XJENZA
ONLINE

www.xjenza.org

Editor-in-Chief: Cristiana Sebu

Chronological List of Past and Present Editors of Xjenza The Journal of the Malta Chamber of Scientists

2018–

Editor: Cristiana Sebu

Senior Editors: Sebastiano D'Amico, David Magri

Associate Editors: Sandro Lanfranco, Ian Thornton, Gianluca Valentino, Ian Cassar, Alexandra Bonnici, Joseph Galea, Pierre Vella, Lourdes Farrugia, Godfrey Baldacchino, Liberato Camilleri

Xjenza Online Vol. 6 Iss. 2 - 2018

Xjenza Online Vol. 6 Iss. 1 - 2018

2013–2017

Editor: Giuseppe Di Giovanni

Associate Editors: David Magri, Ian Thornton, Ian Cassar, Philip Farrugia, Sebastiano D'Amico, Nicholas Sammut, David Mifsud, Godfrey Baldacchino, Liberato Camilleri, Carmel Cefai

Xjenza Online Vol. 5 Iss. 2 - 2017

Xjenza Online Vol. 5 SI MNS Proceedings - 2017

Xjenza Online Vol. 5 Iss. 1 - 2017

Xjenza Online Vol. 5 Virtual Issue COST - 2017

Xjenza Online Vol. 4 Iss. 2 - 2016

Xjenza Online Vol. 4 Iss. 1 - 2016

Xjenza Online Vol. 3 Iss. 2 - 2015

Associate Editors: David Magri, Ian Thornton, Ian Cassar, Philip Farrugia, Sebastiano D'Amico, Nicholas Sammut, Joseph Galea, David Mifsud, Sandro Lanfranco, Mario Valentino, Godfrey Baldacchino, Liberato Camilleri

Xjenza Online Vol. 3 Iss. 1 - 2015

Xjenza Online Vol. 2 Iss. 2 - 2014

Xjenza Online Vol. 2 Iss. 1 - 2014

Xjenza Online Vol. 1 Iss. 2 - 2013

Xjenza Online Vol. 1 Iss. 1 - 2013

2003–2007

Editors: Joseph N. Grima and Richard Muscat

Xjenza Vol. 12 - 2007

Xjenza Vol. 11 - 2006

Xjenza Vol. 10 - 2005

Xjenza Vol. 9 - 2004

Xjenza Vol. 8 - 2003

1996–2002

Editor: Angela Xuereb

Associate Editor: Richard Muscat

Xjenza Vol. 7 - 2002

Xjenza Vol. 6 - 2001

Associate Editors: Martin Ebejer and Richard Muscat

Xjenza Vol. 5 - 2000

Xjenza Vol. 4 Iss. 2 - 1999

Xjenza Vol. 4 Iss. 1 - 1999

Associate Editors: Martin Ebejer, Richard Muscat, and Christian A. Scerri

Xjenza Vol. 3 Iss. 2 - 1998

Xjenza Vol. 3 Iss. 1 - 1998

Associate Editors: Martin Ebejer, Richard Muscat, Christian A. Scerri and Emmanuel Sinagra

Xjenza Vol. 2 Iss. 2 - 1997

Xjenza Vol. 2 Iss. 1 - 1997

Xjenza Vol. 1 Iss. 2 - 1996

Xjenza Vol. 1 Iss. 1 - 1996

Scope of Journal

Xjenza Online is the Science Journal of the Malta Chamber of Scientists and is published in an electronic format. Xjenza Online is a peer-reviewed, open access international journal. The scope of the journal encompasses research articles, original research reports, reviews, short communications and scientific commentaries in the fields of: mathematics, statistics, geology, engineering, computer science, social sciences, natural and earth sciences, technological sciences, linguistics, industrial, nanotechnology, biology, chemistry, physics, zoology, medical studies, electronics and all other applied and theoretical aspect of science.

The first printed issue of the journal was published in 1996 and the last (Vol. 12) in 2007. The publication of Xjenza was then ceased until 2013 when a new editorial board was formed with internationally recognised scientists, and Xjenza was relaunched as an online journal, with two issues being produced every year. One of the aims of Xjenza, besides highlighting the exciting research being performed nationally and internationally by Maltese scholars, is to provide a launching platform into scientific publishing for a wide scope of potential authors, including students and young researchers, into scientific publishing in a peer-reviewed environment.

Instructions for Authors

Xjenza is the Science Journal of the Malta Chamber of Scientists and is published by the Chamber in electronic format on the website: http://www.mcs.org.mt/index.php/xjenza. Xjenza will consider manuscripts for publication on a wide variety of scientific topics in the following categories

1. Research Articles
2. Communications
3. Review Articles
4. Notes
5. Research Reports
6. Commentaries
7. News and Views
8. Invited Articles and Special Issues
9. Errata

Research Articles form the main category of scientific papers submitted to Xjenza. The same standards of scientific content and quality that applies to Communications also apply to Research Articles.

Communications are short peer-reviewed research articles (limited to three journal pages) that describe new important results meriting urgent publication. These are often followed by a full Research Article.

Review Articles describe work of interest to the wide community of readers of Xjenza. They should provide an in-depth understanding of significant topics in the sciences and a critical discussion of the existing state of knowledge on a topic based on primary literature sources. Review Articles should not normally exceed 6000 words. Authors are strongly advised to contact the Editorial Board before writing a Review.

Notes are fully referenced, peer-reviewed short articles limited to three journal pages that describe new theories, concepts and developments made by the authors in any branch of science and technology. Notes need not contain results from experimental or simulation work.

Research Reports are extended reports describing research of interest to a wide scientific audience characteristic of Xjenza. Please contact the editor to discuss the suitability of topics for Research Reports.

Commentaries: Upon Editor's invitation, commentaries discuss a paper published in a specific issue and should set the problems addressed by the paper in the wider context of the field. Proposals for Commentaries may be submitted; however, in this case authors should only send an outline of the proposed paper for initial consideration. The contents of the commentaries should follow the following set of rules: 3000 words maximum, title 20 words maximum, references 10 maximum (including the article discussed) and figures/tables 2 maximum.

News and Views: The News section provides a space for articles up to three journal pages in length describing leading developments in any field of science and technology or for reporting items such as conference reports. The Editor reserves the right to modify or reject articles for consideration as News.

Invited Articles and Special Issues: Xjenza regularly publishes Invited Articles and Special Issues that consist of articles written at the invitation of the Editor or another member of the editorial board.

Errata: Xjenza also publishes errata, in which authors correct significant errors of substance in their published manuscripts. The title should read: Erratum: "Original title" by ***, Xjenza, vol. *** (year). Errata should be short and consistent for clarity.

Submission of Manuscripts

Manuscripts should be sent according to the guidelines given hereafter to xjenza@mcs.org.mt.

Referees All manuscripts submitted to Xjenza are peer reviewed. Authors are requested to submit with their manuscript the names and addresses of three referees, preferably from overseas. Every effort will be made to use the recommended reviewers; however the editor reserves the right to also consult other competent reviewers.

Conflict of Interest Authors are expected to disclose any commercial or other types of associations that may pose a conflict of interest in connection to with the submitted manuscript. All funding sources supporting the work, and institutional or corporate affiliations of the authors, should be acknowledged on the title page or at the end of the article.

Policy and Ethics The work presented in the submitted manuscript must have been carried out in compliance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans (`http://www.wma.net/en/30publications/10policies/b3/index.html`); EU Directive 2010/63/EU for animal experiments (`http://ec.europa.eu/environment/chemicals/lab_animals/legislation_en.htm`); Uniform Requirements for manuscripts submitted to Biomedical journals (`http://www.icmje.org`). This must be stated at an appropriate point in the article.

Submission, Declaration and Verification Author(s) must only submit work that has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that is not under consideration for publication elsewhere, that has been approved for publication by all authors, and tacitly or explicitly, by the responsible authorities where the work was carried out, and that, if accepted, will not be published elsewhere in the same form, in English or in any other language, including electronically, without the written consent of the copyright-holder.

Permissions It is the responsibility of the corresponding author of a manuscript to ensure that there is no infringement of copyright when submitting material to Xjenza. In particular, when material is copied from other sources, a written statement is required from both the author and/or publisher giving permission for reproduction. Manuscripts in press, unpublished data and

personal communications are discouraged; however, corresponding authors are expected to obtain permission in writing from at least one author of such materials.

Preparation of Manuscripts

Xjenza accepts submissions in MS Word, Libre Office Writer and LATEX, the latter being the preferred option. Anyone submitting in LATEX should use the journal template, the latest version of which can be found at `http://github.com/hicklin/Xjenza-Journal-Template`. All the necessary files to run the LATEX document should be supplied together with the rendered PDF.

If a word processor is used the styling should be kept to a minimum. Bold face and italic fonts, as well as subscript and superscript text may be used as required by the context. Text should be in single-column format and the word processor options should not be used in order to justify text or hyphenate words. Alongside the native format of the word processer, a PDF file, generated by the word processor, must be provided. Furthermore, artwork should be in accordance with the artwork guidelines given below and must be submitted separately from the word processor file. Similarly, the bibliographic data of the cited material should be submitted separately as an Endnote (*.xml), Research Information Systems (*.ris), Zotero Library (zotero.splite) or a BIBTEX (*.bib) file.

Article Structure

A manuscript for publication in Xjenza will typicall have the following components: Title page, Abstract, Keywords, Abbreviations, Introduction, Materials and Methods, Results, Discussion, Conclusions, Appendices and References.

The manuscript will be divided into clearly defined and numbered sections. Each numbered subsection should have a brief heading. Each heading should appear on its own separate line. Subsections should be used as much as possible when cross-referencing text, i.e. refer to the subsection by the section number.

Title page
- The title should be concise yet informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.
- Author names and affiliations. Indicate the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lower-case superscript number immediately after each author's name and in front of the appropriate address. Provide the full postal address of each affiliation, including the country name and, if available, the e-mail address.
- Corresponding author. Clearly indicate who will handle correspondence at all stages of refereeing and publication, including post-publication. Ensure that telephone and fax numbers (with country and area code) are provided in addition to the e-mail address and complete postal address. Contact details must be kept up to date by the corresponding author.
- Present/permanent address. If an author has changed the address since the work described, this can be indicated as a footnote to the author's name. The address at which the author actually did the work must be retained as the main, affiliation address. Superscript Arabic numerals are used for such footnotes.

