteorological stations [14].

Several measurements or variables are taken at each station. These variables differ, depending on if the station is in a buoy or not. For this work, a common set of variables, available at all stations, was considered: atmospheric pressure ($PRES$) [mbar], air temperature ($ATMP$) [°C], wind speed ($WSPD$) [m/s] and peak gust speed (GST) [m/s]. Depending on the variable, the measurements are taken every 6 min or correspond to averages over a 6 min period.

Some basic properties of these variables may be seen in Table 1, namely the global average value in the dataset and the availability, for each station, between 2010 until the end of the year 2019. In this study, only variables with at least 85% of availability (in bold) were used. Moreover, because some stations are unable to comply with this degree of availability, the combination of stations used for each variable may be different.

To decide which variables will later be combined in the experiments, a preliminary study must be performed on the correlation between variables and stations. This is important because when variables are sufficiently correlated, the PCA method may be used to retain more data in fewer dimensions.

Table 1. Meteorological dataset characterization.

Station	**WSPD**		**GST**		**PRES**		**ATMP**	
	Mean	Avail.(%)	Mean	Avail.(%)	Mean	Avail.(%)	Mean	Avail.(%)
WDS	5.7	**97.5**	6.6	**97.5**	1017.4	**93.6**	16.5	**87.9**
YKR	5.9	**98.0**	6.9	**98.0**	1017.4	**98.6**	15.9	**98.5**
YKT	4.3	**97.7**	5.4	**97.7**	1017.3	**98.4**	16.0	**98.2**
MNP	2.6	**96.4**	4.1	**96.5**	1017.5	**97.9**	16.8	**97.7**
CHY	5.4	**95.5**	6.9	**95.5**	1017.0	31.1	16.1	**97.0**
DOM	3.9	**97.5**	5.3	**97.5**	1017.8	**98.3**	16.1	**98.2**
KPT	4.7	**97.4**	6.0	**97.5**	NA	0	NA	0
SWP	NA	0	NA	0	1017.7	**96.1**	NA	0
CRY	4.1	82.5	15.6	80.5	1017.6	82.8	16.5	34.3

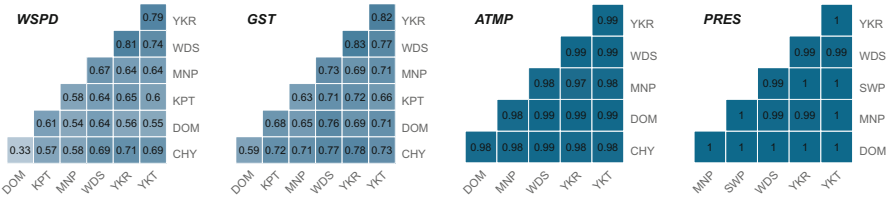


Fig. 2. Correlation between stations for each variable.

The correlations between the stations for each variable, relative to the same time period of Table 1, are shown in Fig. 2. In general, all stations are somehow correlated across all variables, with the $ATMP$ and $PRES$ variables showing the strongest correlations (1 or near 1).

Figure 3 presents the correlations between different meteorological variables within the same station. Due to the percentage of data available in the KPT and CHY stations, only records of two ($WSPD$ and GST) and three ($WSPD$, GST and $ATMP$) variables, respectively, are used for these stations. In the other stations, the records of all variables are available (see Table 1). The correlation between $WSPD$ and GST is strong across all stations. Between $ATMP$ and $PRES$, there is a minor inverse correlation. The other variables at different stations showed a weak correlation. Thus, only $WSPD$ and GST are used together in the experiments, once they are the only strongly correlated variables.

3 Determination of the Principal Components

Following the study on the correlation between variables and stations, the PCA approach may then be applied to reduce the dimensionality of the datasets.