Abstract A concise and factual abstract is required of up to about 250 words. The abstract should state briefly the background and purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, references and non-standard abbreviations should be avoided. If essential, these must be defined at first mention in the abstract itself.

Abbreviations Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at their first mention as well as in the footnote and should be used consistently throughout the text.

Introduction State the objectives of the work and provide an adequate background, avoid a detailed literature survey or a summary of the results.

Materials and Methods Provide sufficient detail to allow the work to be reproduced. Methods already published should be indicated by a reference: only relevant modifications should be described.

Results Results should be clear and concise. Numbered/tabulated information and/or figures should also be included.

Discussion This should explore the significance of the results of the work, yet not repeat them. Avoid extensive citations and discussion of published literature. A combined section of Results and Discussion is often appropriate.

Conclusions The main conclusions based on results of the study may be presented in a short Conclusions section. This may stand alone or form a subsection of a Discussion or Results and Discussion section.

Appendices Formulae and equations in appendices should be given separate numbering: Eq. (A.1), Eq. (A.2), etc.; in a subsequent appendix, Eq. (B.1) and so on. Similarly for tables and figures: Table A.1; Fig. A.1, etc.

Acknowledgements Collate acknowledgements in a separate section at the end of the article before the references. Do not include them on the title page, as a footnote to the title or otherwise. List here those individuals who provided assistance during the research (e.g., providing language help, writing assistance or proof reading the article, etc.).

Units Follow internationally accepted rules and conventions: use the international system of units (SI). If other units are mentioned, please give their equivalent in SI. Anyone using LATEX should use the package siunitx in all cases.

Footnotes Footnotes should be used sparingly. Number them consecutively throughout the article, using superscript Arabic numbers. Many word processors build footnotes into the text, and this feature may be used. Should this not be the case, indicate the position of footnotes by a superscript number in the text and list the footnotes separately at the end of the article. Do not include footnotes in the Reference list.

Table Footnotes Indicate each footnote in a table with a superscript lower case letter.

Artwork Electronic artwork general instructions:
- Make sure you use uniform lettering and sizing of your original artwork.
- Save text in illustrations as 'graphics' or enclose the font.
- Only use the following fonts in your illustrations: Arial, Courier, Times, Symbol or Computer Modern Roman, the latter is preferred.
- Number the illustrations according to their sequence in the text.
- Name your artwork files as 'fig$x$' or 'tab$x$' where x corresponds to the sequence number in your document.
- Provide captions to illustrations separately.
- Produce images near to the desired size of the printed version or grater.
- Make sure that the artwork has no margins and borders.

- Submit each figure as a separate file.

Formats Regardless of the application used, when your electronic artwork is finalised its file format should be one of the following (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):

- PDF or SVG: Vector drawings. Embed the font or save the text as 'graphics'.
- JPEG or PNG: Color or grayscale photographs (halftones): always use a minimum of 300 dpi.
- JPEG or PNG: Bitmapped line drawings: use a minimum of 1000 dpi.
- JPEG or PNG: Combinations bitmapped line/half-tone (color or grayscale): a minimum of 500 dpi is required.

Where possible use a vector format for your artwork (PDF or SVG). If this is not possible, supply files that have and adequate resolution.

Colour Artwork Make sure that color artwork files are in an acceptable format (JPEG, PNG, PDF or SVG) and have the correct resolution.

Figure Captions Ensure that each illustration has a caption. Supply captions separately, not attached to the figure. A caption should comprise a brief title (not on the figure itself) and a description of the illustration. Keep text in the illustrations themselves to a minimum, but explain all symbols and abbreviations used.

Tables Number tables consecutively in accordance with their appearance in the text. Place footnotes to tables below the table body and indicate them with superscript lowercase letters. Avoid vertical rules. Be moderate with the use of tables and ensure that the data presented in tables do not duplicate results described elsewhere in the article. Large tables should be submitted in CSV format.

Citations and References Reference and citation styles for manuscripts submitted to Xjenza should be in accordance to the APA v6 style.

Citation in text References to cited literature in the text should be given in the form of an author's surname and the year of publication of the paper with the addition of a letter for references to several publications of the author in the same year. For further information regarding multiple authors consult the APA v6 guidelines. Citations may be made directly

Kramer et al. (2010) have recently shown . . .

or parenthetically

as demonstrated (Allan, 2000a, 2000b, 1999; Allan and Jones, 1999).

Groups of references should be listed first alphabetically, then chronologically. When writing in LaTeX use \textcite{} and \parencite{} for the respective cases mentioned.

The reference section Every reference cited in the text should also be present in the reference list (and vice versa). The reference list should also be supplied as an Endnote (*.xml), Research Information Systems (*.ris), Zotero Library (zotero.splite) or a BiBTeX (*.bib) file. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

References should be arranged first alphabetically and then further sorted chronologically if necessary. More than one reference from the same author(s) in the same year must be identified by the letters 'a', 'b', 'c', etc., placed after the year of publication.

Consult the APA v6 guidelines for multiple authors. Below are some examples of referencing different bibliographic material.

Reference to a Journal Publication:

Agree, E. M. and Freedman, V. A. (2011). A Quality-of-Life Scale for Assistive Technology: Results of a Pilot Study of Aging and Technology. *Phys. Ther.*, 91(12):1780–1788.

McCreadie, C. and Tinker, A. (2005). The acceptability of assistive technology to older people. *Ageing Soc.*, 25(1):91–110.

Reference to a Book:

Brownsell, B. (2003). *Assistive Technology and Telecare: Forging Solutions for Independent Living.* Policy Press, Bristol.

Fisk, M. J. (2003). *Social Alarms to Telecare: Older People's Services in Transition.* Policy Press, Bristol, 1st edition.

Reference to a Chapter in an Edited Book:

Brownsell, S. and Bradley, D. (2003). New Generations of Telecare Equipment. In *Assist. Technol. Telecare Forg. Solut. Indep. Living*, pages 39–50.

Web references The full URL should be given together with the date the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately or can be included in the reference list.

References in a Special Issue Please ensure that the words 'this issue' are added to any references in the list (and any citations in the text) to other articles in the same Special Issue.

Journal Abbreviations Journal names should be abbreviated according to:

-Index Medicus journal abbreviations: `http://www.nlm.nih.gov/tsd/serials/lji.html`;

-List of title word abbreviations: `http://www.issn.org/2-22661-LTWA-online.php`;

-CAS (Chemical Abstracts Service): `http://www.cas.org/sent.html`.

Video data Xjenza accepts video material and animation sequences to support and enhance the presentation of the scientific research. Authors who have video or animation files that they wish to submit with their article should send them as a separate file. Reference to the video material should be clearly made in text. This will the modified into a linked to the paper's supplementary information page. All submitted files should be properly labelled so that they directly relate to the video files content. This should be within a maximum size of 50 MB.

Submission check list

The following list will be useful during the final checking of a manuscript prior to sending it to the journal for review. Please consult the Author Guidelines for further details of any item.

- One author has been designated as the corresponding author with contact details:
    - E-mail address.
    - Full postal address.
    - Telephone and fax numbers.
- All necessary files have been sent, and contain:
    - All figures are given separately in PDF, SVG, JPEG of PNG format.
    - Caption for figures is included at the end of the text.
    - All tables (including title, description, footnotes) are included in the text and large tables have been given separately as CSV.
    - The reference list has been given in XML, RIS, zotero.splite or BIB file format.
- Further considerations

- Abstract does not exceed about 250 words.
- Manuscript has been 'spell-checked' and 'grammar-checked'.
- References are in the required format.
- All references mentioned in the reference list are cited in the text, and vice versa.
- Bibliographic data for all cited material has been provided.
- Permission has been obtained for use of copyrighted material from other sources (including the Web).
- A PDF document generated from the word processor used is submitted.

After Acceptance

Use of the Digital Object Identifier The Digital Object Identifier (DOI) may be used to cite and link to electronic documents. The DOI consists of a unique alpha-numeric character string which is assigned to a document by the publisher upon the initial electronic publication. The assigned DOI never changes. Therefore, it is an ideal medium for citing a document, particularly 'Articles in press' because they have not yet received their full bibliographic information. When you use a DOI to create links to documents on the web, the DOIs are guaranteed never to change.

Proofs, Reprints and Copyright The corresponding author will receive an electronic proof of the article and have an opportunity to review editorial changes and to double-check accuracy of content, tables, and statistics before publication. A list of any necessary corrections should be sent by email to the managing editor within a week of proof receipt to avoid unnecessary delays in the publication of the article. Alterations, other than essential corrections to the text of the article, should not be made at this stage. Manuscripts are accepted for publication on the understanding that exclusive copyright is assigned to Xjenza. However, this does not limit the freedom of the author(s) to use material in the articles in any other published works.

Editorial

A New Home for Xjenza Online

Cristiana Sebu[*1]
[1]*Department of Mathematics, University of Malta, Msida, Malta*

Dear readers of Xjenza Online, as Editor, I am pleased to inform you that the release of the first issue of 2019 coincides with the launch of a brand new website for the journal, www.xjenza.org.

One of my main aims since I took over the responsibility of acting as Editor-in-Chief for Xjenza Online has been to continue to increase the journal's quality, visibility, availability and readership. The new website is not only going to achieve this, but it provides a completely new look and feel experience to all our readers, authors, expert reviewers, and members of the Editorial Board. It is not the first time when I mention the fact that the journal has evolved alongside the community it serves and has become the home to a rich and exciting scientific research. I hope that the new website will fully reflect this transformation of the journal over the years. The website also has its own manuscript submission and management system and an online article repository, thus meeting one of the indexing requirements set by internationally leading databases. Of course, this achievement would not have been possible without the assistance of the members of the Editorial Board of Xjenza Online, in general, and of Dr Sebastiano D'Amico, in particular; the hard work of the website developer, Giovanni Faraone; and, most importantly, the financial support of the Research, Innovation and Development Trust of the University of Malta.

As today's research is a global endeavour, so remains our focus in this issue of Xjenza Online, both in the original research we publish and in our news coverage.

The issue opens with an article co-authored by Zerafa, Galea and Sebu, which proposes a novel approach which uses mathematical modelling and deep neural networks to learn how to invert pseudo-spectral data for seismic waveforms. Needless to mention that deep learning techniques have recently gained great popularity for their widespread success in pattern recognition and computer vision, and they are now emerging as excellent tools for solving inverse problems and optimization of parameters.