Firstly, the dimensions with most data dispersion are identified. This enables to identify the principal components (PCs) that best distinguish the dataset

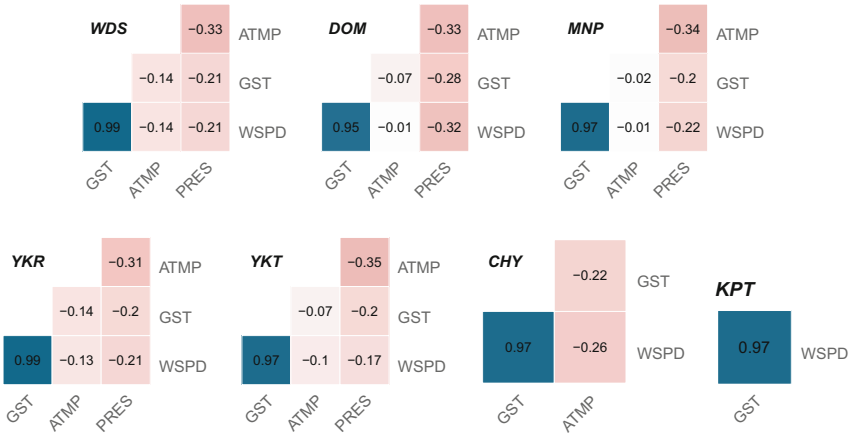


Fig. 3. Correlation between variables at each station.

under study. Consider that the dataset corresponding to the multiple predictor stations is represented by the data matrix $H \in \mathbb{R}^{m \times n}$, where each column h_i, with $i = 1, \ldots, n$, includes the scaled and normalized records of a single variable. Then, the thin singular value decomposition of H gives $H = U\Sigma V^T$, where $U \in \mathbb{R}^{m \times n}$, $\Sigma \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times n}$ (for details see [10]). The diagonal matrix Σ contains the singular values σ_i of H, for $i = 1, \ldots, n$, where $\sigma_1 > \sigma_2 > \ldots > \sigma_n$. The right singular vectors v_i are the *principal components directions* of H.

The vector $z_1 = Hv_1$ has the largest sample variance (σ_1^2/m) amongst all linear combinations of the columns of H, and so z_1 is the first principal component (PC_1). The second principal component (PC_2) is $z_2 = Hv_2$, once v_2 corresponds to the second largest variance (σ_2^2/m). The remaining principal components are defined similarly. The new variables are linear combinations of the columns of H, i.e., they are linear combinations of the normalized original variables $h_1, h_2, \ldots, h_n$, given by

$$z_j = v_{1j}h_1 + v_{2j}h_2 + \ldots + v_{nj}h_n \quad \text{for} \quad j = 1, 2, \ldots, n \tag{1}$$

where the coefficients v_{ij} (called *loadings*), with $i = 1, 2, \ldots, n$, are the elements of the vector v_j. The value of a coefficient is proportional to how significant a particular variable is in the principal component. It is expected that a few of the first principal components accurately reflect the original dataset, since they are likely to account for a significant proportion of the overall variation [17].

Figures 4 and 5 show the standard deviations of each PC for different amounts of input stations (# Stations). The standard deviation threshold of 1 is shown by a dotted line. PCs with standard deviation values above this line have more variance and, consequently, more information than the original normalized variables, whose standard deviation is equal to 1. Note that the variables from the WDS station were not included in the original variables because WDS was used only as the predicted station.