The following article by Bellia and Lanfranco presents a preliminary assessment of the efficiency of drones in surveying land cover from large spatial scales of about $10\,\mathrm{ha}$ to smaller ones of $1\,\mathrm{m}^2$. Accurate land cover maps are a fundamental prerequisite for vegetation studies, ecological monitoring, geographical mapping, land use planning, etc. The data collection is generally carried out on the ground by a team of expert field workers, and the process is commonly time-consuming, effort-intensive and costly. This article suggests that consumer-level drones could become a versatile and viable tool for mapping, since they can be used for surveying land-cover at landscape scale as well as scales comparable to the dimensions of individual shrubs or trees.

Next, the article by Talay and Erkan from Yildiz Technical University, Turkey, illustrates the design of axial flux permanent magnet synchronous motors. This is yet another application of artificial neural networks to find the most efficient design parameters for the engine.

Then, Silvio Attard from Central Bank of Malta analyses the historic developments in the local tourism industry, focusing on the changing characteristics of demand and supply, especially since the recent surge in inbound tourism (5.3 visitors per inhabitant in 2017, the second highest ratio worldwide after Iceland) is largely driven by the increased air connectivity to and from Malta. The author is also sharing his own informed views on sustainable tourism and considerations in devising tourism policies.

The article by Tonna and Saliba is an overview of the research work related to the design of bone regenerations scaffolds made of iron alloys including proposed strategies for tailoring the corrosion, the analysis of mechanical and cytotoxic responses as well as the promising processing methods for the production of iron-

Correspondence to: Cristiana Sebu (cristiana.sebu@um.edu.mt)

based foam structures. Although *in vivo* trials of iron alloys are still rather limited, their consistent *in vitro* mechanical performance shows great promise for future applications in orthopedic and trauma surgery.

The last article by Borg and Di Giovanni reviews the existing scientific literature concerning the pathophysiological modifications of the endocannabinoid system in patients with Parkinson disease and its role as a potential target for pharmacological therapies aimed at ameliorating both motor and non-motor symptoms associated to this condition.

The issue concludes with a News article which promotes Malta's Science and Arts Festival 2019. This year the festival focuses on the theme 'The Science of YOU' probing the science of human development, and covers topics from personalized genomic medicine and bio banking to Artificial Intelligence and the Environment. It has to be noted that Malta organizes one of the largest celebrations of researchers around Europe and it is simply an event not to be missed by anyone!

Although we now have a new home, the main goals of the coverage remain the same: to serve the local professional scientific community, to publish high-quality original findings in a peer-reviewed environment, and to help early-career researchers to advance their scientific discourse in the community.

Research Article

Learning to Invert Pseudo-Spectral Data for Seismic Waveforms

Christopher Zerafa[*1], **Pauline Galea**[1] **and Cristiana Sebu**[2]
[1]*Department of Geosciences, University of Malta, Msida, Malta*
[2]*Department of Mathematics, University of Malta, Msida, Malta*

Abstract. Full-waveform inversion (FWI) is a widely adopted technique used in seismic processing to produce high resolution Earth models, that fully explain the recorded seismic data. FWI is a local optimisation problem which aims to minimise, using a least-squares approach, the misfit between recorded and modelled data. The inversion process begins with a best-guess initial model which is iteratively improved using a sequence of linearised local inversions to solve a fully non-linear problem. Deep learning has gained widespread popularity in the new millennium. At the core of these tools are Neural Networks (NN), in particular Deep Neural Networks (DNN), which are variants of these original NN algorithms with significantly more hidden layers, resulting in efficient learning of a non-linear function between input and output pairs. The learning process within DNN involves repeatedly updating network neuron weights to best approximate input-to-output mappings. There is clear similarity between FWI and DNN as both approaches attempt to solve non-linear mapping in an iterative sense. However, they are fundamentally different in that FWI is knowledge-driven, whereas DNN is data-driven. This article proposes a novel approach which learns pseudo-spectral data-driven FWI. We test this methodology by training a DNN on 1D multi-layer, horizontally-isotropic data and then apply this to previously unseen data to infer the surface velocity. Results are compared against a synthetic model and success and failures of this approach are hence identified.

Keywords: Deep Neural Networks, Full-waveform Inversion, Machine Learning, Computational Geophysics, Pseudo-Spectral Inversion

1 Introduction

1.1 Preliminaries

The seismic reflection method uses artificially generated seismic waves that excite the Earth and propagate through the subsurface. They are attenuated by interactions with their medium of propagation and are partially reflected back across a high contrasting acoustic impedance layer. A simple 2D two-layer example of an acoustic forward propagation through the subsurface is given in Fig. 1. The model contains a high acoustic impedance layer between 1 and 1.5 km depth. When hitting the interface between different velocity layers, the wave is reflected back to the surface and recorded by receivers (geophones or hydrophones) located at or close to the surface. The internal structure of the subsurface can then be inferred from the total travel time of the recorded wave.

Full-waveform inversion (FWI) is a technique which attempts to exploit the information contained in the reflected seismic wave-field as much as possible. It goes beyond refraction and reflection tomography techniques, which only use the travel time kinematics of the seismic data. It honours the Physics of the finite-frequency wave equation and uses the additional information provided by the amplitude and phase of the seismic waveform (Tarantola, 1987). FWI seeks to achieve a high-resolution geological model of the subsurface through application of multivariate optimisation to the seismic inverse problem (Lailly, 1983; Tarantola, 1984; Virieux & Operto, 2009). The inversion process begins with a best-guess initial model which is iteratively improved using a sequence of linearised local inversions to solve a fully non-linear problem. Fig. 2 illustrates the imaging uplift which is achievable through FWI. In situations of more complex structures, such as complicated salt structures with convoluted ray-paths in the overbur-

*Correspondence to*: Christopher Zerafa (christopher.zerafa.08@um.edu.mt)

(a) Synthetic p-wave velocity model.

(b) Time-step 300 ms.

(c) Time-step 600ms.

(d) Time-step 700 ms.

(e) Time-step 800 ms.

(f) Time-step 950 ms.
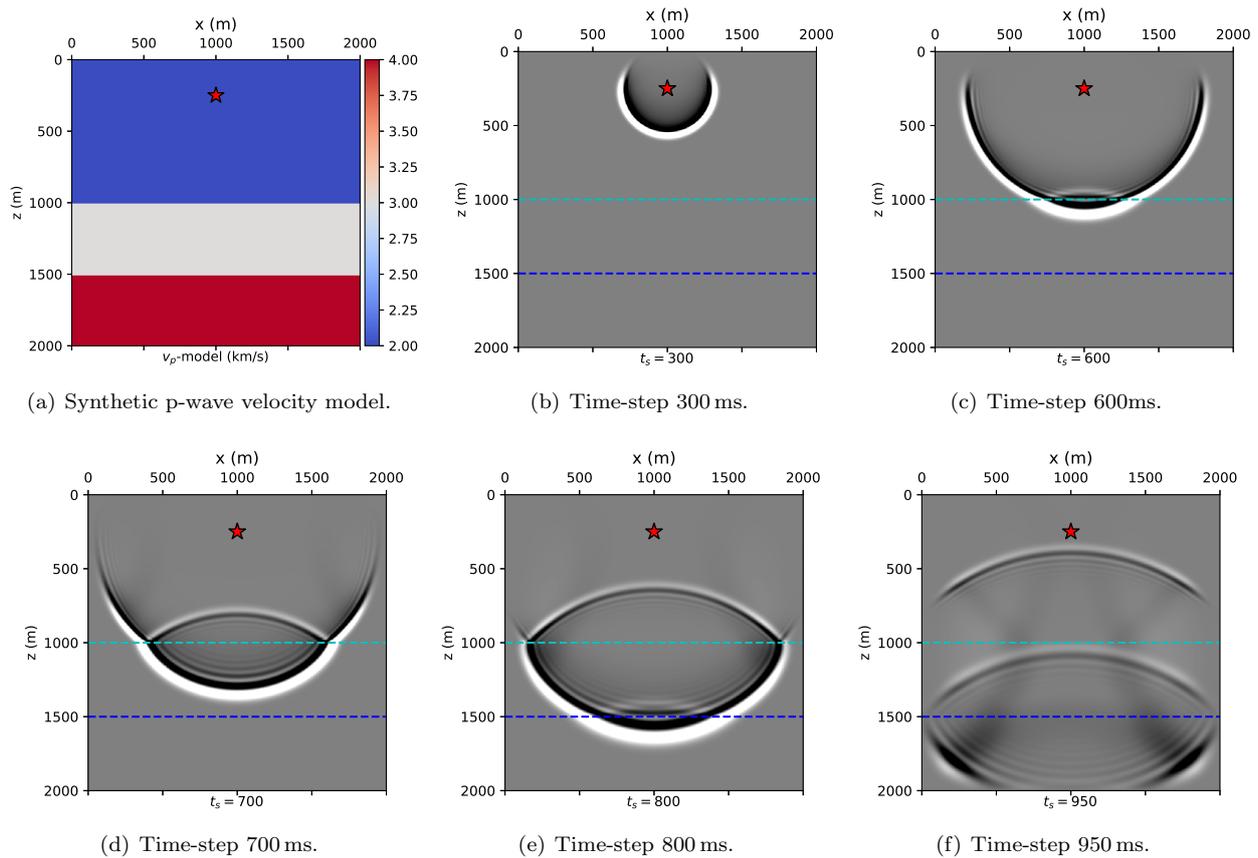
Figure 1: Simple 2D two-layer model used for forward propagation of seismic waves. The red star marks the source location at time-step 0 ms. Figure (a) is the ground truth velocity. Figures (b) to (f) illustrate the propagation of an acoustic wave through (a).

den, the inversion becomes more difficult and computationally more expensive. Fig. 3 illustrates an example of FWI on the 2004 BP synthetic data. The zoomed sections in Fig. 3(d) clearly illustrate a lack of resolution of FWI.

1.2 Aims & Objectives

Optimization theory is fundamental to FWI. The parameters of the system under investigation are reconstructed from indirect observations that are subject to a forward modelling process (Tarantola, 2005). The accuracy of this forward modelling depends on the validity of physical theory that links ground-truth, to the measured data (Innanen, 2014). Moreover, solving for this inverse problem involves learning the inverse mapping from measurements to the ground-truth which is based on a subset of degraded best-estimated data (Tarantola, 2005; Tikhonov & Arsenin, 1977). Two limitations within inverse theory can be identified: (i) solving the forward problem and (ii) training the data.

Choice of the numerical method used to solve the forward problem will crucially impact the accuracy of the FWI result. Challenging environments require

more complex assumptions to explain the physical link between data and observations, with not necessarily improved levels of accuracy (Morgan et al., 2013). Secondly, the data being used to reconstruct the mapping of measurements for the ground-truth are not optimal. Very wide angle and multi-azimuth data are required to enable full inversion (Morgan et al., 2016); this information might not necessarily have been recorded in the acquisition stages of the data. Furthermore, pre-conditioning of data is a necessity prior to FWI in order to induce well-posedness (Kumar, Ramrez & Butt, 2012; Mothi, Schwarz & Zhu, 2013; Peng, Wang, Chazalnoel & Gomes, 2018; Warner et al., 2013). However, if done incorrectly this can degrade the inversion process (Lines, 2014). Indeed, Lines (2014) shows how FWI remains robust to both random and coherent noise, and his work indicates that with the inclusion of multiple data, FWI proves useful at estimating a better solution in some situations.

Recently, deep learning (DL) techniques have emerged as excellent models and gained great popularity for their widespread success in pattern recognition (Cireşan,

(a) Conventional method.

(b) <10 Hz FWI velocity model result.
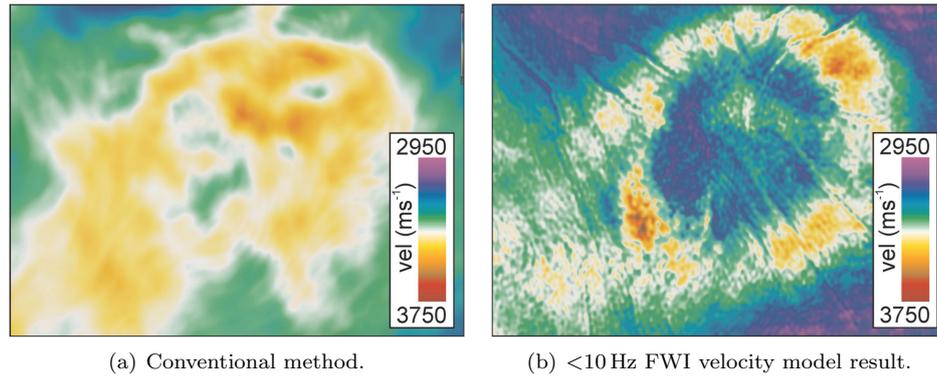
Figure 2: Horizontal slices though the Samson Dome at 1350 m. From Morgan et al. (2013).

(a) Original 2004 BP synthetic for FWI.

(b) Zoomed section.

(c) 2D FWI result.

(d) Lack of resolution.
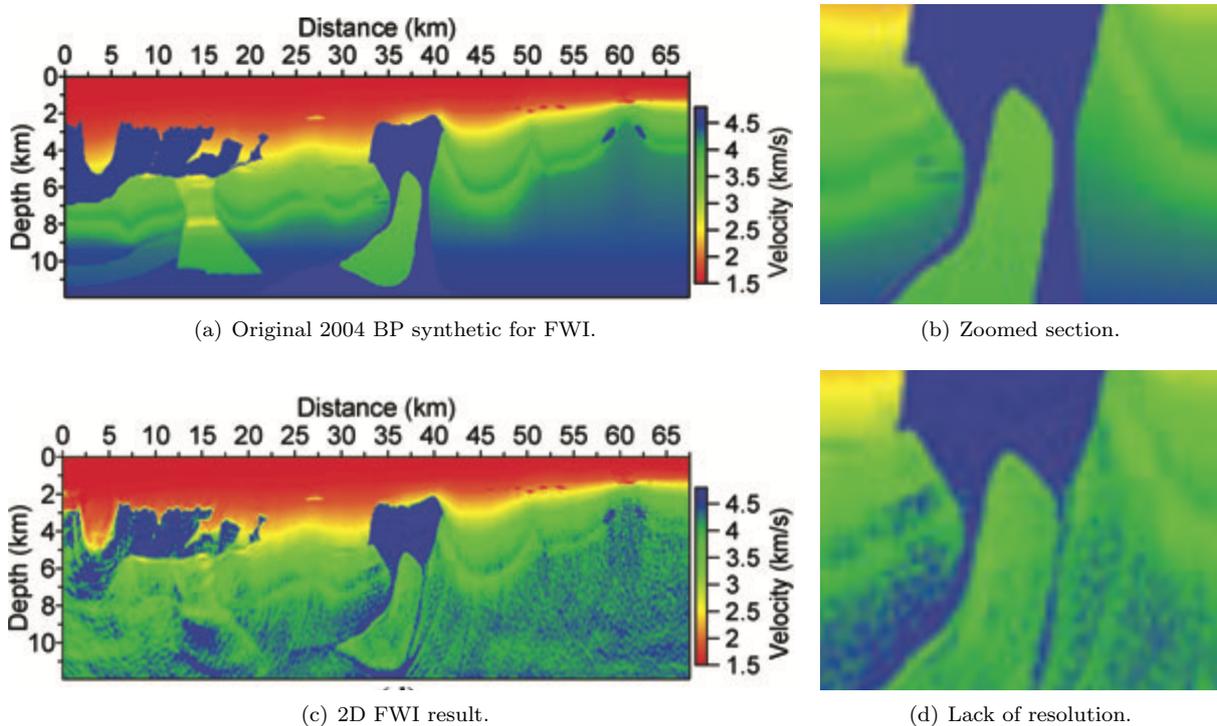
Figure 3: Limitations of FWI in complicated geology. From Shin, Koo, Cha and Park (2010).

Meier, Masci & Schmidhuber, 2012, 2011), speech recognition (Hinton et al., 2012) and computer vision (Krizhevsky, Sutskever, Hinton, Tasci & Kim, 2015; Deng & Yu, 2013). The use of Deep Neural Networks (DNN) to solve inverse problems has been explored by Elshafiey (1991), Adler and Öktem (2017), Chang, Li, Póczos, Kumar and Sankaranarayanan (2017), as well as Wei, Fai and Carin (2017), and has achieved state-of-the-art performance in image reconstruction (Kelly, Matthews & Anastasio, 2017; Petersen, Bölcskei, Grohs & Kutyniok, 2017; Adler, Ringh, Öktem & Karlsson, 2017), super-resolution (Bruna, Sprechmann &

LeCun, 2015; Galliani, Lanaras, Marmanis, Baltsavias & Schindler, 2017) and automatic-colorization (Larsson, Maire & Shakhnarovich, 2016).

In Geophysics, the applications of DL techniques have focused on the identification of features and attributes in migrated seismic sections, with few studies looking into velocity inversion. Zhang, Frogner, Araya-Polo and Hohl (2014) used a kernel regularized least-squares method for fault detection from seismic records on numerical experiments. W. Wang, Yang and Ma (2018) employed a fully convolutional neural network (FCN) to perform salt-detection from raw multi-shot gathers,

which was found to be much faster and more efficient than traditional migration and interpretation. Lewis and Vigh (2017) combined DL and FWI to improve the performance for salt inversion, by generating a probability map from learned abstractions of the data and incorporating these in the FWI objective function. These tests results showed promise for automated salt body reconstruction using FWI. Mosser et al. (2018) used a generative adversarial network (Goodfellow, Bengio & Courville, 2016) with cycle-constraints to perform seismic inversion, by reformulating the inversion problem as a domain transfer problem. The mapping between post-stack seismic traces and p-wave velocity models was approximated through DL. More recently, Yang and Ma (2019) developed a supervised FCN for velocity-model, building directly from raw seismograms using a DNN architecture based on U-Net (Ronneberger, Fischer & Brox, 2015). Their training data was obtained from modelling of the acoustic wave equation via a time-domain staggered-grid finite-difference scheme, with numerical experiments showing good potential of DL for seismic velocity inversion.

In this work, we are re-casting the mathematical formulation of FWI within a DL framework. The conventional least-squares formulation of FWI can be expressed as:

$$\min_{\mathbf{m}} J(\mathbf{m}) = ||\mathbf{d} - F(\mathbf{m})||_2^2, \tag{1}$$

where $\mathbf{m} \in M$ is the subsurface model, $F : M \to D$ is the forward wave equation model, and $\mathbf{d} \in D$ is the observed data. This inversion is non-linear and ill-posed since $\mathbf{d}$ does not contain all subsurface information to define a velocity model explicitly (Biondi, 2006). Based on the Universal Approximation Theorem (Hornik, Stinchcombe & White, 1989), a DNN can be used to approximate the non-linear inverse operator $F^{-1} : D \to M$ by a pseudo-inverse operator or mapping function g_θ which minimizes the functional:

$$J(\mathbf{m}) = ||\mathbf{m} - g_\theta(\mathbf{d})||^2, \tag{2}$$

where θ is a large simulated dataset of pairs $(\mathbf{m}, \mathbf{d})$ used for learning the process function g_θ (Hastie, Friedman & Tibshirani, 2001). In particular, based on the work of Falsaperla, Graziani, Nunnari and Spampinato (1996), DNN utilizing pseudo-spectral transformed data $\mathcal{F}$, facilitates the learning process due to better sparsity in the transformed domain, as compared to the time domain. The novelty of this approach is the combination of both DL, signal processing and inverse theory for subsurface velocity inversion. This paper aims to prove this theoretical potentially viable solution via a practical implementation to a 1D synthetic model.

The structure of this manuscript is as follows. Section 1 introduces the subject of FWI and its importance

within current workflows for seismic exploration. Limitations within the current formulation are identified and a novel approach to devise better velocity models of the subsurface is proposed. In Section 2, mathematical fundamentals for FWI and DNN are derived respectively. These are then compared and their differences are highlighted. In particular, FWI is recast as a DL problem. Based on the derived formulation in Section 2, numerical results of this novel approach are presented in Section 3 and a 1D synthetic highlights the success and failures of this approach. In Section 4, concluding remarks are presented.

2 Theoretical Framework and Methodology

2.1 Inverse Problem Formulation

The aim of inversion is to estimate the parameters of a physical system based on the measurements available. In the case of Geophysics, the physical system is the Earth and data are the recorded wave-field.
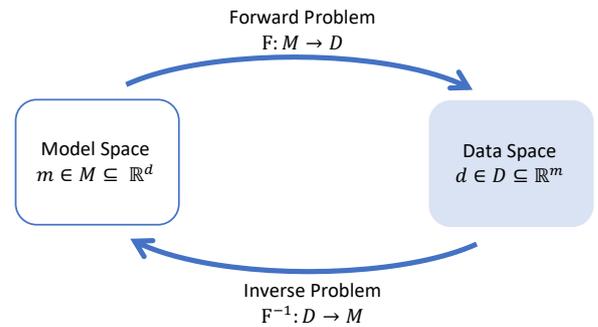
Figure 4: Visual representation of the mapping between the Forward and the Inverse problem.