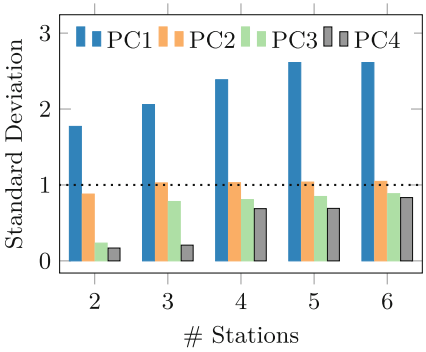
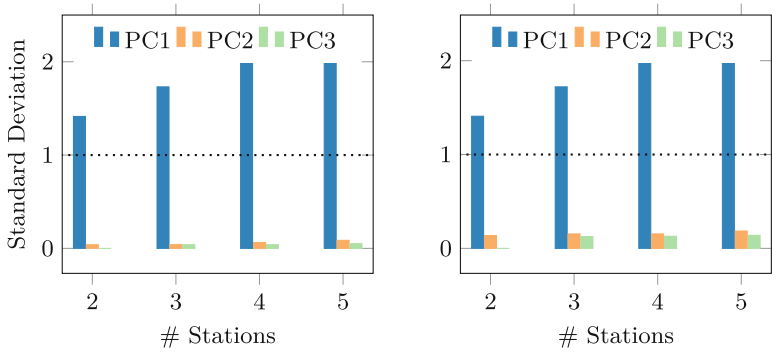


Fig. 4. Standard deviation of the *PC*s from the variables $WSPD$ and GST.



a) Variable $PRES$. b) Variable $ATMP$.

Fig. 5. Standard deviation of the *PC*s from the variables $PRES$ and $ATMP$.

In Fig. 4 the Principal Component Analysis is conducted from a data matrix that includes two meteorological variables, $WSPD$ and GST, from distinct stations. Except for the 2-station arrangement, both $PC1$ and $PC2$ exhibited values of standard deviation higher than one. These PCs with standard deviation greater than one were selected for the training phase.

The standard deviations of the PCs computed from the $PRES$ and $ATMP$ variables are shown in Fig. 5. It is crucial to highlight that, unlike the variables $WSPD$ and GST (see Fig. 4), the $PRES$ and $ATMP$ variables were examined individually because they do not correlate sufficiently (see Fig. 3). Both variables showed the same pattern of standard deviation values, indicating that the $PC1$ was enough to capture most information of the data in all configurations of stations. Furthermore, in contrast to the $WSPD$ and GST analysis, more information was contained in the $PC1$, since the relative concentration of standard deviation in this PC was significantly larger.

4 Analog Ensemble Combined with PCA

The AnEn method is able to reconstruct missing data in a time series. In this work, the time series builds on meteorological data and the reconstruction at a *predicted* station is carried out using data from neighbor *predictor* stations.

The process starts by identifying, in each predictor station, the predictor value for the same moment in time for which the predicted value must be reconstructed. Then, past historical values are found, in the predictor stations dataset, that are similar to the predictor value. These past values are called *analogs*. In the next step, the analogs are matched in time with corresponding observations (also historical measurements) of the predicted station. Finally, the missing value is predicted (reconstructed) by averaging the matching observations. The reconstruction error can then be evaluated if the real observed value for that instant is indeed available (as it is the case in this work).

The previous simplified description considers single numerical values for the predictor and analogs. In fact, these are vectors of $2k+1$ values (measurements), recorded at successive instants of the same time window, and $k > 0$ is an integer representing the width of each half-window (past and future) around the central instant. Therefore, analogs are established based on the similarity of vectors, and not single values, which enables the selection of analogs based on similar weather trends rather than single similar values [5]. The mapping of analogs (vectors) into observations (single values) is then based on the analogs central values.

In addition, when using multiple predictor stations, the analogs identified for each station may be required to overlap in time (dependent approach), or not (independent approach). In the current work, it was used the first alternative.

When the PCA technique is combined with the AnEn method (PCAnEn), the principal components (PCs) generated from the datasets of the predictor stations are used instead of the original datasets. This allows to use data from a larger number of stations, without increasing the computational effort.

The same idea may be applied to the K-means variant of the AnEn method. This variant enables to reduce the number of operations needed to determine the analogs of a given predictor. This is accomplished by replacing the comparison with all possible analogs, by a comparison with the clusters produced using the K-means clustering method. More precisely, the comparison is made with the centroids of the clusters, whose number is much less than the number of possible analog vectors (see [4] for details). Thus, similarly to what happens in the PCAnEn method, the PCClustAnEn method involves replacing the original predictor datasets with corresponding PCs prior to the clustering.