The recorded wave-field is known, while the physical properties of the medium which the wave-field propagated through are the unknowns. The wave-field will be a function of these medium properties and the function for the forward problem can be as expressed as:

$$\mathbf{d} = F(\mathbf{m}), \tag{3}$$

where $F : M \to D$, $F \in \mathbb{R}^{(m \times d)}$ is the operator applied on the model space $\mathbf{m} \in M \subseteq \mathbb{R}^d$ to recover measurements $\mathbf{d} \in D \subseteq \mathbb{R}^m$. The forward problem is well-posed, that is, a unique solution exists that depends continuously on the model in some appropriate topology.

The opposite to forward modelling is the inversion. This involves making assumptions on the physical properties of the object we want to image, to be able to compute the wave-field at any given time and location to a certain degree of accuracy. If F is invertible, the inverse problem is given by:

$$\mathbf{m} = F^{-1}(\mathbf{d}), \tag{4}$$

This aims to extract all the information contained within the data.

2.2 FWI as Local Optimisation

Lailly (1983) and Tarantola (1984) re-cast the migration imaging principle introduced by Claerbout (1971) as a local optimisation problem. The forward problem is based on the wave equation, which is one of the most fundamental equations in Physics used for the description of wave motion. It is a second order, partial differential equation involving both time and space derivatives.

The particle motion for an isotropic medium is given by:

$$\frac{1}{c(\mathbf{m})^2}\frac{\partial^2 p(\mathbf{m},t)}{\partial t^2} - \nabla^2 p(\mathbf{m},t) = s(\mathbf{m},t), \qquad (5)$$

where $p(\mathbf{m},t)$ is the pressure wave-field, $c(\mathbf{m})$ is the acoustic p-wave velocity and $s(\mathbf{m},t)$ is the source (Igel, 2016). To solve the wave equation numerically, it can be expressed as a linear operator. Although the data $\mathbf{d}$ and model $\mathbf{m}$ are not linearly related, the wave-field $p(\mathbf{m},t)$ and the sources $s(\mathbf{m},t)$ are linearly related by the equation:

$$\mathbf{A}p(\mathbf{m},t) = s(\mathbf{m},t), \qquad (6)$$

where $p(\mathbf{m},t)$ is the pressure wave-field produced by a source $s(\mathbf{m})$ and $\mathbf{A}$ is the numerical implementation of the operator:

$$\frac{1}{c(\mathbf{m})^2}\frac{\partial^2}{\partial t^2} - \nabla^2, \qquad (7)$$

A common technique employed within the forward modelling stage is to perform modelling in a pseudo-spectral domain ($\mathcal{F}$) rather than the time domain ($\mathcal{T}$). The most common domain is the Fourier domain (Igel, 2016); computational implementation is generally achieved via the Fast Fourier Transform (FFT) developed by Cooley and Tukey (1965), as it utilises the fact that $e^{-2\pi i/N}$ is an N-th primitive root of unity and allows for the reduction of computational costs from $O(N^2)$ to $O(N\log N)$.

After forward modelling the data in a pseudo-spectral domain, the objective is to seek to minimize the difference between the observed data and the modelled data. The metric for the difference or misfit between the two datasets is known as the misfit-, objective- or cost-function J. The most common cost function is given by the L_2-norm of the data residuals:

$$J(\mathbf{m}) = \frac{1}{2}||d - F(\mathbf{m})||_D^2, \qquad (8)$$

where D indicates the data domain given by n_s sources and n_r receivers (Igel, 2016). The misfit function J can be minimized with respect to the model parameters d if the gradient is zero, namely:

$$\nabla J = \frac{\partial J}{\partial \mathbf{d}} = 0, \qquad (9)$$

Minimising the misfit function is generally achieved via a linearised iterative optimisation scheme, based on the Born approximation in scattering theory (Born & Wolf, 1980; Clayton & Stolt, 1980). The inversion algorithm starts with an initial estimate of the model $\mathbf{m}_0$. After the first pass via forward modelling, the model is updated by the model parameter perturbation $\Delta\mathbf{m}_0$. This newly updated model is then used to calculate the next update and the procedure continues iteratively until the computed model is close enough to the observations, based on a residual threshold criterion. At each iteration k, the misfit function $J(\mathbf{m}_k)$ is calculated from the previous iteration model $\mathbf{m}_{k-1}$ by:

$$J(\mathbf{m}_k) = J(\mathbf{m}_{k-1} + \Delta\mathbf{m}_0), \qquad (10)$$

Assuming that the model perturbation is small enough with respect to the model, Eq. (10) can be expanded via Taylor series up to second orders as:

$$\begin{aligned} J(\mathbf{m}_k) &= J(\mathbf{m}_{k-1} + \Delta\mathbf{m}_0) \\ &= J(\mathbf{m}_{k-1}) + \delta\mathbf{m}_{k-1}^T\frac{\partial J}{\partial \mathbf{m}_{k-1}} + \frac{1}{2}\delta\mathbf{m}_{k-1}^{2T}\frac{\partial^2 J}{\partial \mathbf{m}_{k-1}^2}, \end{aligned} \qquad (11)$$

Taking the derivative of Eq. (11) and minimizing to determine the model update leads to:

$$\delta\mathbf{m}_{k-1} \approx -\mathbf{H}^{-1}\nabla_{\mathbf{m}_{k-1}}J, \qquad (12)$$

where $\mathbf{H} = \frac{\partial^2 J}{\partial \mathbf{m}_{k-1}^2}$ is the Hessian matrix and $\nabla_{\mathbf{m}_{k-1}}J$ the gradient of the misfit function. The Hessian matrix is a symmetric matrix of size $N \times N$ where N is the number of model parameters and represents the curvature trend of the quadratic misfit function.

FWI is an ill-posed problem, implying that an infinite number of models that fit the observations exist. Well-posedness can be introduced with the addition of Tikhonov L_2-norm regularization (Tikhonov, 1963, 3; Tikhonov & Arsenin, 1977):

$$J(\mathbf{m}) = \frac{1}{2}\left[||d - F(\mathbf{m})||_D^2 + \lambda||\mathbf{m}||_M^2\right], \qquad (13)$$

where λ is the regularization parameter which signifies the trade-off between the data and model residuals.

2.3 FWI Algorithm Summary

A summary of FWI as a local optimisation problem is given in Algorithm 1 and a schematic is illustrated in Fig. 5.

Algorithm 1 FWI as a local optimisation problem

(I) Choose an initial model $\mathbf{m}_0$ and source wavelet $s(\mathbf{m})$.

(II) For each source location, solve the forward problem $F : M \to D$ using pseudo-spectral forward modelling everywhere in the model space to get a predicted wave-field $\mathbf{d}_k$. This is sampled at receivers $r(\mathbf{m})$.

(III) At every receiver $r(\mathbf{m})$, data residuals are calculated between the modelled wave-field $\mathbf{d}_k$ and the observed data $\mathbf{d}$.

(IV) These data residuals are back-propagated from the receivers to produce a back-propagated residual wave-field.

(V) For each source location, the misfit function $J(\mathbf{m})$ is calculated for the observed data and back-propagated residual wave-field in order to generate the gradient ∇J required at every point in the model.

(VI) The gradient is scaled based on the step-length α, applied to the starting model and an updated model is obtained $\mathbf{m}_{(k+1)}$.

(VII) The process is iteratively repeated from Step 2 until the convergence criterion is satisfied.

2.4 Deep Neural Networks for FWI

Neural Networks (NN) are a subset of tools in machine learning, which when applied to inverse problems, can approximate the non-linear functional of the inverse problem $F^{-1} : D \to M$. That is, using a NN, a non-linear mapping can be learned to minimize

$$||\mathbf{m} - g_\theta(\mathbf{d})||^2, \tag{14}$$

where θ the large data set of pairs $(\mathbf{m}, \mathbf{d})$ used for the learning process (Lucas, Iliadis, Molina & Katsaggelos, 2018).

The most elementary component in a NN is a neuron. This receives excitatory input and sums the result to produce an output or activation, representing a neuron's action potential which is transmitted along its axon (Raschka & Mirjalili, 2017). For a given artificial neuron, consider n inputs with signals m and weights w. The output d of the kth neuron from all input signals is given by:

$$d_k = \sigma \left(b + \sum_{j=0}^{m} w_{kj} m_j \right), \tag{15}$$

where σ is the activation function and b is a bias term enabling the activation functions to shift about the origin. When multiple neurons are combined together they form a NN. The architecture of a NN refers to the number of neurons, their arrangement and their connectivity (Šíma & Orponen, 2003). The initial layer of nodes $\mathbf{m}$ are referred to as the Input Layer. These are connected to a sequence of hidden layers of neurons. The final layer of the neurons is not a hidden layer and is referred to as the Output Layer. Communication proceeds layer by layer from the input layer, via the hidden layers, up to the output layer. If a NN has two or more hidden layers, it is called a DNN. Fig. 6 shows a NN
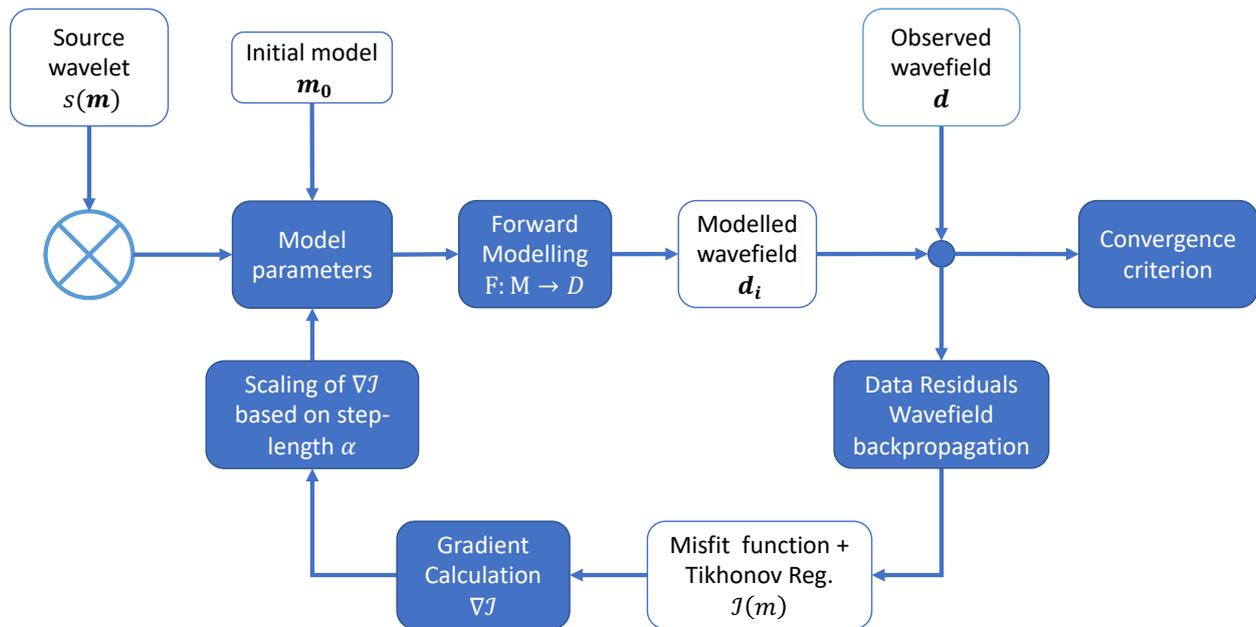

Figure 5: Schematic of a FWI workflow solved as an iterative optimisation process.
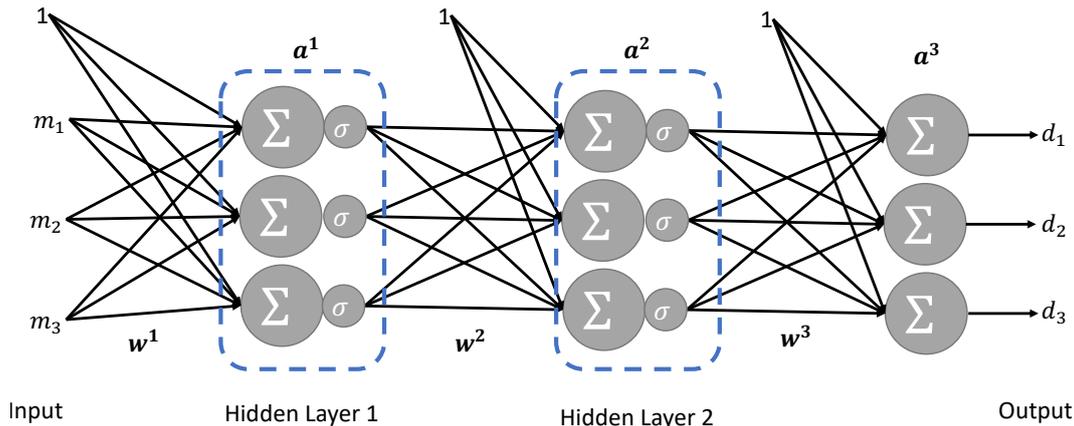
Figure 6: An example of a fully connected NN with 2 hidden layers. All weights w and bias b are learned during the training phase. The 1's connected to each hidden layer represents bias nodes which help the NN learn patterns by allowing the output of an activation function to be shifted. Adapted from Lucas, Iliadis, Molina and Katsaggelos (2018).

consisting of 2 hidden layers. The output of the unit in each layer is the result of the weighted sum of the input units, followed by a non-linear element-wise function. The weights between each unit are learned as a result of a training procedure.

When training a DNN, the forward propagation through the hidden layers from input $\mathbf{m}$ to output $\mathbf{d}$ needs to be measured for its misfit. The most commonly used cost function is the Sum of Squared Errors (SSE), defined as:

$$J(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^{J} \left(\mathbf{m} - g_\theta(\mathbf{d}^{(i)}) \right)^2, \qquad (16)$$

where $\mathbf{d}$ is the labelled true dataset, $\mathbf{d}^{(i)}$ is the output from the ith forward pass through the network and the summation is across all neurons in the network. The objective is to minimize the function J with respect to the weights w of the neurons in the NN. Employing the Chain Rule and after a series of recursive formulations, the error signals for all neurons in the network can be recursively calculated throughout the network and the derivative of the cost function with respect to all the weights w can be calculated. Training of the DNN is then achieved via a Gradient Descent algorithm, referred to as back-propagation training algorithm (Rumelhart, Hinton & Williams, 1985). The reader is referred to Goodfellow et al. (2016) and citations therein for a full mathematical formulation.

2.5 Outline for Solving FWI Using DNN

Algorithm for training of a DNN for FWI is given in Algorithm 2 and a schematic is given in Fig. 7.

Algorithm 2 FWI as a DNN problem

(I) Setup a deep architecture for the NN.

(II) Initialise the set of weights w^l and biases b^l.

(III) Forward propagate through the network connections to calculate input sums and activation function for all neurons and layers.

(IV) Calculate the error signal for the final layer δ^L by choosing an appropriate differentiable activation function.

(V) Back-propagate the errors (δ^l) for all neurons in layer l.

(VI) Differentiate the cost function with respect to biases $\left(\frac{\partial J}{\partial b^l} \right)$.

(VII) Differentiate the cost function with respect to weights $\left(\frac{\partial J}{\partial w^l} \right)$.

(VIII) Update weights w^l via gradient descent.

(IX) Recursively repeat from Step 3 until the desired convergence criterion is met.

---

3 Numerical Example

3.1 Experiment Setup

The hypothesis we would like to prove is as follows:

"Given a seismic trace in the time domain, invert for the seismic velocity (v_p) via a DNN which transforms the input data into pseudo-spectral domain and learns to invert for a velocity estimate."

3.2 Training Data

Learning of the inversion from time to pseudo-spectral domain requires a training dataset which maps time to Fourier components of magnitude and phase, and their respective velocity profile. For our numeric example, 500,000 randomly generated mappings from time $(\mathcal{T})$ to Fourier components $(\mathcal{F})$ for a 2000 ms time window
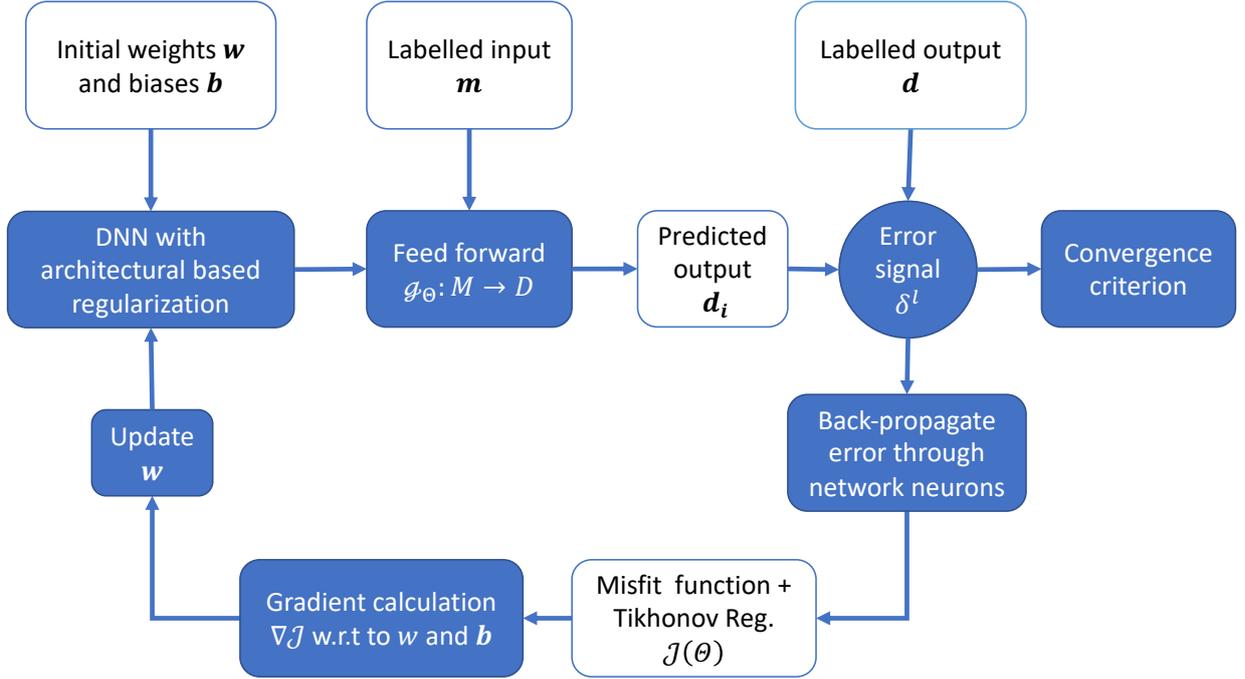
Figure 7: Schematic of a FWI workflow solved as learned optimisation process.

were created. The steps involved in the creation of the synthetic are shown in Fig. 8 for a sample velocity profile, and the steps involved in creating the dataset are given as:

   i Randomly create a v_p velocity profile for a 2000 ms time duration, with values ranging from $1400\,\text{ms}^{-1}$ to $4000\,\text{ms}^{-1}$. The lower bound of $1400\,\text{ms}^{-1}$ was selected as in normal off-shore seismic exploration conditions, the smallest observed velocity is that of the water which ranges from $1450\,\text{ms}^{-1}$ to $1460\,\text{ms}^{-1}$ (Cochrane & Cooper, 1991). Following the assumption that limestones, carbonates and salt deposits are not present in the subsurface model being inverted, as these have velocity ranges in excess of $4000\,\text{ms}^{-1}$, the upper bound of $4000\,\text{ms}^{-1}$ was selected as this is the upper limit of velocity in porous and saturated sandstones (Lee, Hutchinson, Collett & Dillon, 1996).

  ii Calculate the density ρ based on Gardner's equation (Gardner, Gardner & Gregory, 1974) given by $\rho = \alpha v_p{}^\beta$ where $\alpha = 0.31$ and $\beta = 0.25$ are empirically derived constants that depend on the Geology.

 iii At each interface, calculate the Reflection Coefficient $\mathcal{R} = \frac{\rho_2 v_{p_2} - \rho_1 v_{p_1}}{\rho_2 v_{p_2} + \rho_1 v_{p_1}}$ where ρ_i is density of medium i and v_{p_i} is the p-velocity in medium i.

 iv For each medium, calculate the Acoustic Impedance $\mathcal{Z}_i = \rho_i v_{p_i}$.

  v Define a wavelet $\mathcal{W}$. This was selected to be a Ricker wavelet at $10\,\text{Hz}$ (Ricker, 1943). The Ricker

wavelet is a theoretical waveform that takes into account the effect of Newtonian viscosity and is representative of seismic waves propagating through visco-elastic homogeneous media (Y. Wang, 2015), thus making it ideal for this numerical simulation. Based on literature results, the central frequency of $10\,\text{Hz}$ was chosen as a nominal value to be representative of normal FWI conditions (Morgan et al., 2013). Beyond $10\,\text{Hz}$ would be considered to be super-high-resolution FWI (Mispel, Furre, Sollid & Maaø, 2019), which goes beyond the scope of this manuscript.

 vi The Reflection Coefficient and wavelet are convolved to produce the seismic trace $\mathcal{T}$

vii Fourier coefficients for magnitude $\mathcal{F}(\zeta)$ and phase $\mathcal{F}(\phi)$ are derived based on the FFT.