5 Experimental Evaluation

Both the PCAnEn and PCClustAnEn methods were put to the test for the reconstruction in the WDS station of the four variables selected for this study ($WSPD$, GST, $ATMP$ and $PRES$), every 6m, from 10m to 6pm period, during the full year of 2019 (prediction period). The remaining stations (within a 30 km

radius around WDS, and with more than 85% of data available), were used as predictor stations, considering the training period of 2011 to 2018.

The reconstruction was performed by two separate implementations of the methods, one in R [15] and another in MATLAB [11]. This allowed for the mutual verification of the numerical results and provided an opportunity to compare the implementations performance-wise. The computing system used to execute the methods was a KVM-based virtual machine (with 16 virtual cores of a Intel Xeon W-2195 CPU, 64 GB of RAM and 256 GB of SSD) hosted on the CeDRI cluster, running Linux Ubuntu 20.04 LTS, R 4.2.2 and MATLAB R2021a.

Besides testing the PCAnEn and PClustAnEn methods, the corresponding non-PCA variants (AnEn and ClustAnEn) applied to the original datasets were also tested. This way, the specific impact of the PCA technique may also be accessed. To make the comparison fair, AnEn and ClustAnEn were tested using as predictors the same two stations used to test the PCAnEn and PClustAnEn with a 2-station configuration (#Stations = 2). This means that the variable is predicted from the same variable located in the two closest stations, thus ensuring the most favourable configuration to the AnEn and ClustAnEn methods.

The accuracy of the predicted/reconstructed values is assessed by comparing them to the exact values recorded at station WDS during the prediction period. The comparison is done by means of the Root Mean Square (RMSE) error that measures, simultaneously, the systematic and the random error [8].

Table 2. RMSE of the reconstruction with different methods.

Method	# Stations	**MATLAB**				**R**			
		WSPD	*GST*	*ATMP*	*PRES*	*WSPD*	*GST*	*ATMP*	*PRES*
PCAnEn	2	1.65	1.95	1.01	0.51	1.65	1.97	1.01	0.51
	3	1.32	1.52	0.84	0.48	1.32	1.52	0.84	0.48
	4	1.27	1.46	0.78	**0.45**	1.27	1.45	0.78	**0.45**
	5	**1.19**	**1.36**	**0.71**	0.61	**1.19**	**1.36**	**0.71**	0.61
	6	1.24	1.42	—	—	1.24	1.44	—	—
AnEn	2	1.68	1.77	0.86	0.59	1.67	1.76	0.86	0.59
PCClustAnEn	2	1.65	1.95	1.01	0.52	1.65	1.95	1.01	0.52
	3	1.32	1.50	0.84	0.48	1.32	1.51	0.84	0.48
	4	1.27	1.45	0.78	**0.45**	1.28	1.45	0.78	**0.45**
	5	**1.20**	**1.35**	**0.72**	0.61	**1.20**	**1.36**	**0.72**	0.60
	6	1.27	1.40	—	—	1.25	1.42	—	—
ClustAnEn	2	1.69	1.73	0.88	0.60	1.69	1.74	0.87	0.56

Table 2 presents the RMSE values for all tests performed. For each test, the number of *PC*s used was 1 or 2, after the values of the respective standard deviation, as explained in Sect. 3. Between PCAnEn and PCClustAnEn, there were no noteworthy changes in accuracy. The 5-station setup demonstrated a

lower RMSE than the non-PCA approaches for the majority of variables. The higher errors are obtained with the 2-station configurations, in which case there's no sensible advantage in using the PCA variants over the non-PCA ones. The reductions in error rates from the PCA implementations ranged from ≈18% to ≈30%, for the best setting of each variable, compared to the non-PCA methods. These considerations apply to both implementations (R and MATLAB).

Regarding the computational performance, Fig. 6 shows the processing times of the MATLAB (M) and R (R) codes, with different amounts of stations, for the reconstruction of the $WSPD$ variable, using all the CPU cores (16) available in the test bed computational system. The $WSPD$ variable was chosen for the performance evaluation because a) it is available for more stations (recall Table 2), and b) it requires 2 PCs to represent the original variables when using 3 or more stations. The same is also valid for the GST variable, whose processing times are either the same (PCA-based approaches) or similar (other approaches).

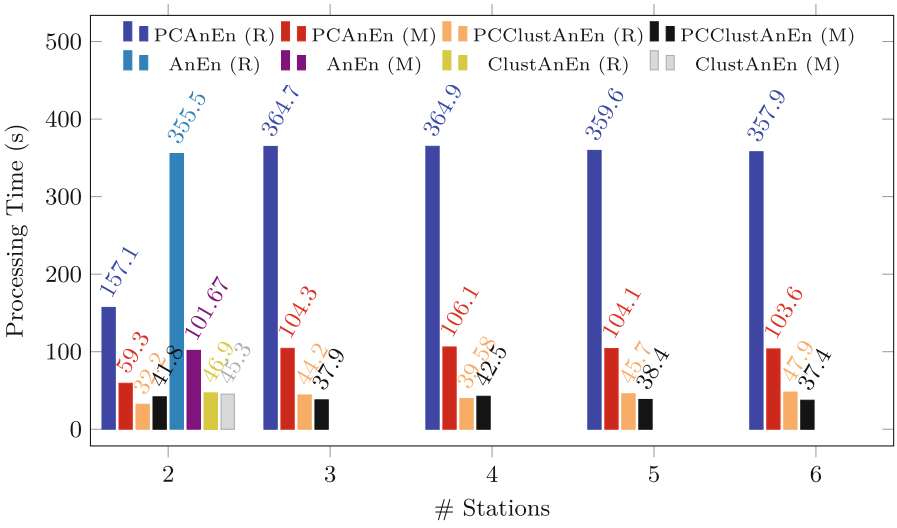


Fig. 6. Reconstruction time of $WSPD$ with 16 cores (2 to 6 stations).

When using clustering (ClustAnEn and PCClustAnEn), the reconstruction times are the lowest, and the variations are small for different numbers of stations, whether using PCA or not; also, for the only scenario where it makes sense to use the non-PCA approaches (2 stations), using clustering alone (ClustAnEn) is slower than combining it with PCA (PCClustAnEn).

Without clustering (AnEn and PCAnEn), the processing times are noticeably higher. When applying PCA (PCAnEN), the highest times are obtained with 3 or more stations (using 2 PCs), and they are similar; these times roughly double the time with 2 stations (using 1 PC); thus, without clustering, the number of PCs used has a noticeable influence (direct proportionality) on the processing

times. For the 2-stations scenario, not using PCA (AnEn) doubles the processing times compared to using PCA (PCAnEn), being equivalent to using PCA with more than 2 stations, once it is also using two time series.

Lowering the processing times is important, but it shouldn't be at the expense of higher reconstruction errors. Ideally, the reconstruction should be faster and also more accurate. The smallest RMSE errors for the $WSPD$ variable are obtained with PCA-based methods using 5 stations, whether clustering is used (PCClustAnEn) or not (PCAnEn) – recall Table 2. However, clustering ensures much lower processing times, with a speedup between $\gtrapprox$2,7 (MATLAB code) and $\gtrapprox$7,8 (R code). Comparing the processing times of PCClustAnEn with 5 stations, with the ones of ClustAnEn with 2 stations (the best provided by not using PCA) yields almost none speedup (46,9/45,7 = 1,03 and 45,3/38,4 = 1,18); however, the RMSE error of PCClustAnEn with 5 stations is only 1,2/1,69≈70% of the error of ClustAnEn with 2 stations, thus favouring the first approach.

The impact on performance of using or not the PCA method is perceivable in the 2-stations scenario. Here, using PCA provides speedups ranging from 2,26 to 1,08, for comparable methods (AnEn vs PCAnEn, and ClustAnEn vs PCClustAnEn).