3.3 DNN Architecture

Fig. 9 illustrates the NN architecture used to first invert for the Fourier coefficients from the time domain and then invert for velocity profile. The complete workflow had 5 modules, with each module consisting of NN with 5 fully-connected hidden layers. The layer distributions consisted of an input layer of 2000 neurons, then a set of 5 hidden layers of sizes 1000, 500, 250, 500, 1000 neurons, and an output layer of 2000 neurons. This hour-glass design can be considered representative of multi-scale FWI (Bunks, Saleck, Zaleski & Chavent, 1995) since at each hidden layer, the NN learns an ab-
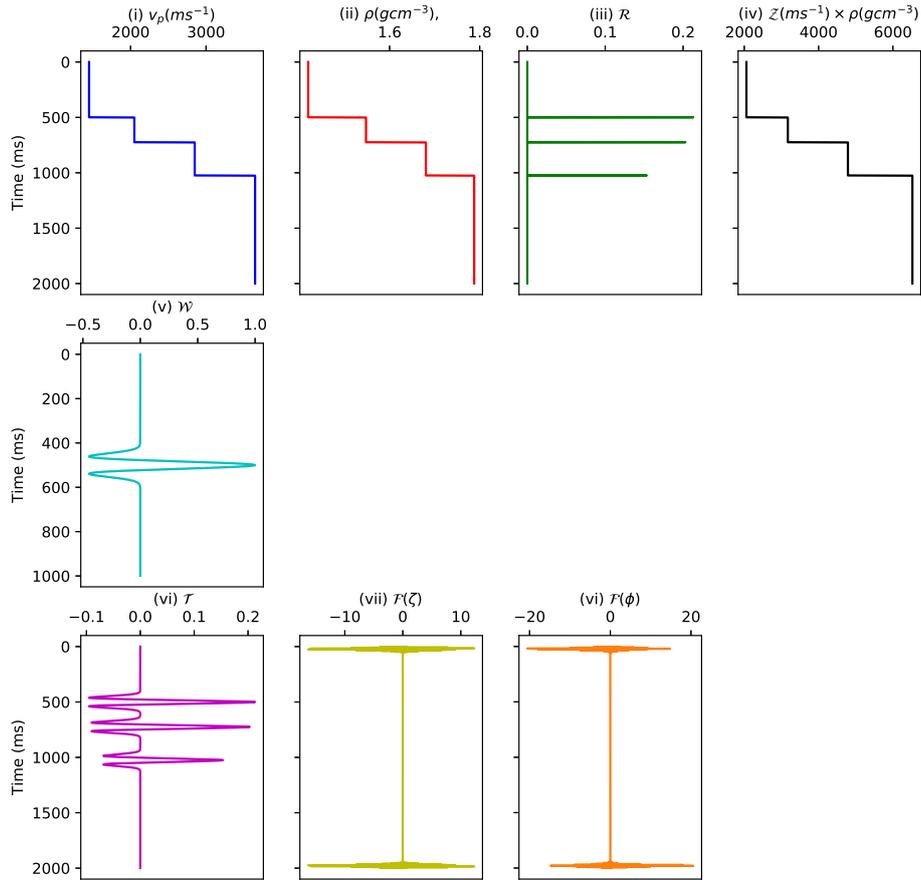
Figure 8: Workflow for creating a pseudo-spectral synthetic trace. This was repeated 500,000 times with random parameters generated within the pre-defined limits stated in Section 3.2 in order to create the learning dataset.
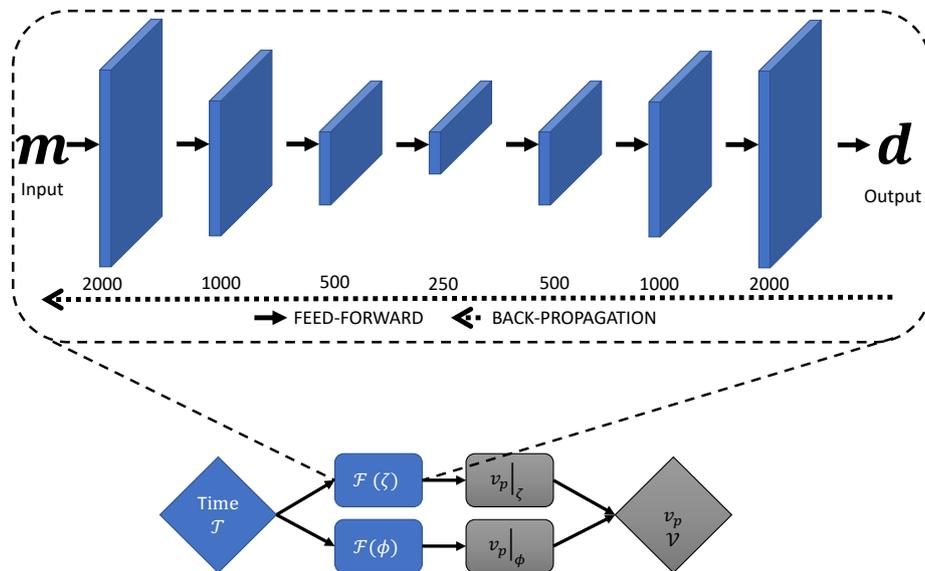
Figure 9: Pseudo-spectral FWI DNN architecture. The highlighted section indicates the set-up employed in each of the 5 modules. Each network has an hour-glass shape with layers of sizes 2000-1000-500-250-500-1000-2000 neurons which can be related to multi-scale FWI. The bottom section illustrates the DNN workflow, where $\mathcal{T}$ is the input time domain, $\mathcal{V}$ is the output v_p velocity and $\mathcal{F}$ is the Fourier domain, with magnitude ζ and phase ϕ.
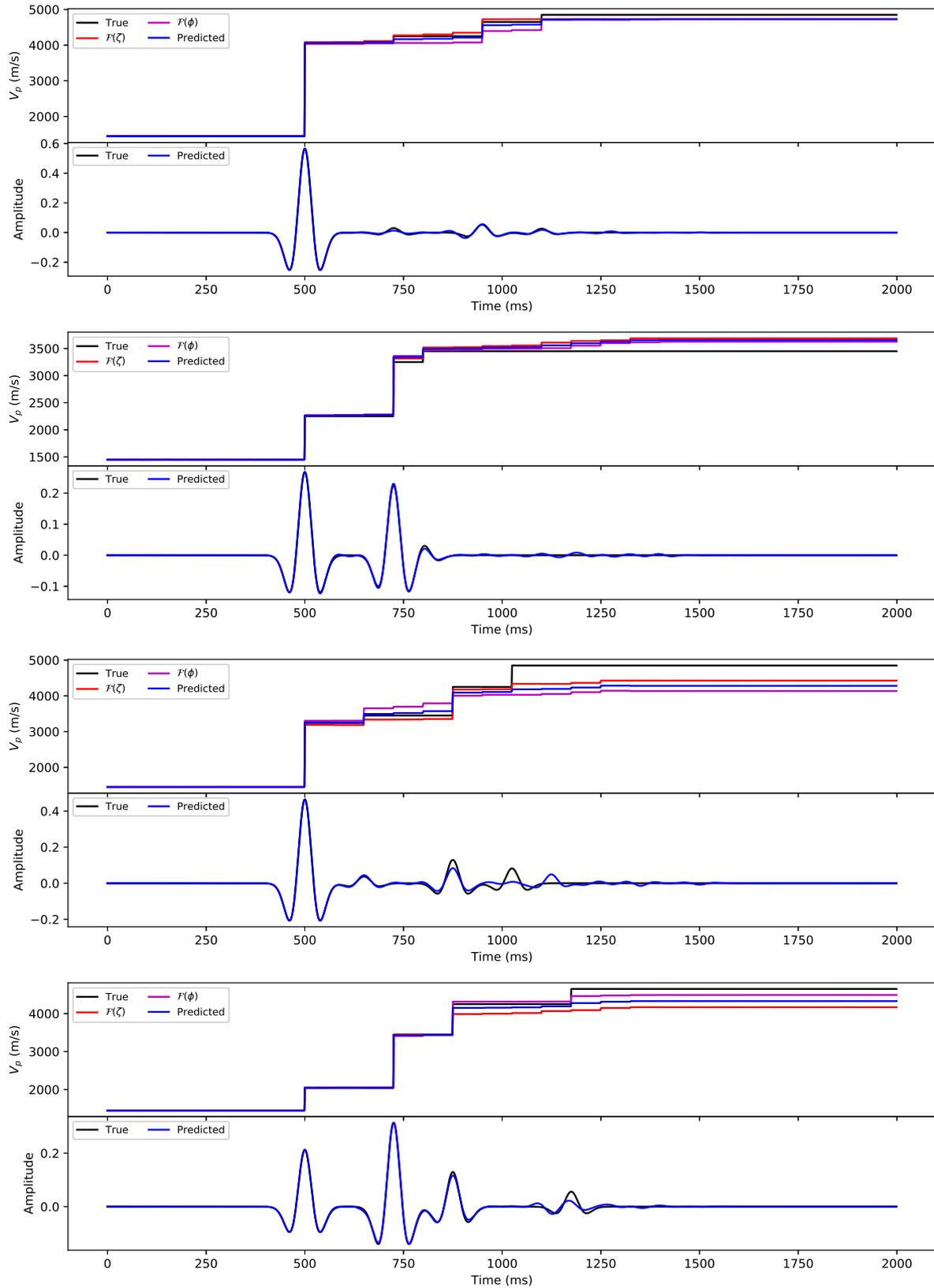
Figure 10: Four different predictions obtained from learned weights of the DNN on unseen data. The top panels are the velocity profile reconstructions from the two NN architecture branches ($\mathcal{F}(\zeta)$ and $\mathcal{F}(\phi)$) and the combined result. Bottom panels are the observed and inverted waveforms.
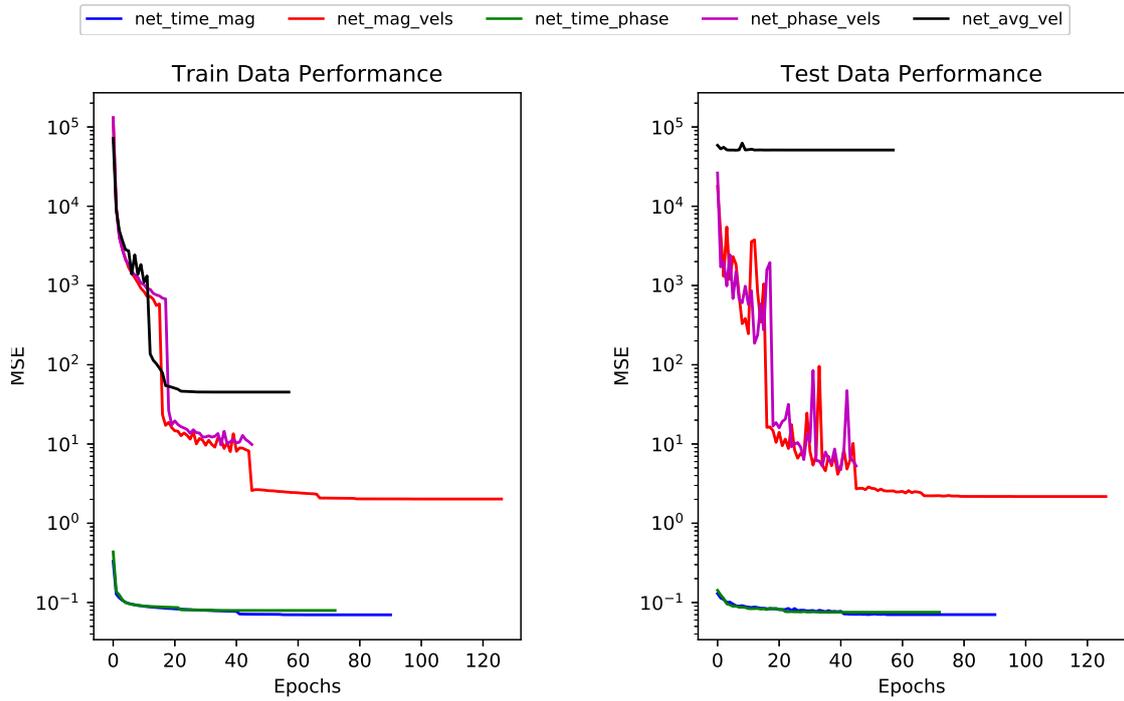
(a) Training dataset MSE over the different epochs per DNN component. Overall performance is decreasing per epoch, indicating that the DNN is learning to invert.