Another advantage of adding PCA emerges when two variables, like $WPSD$ and GST, are used together in the analysis. Once they share the same time series, PCA-based methods can predict both variables in a single run, unlike the non-PCA approaches, which would require two runs of the reconstruction code.

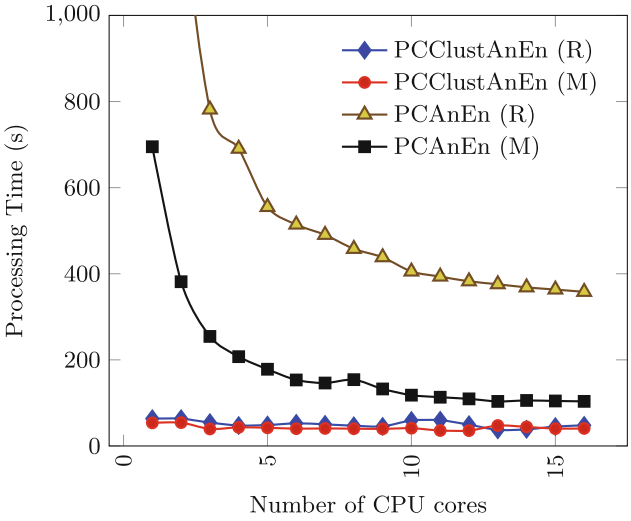


Fig. 7. Reconstruction time of $WSPD$ with 1 to 16 cores (6 stations).

The MATLAB code was found consistently faster than the R code. This is visible in Fig. 6 for 16 cores, and can also be seen in Fig. 7 for a variable number of cores. However, under PCClustAnEn the differences were minor, meaning

both implementations are equally efficient when applying the K-means clustering. More important, PCClustAnEn required much less processing times, in all configurations, compared to PCAnEn. Also, PCClustAnEn mostly doesn't benefit from the extra cores, in opposition to the PCAnEn method, where the search for analogues is the biggest code hotspot and is easily parallelizable.

It should also be stressed that both R and MATLAB were used with default configurations, without any extra performance tuning to optimize their behavior.

6 Conclusion

This paper describes the combination of Analog Ensemble (AnEn) methods with Principal Component Analysis (PCA), allowing for data from several stations to be reduced to a smaller number of time series, corresponding to the Principal Components (*PC*s). These are then used, instead of the original variables records, to reconstruct data missing in the records of a meteorological site.

In our experiments, the PCA technique improved the assertiveness of prediction without compromising the computational performance, since it is possible to increase the number of stations without increasing the quantity of input time series. It was also shown that the efficacy of PCA is heavily influenced by the correlation between the time series of several predictors, as higher correlation allows for a high proportion of information/variance in the first components.

Furthermore, two different implementations of the methods studied were used and compared, one in MATLAB and other in R. This allowed to double-check the numerical results and gain insight on the potential performance impact of choosing either implementation. The scalability of both codes was also studied, in a medium-scale multicore system. The performance evaluation showed the superiority of the AnEn methods where PCA is combined with clustering.

Future applications of the same methodology with larger datasets are planned, to assess more accurately the effects of the quantity, correlation and closeness of predictor stations. The R code will also be tuned to improve its performance.

Acknowledgements. This work was supported by national funds through FCT/MCTES (PIDDAC): CeDRI, UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and UIDP/05757/2020 (DOI: 10.54499/UIDB/05757/2020); and SusTEC, LA/P/0007/2020 (DOI: 10.54499/LA/P/0007/2020).

References

1. Alessandrini, S.: Predicting rare events of solar power production with the analog ensemble. Solar Energy **231**, 72–77 (2022). https://doi.org/10.1016/j.solener.2021.11.033. https://www.sciencedirect.com/science/article/pii/S0038092X21009920
2. Balsa, C., Breve, M.M., André, B., Rodrigues, C.V., Rufino, J.: PCAnEn - Hindcasting with Analogue Ensembles of Principal Components. In: Garcia, M.V., Gordón-Gallegos, C. (eds.) CSEI 2022. LNNS, vol. 678, pp. 169–183. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-30592-4_13

3. Balsa, C., Rodrigues, C.V., Araújo, L., Rufino, J.: Hindcasting with cluster-based analogues. In: Guarda, T., Portela, F., Santos, M.F. (eds.) ARTIIS 2021. CCIS, vol. 1485, pp. 346–360. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-90241-4_27
4. Balsa, C., Rodrigues, C.V., Araújo, L., Rufino, J.: Cluster-based analogue ensembles for hindcasting with multistations. Computation **10**(6), 91 (2022). https://doi.org/10.3390/computation10060091
5. Balsa, C., Rodrigues, C.V., Lopes, I., Rufino, J.: Using analog ensembles with alternative metrics for hindcasting with multistations. ParadigmPlus **1**(2), 1–17 (2020). https://journals.itiud.org/index.php/paradigmplus/article/view/11
6. Birkelund, Y., Alessandrini, S., Byrkjedal, Ø., Monache, L.D.: Wind power prediction in complex terrain using analog ensembles. J. Phys. Conf. Ser. **1102**(1), 012,008 (2018). https://doi.org/10.1088/1742-6596/1102/1/012008. https://dx.doi.org/10.1088/1742-6596/1102/1/012008
7. Castellano, C.M., DeGaetano, A.T.: Downscaling extreme precipitation from cmip5 simulations using historical analogs. J. Appl. Meteorol. Climatol. **56**(9), 2421 – 2439 (2017). https://doi.org/10.1175/JAMC-D-16-0250.1. https://journals.ametsoc.org/view/journals/apme/56/9/jamc-d-16-0250.1.xml
8. Chai, T., Draxler, R.R.: Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geosci. Model Dev. **7**(3), 1247–1250 (2014). https://doi.org/10.5194/gmd-7-1247-2014
9. Davò, F., Alessandrini, S., Sperati, S., Monache, L.D., Airoldi, D., Vespucci, M.T.: Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting. Sol. Energy **134**, 327–338 (2016). https://doi.org/10.1016/j.solener.2016.04.049
10. Eldén, L.: Matrix methods in data mining and pattern recognition. SIAM, Philadelphia, PA, USA (2007)
11. MATLAB: version 9.10.0.1602886 (R2021a). The MathWorks Inc., Natick, Massachusetts (2021)
12. Meech, S., Alessandrini, S., Chapman, W., Delle Monache, L.: Post-processing rainfall in a high-resolution simulation of the 1994 piedmont flood. Bulletin of Atmospheric Science and Technology **1**(3), 373–385 (2020). https://doi.org/10.1007/s42865-020-00028-z
13. Monache, L.D., Eckel, F.A., Rife, D.L., Nagarajan, B., Searight, K.: Probabilistic weather prediction with an analog ensemble. Mon. Weather Rev. **141**(10), 3498–3516 (2013). https://doi.org/10.1175/mwr-d-12-00281.1
14. National Weather Service: National Data Buoy Center. https://www.ndbc.noaa.gov
15. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2022). https://www.R-project.org/
16. Rozoff, C.M., Alessandrini, S.: A comparison between analog ensemble and convolutional neural network empirical-statistical downscaling techniques for reconstructing high-resolution near-surface wind. Energies **15**(5) (2022). https://doi.org/10.3390/en15051718. https://www.mdpi.com/1996-1073/15/5/1718
17. Spence, L., Insel, A., Friedberg, S.: Elementary Linear Algebra: A matrix Approach. Pearson Education Limited (2013)
18. Zhang, X., Li, Y., Lu, S., Hamann, H.F., Hodge, B.M., Lehman, B.: A solar time based analog ensemble method for regional solar power forecasting. IEEE Trans.on Sustain. Energy **10**(1), 268–279 (2019). https://doi.org/10.1109/TSTE.2018.2832634