(b) Test dataset MSE over the different epochs per DNN component.
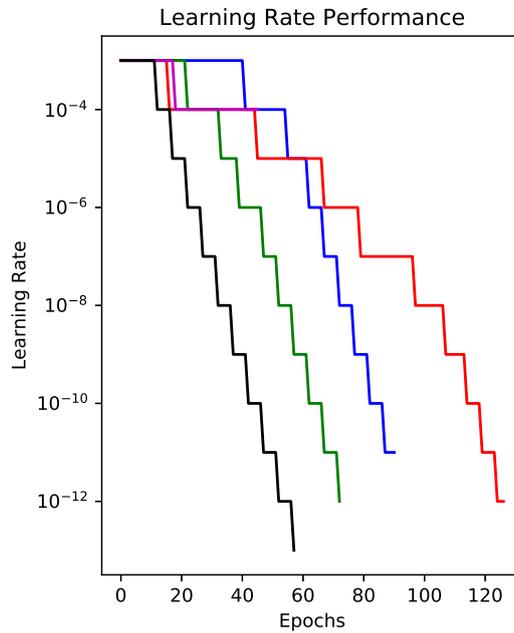
(c) Learning Rate performance over the different epochs per DNN component.

Figure 11: DNN performance metrics.

stracted component of the data at a different scale. The network employed a sum of squared errors loss function, data batching, early stopping, L_2-norm regularization updates and executed for 200 epochs. A Rectified linear unit or ReLU function given by $f(x) = \max(0, x)$ was used as an activation function. This is a nonlinear function which allows for back-propagation of errors. When employed on a network of neurons, the negative component of the function is converted to zero and the neuron is deactivated, thus introducing sparsity within the network and making it efficient and easy for computation. The output from each parallel thread in the flow is fed into another neural network which learns the optimal way of combining the outputs. In total, the DNN had 25 hidden layers. The learning or back-propagation for each network was optimized via an ADAM optimizer (Kingma & Ba, 2014), which is a stochastic gradient descent-based algorithm for first order gradient-based optimisation, which employs on adaptive estimates of lower-order moments. The DNN was implemented in Python 3.7, using Keras 2.2.4 (Chollet, 2015) and TensorFlow 1.13.1 (Abadi et al., 2016) backend.

3.4   Numerical Results

Fig. 10 illustrates the application of DNN architecture in Section 3.3 for a sample of unseen data and the respective reconstruction. Inspection of the first 750 ms indicates that the DNN approach is able to reconstruct both the velocity and the waveform profile with minimal error, irrespective of the number of layers and the magnitude of the acoustic difference in this time range. Beyond 750 ms, reconstructions start suffering from slight degradation. As illustrated in the velocity reconstruction of the middle figure, the inaccuracy is minimal and ranges $\pm 100 \, \text{ms}^{-1}$. However, this leads to perturbations in the reconstruction and does not allow for perfect matching. Further inspection suggests that the main source of error is due to the magnitude component of the network (red). To improve this error component, the network inverting for the magnitude component of the FFT would need to be trained and generalised further.

Fig. 11 shows the DNN metric performance over the different epochs per DNN component. Fig. 11(a) and 11(b) illustrate the MSE performance for the training and testing dataset respectively. Considering the former, the plots indicate that the network is indeed learning, since MSE is decreasing at each epoch. Comparing respective DNN components between the training and the testing dataset metrics, there is evidence of no under-fitting or over-fitting with the pseudo-spectral learning components of the DNN architecture (`net_-time_mag`, `net_mag_vels`, `net_time_phase`, `net_phase_-vel`). Furthermore, there is indeed good-fit, since train-

ing and testing MSE both decrease to a point of stability with a minimal difference between the two final MSE values. On the other hand, `net_avg_vel` component, which is learning to average out the velocity from Fourier components, indicates symptoms of an underpresented training dataset. Moreover, these MSE performance plots indicate that the technique might suffer from a *compounding error* issue. The two best performing components are the first layer of learning for the inversion, namely Time-to-FFT-Magnitude (`net_-time_mag`) and Time-to-FFT-Phase (`net_time_phase`), as their MSE performance plateaus at 10^{-1}. In the second phase of the inversion, which converts respective FFT components to velocities (FFT-Magnitute-to-Velocity (`net_mag_vels`) and FFT-Phase-to-Velocity (`net_phase_vels`)), the error plateaus are at 10^1, which is two orders of magnitude greater. The final network component sits even higher on the scale at 10^2. Both the train and the test dataset show drastic decreases in the MSE at different epoch levels. These can be attributed to the step-wise reductions in learning rate shown in Fig. 11(c). This varying learning rate allows the network to move to a deeper optimisation level and approach a more global minima for the optimisation problem.

4   Conclusions

In this manuscript we presented the investigation of direct modelling for seismic waveforms using a DNN, which first converts data to a pseudo-spectral domain and subsequently inverts for velocity profiles. Experimental results demonstrated that the use of synthetically generated data to train a DNN proves to be a viable technique in order to learn how to invert via pseudo-spectral data. Although inversion was successfully achieved in the numerical examples presented, one branch of the DNN architecture was lacking in inversion performance and was resulting in a compounding error effect. To improve the overall performance of the technique, data augmentation will be considered, as it is probable that 500,000 random traces are not sufficient to train the magnitude component of the Fourier transform for the network, in order to achieve a desirable performance. In addition, fine-tuning of the NN architecture in the form of in-between layer regularization, neuron drop-out during epoch training and convolutional layers have yet to be investigated. Moreover, in the next stage, this technique will be used for the inversion of more interesting subsurface structures which have a geological relevance, to evaluate image resolution when compared to standard FWI, and to also consider the case of a sequential input in the form of a Recurrent Neural Network, similar to the work of Sun, Niu, Innanen, Li and Trad (2019), but via a pseudo-spectral approach.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Devin, M. et al. (2016). Tensorflow: large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

Adler, J. & Öktem, O. (2017). Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, *33*(12), 1–24.

Adler, J., Ringh, A., Öktem, O. & Karlsson, J. (2017). Learning to solve inverse problems using Wasserstein loss. *Iclr 2018*, 1–13.

Biondi, B. L. (2006). *3D seismic imaging*. Investigations in Geophysics No. 14. Society of Exploration Geophysicists.

Born, M. & Wolf, E. (1980). Principles of optics. *Pergamon Press*, *6*, 188–189.

Bruna, J., Sprechmann, P. & LeCun, Y. (2015). Super-Resolution with Deep Convolutional Sufficient Statistics. *arXiv preprint arXiv:1511.05666*.

Bunks, C., Saleck, F. M., Zaleski, S. & Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, *60*(5), 1457–1473.

Chang, J. H., Li, C.-L., Póczos, B., Kumar, B. V. K. & Sankaranarayanan, A. C. (2017). One Network to Solve Them All—Solving Linear Inverse Problems using Deep Projection Models. *IEEE International Conference on Computer Vision (ICCV)*.

Chollet, F. et al. (2015). Keras. https://keras.io.

Cireşan, D. C., Meier, U., Masci, J. & Schmidhuber, J. (2011). A committee of neural networks for traffic sign classification. In *The 2011 International Joint Conference on Neural Networks* (pp. 1918–1921). San Jose, CA, USA: IEEE.

Cireşan, D. C., Meier, U., Masci, J. & Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. *Neural networks*, *32*, 333–338.

Claerbout, J. (1971). Toward a unified theoru of relector mapping. *Geophysics*, *36*(June).

Clayton, R. W. & Stolt, R. (1980). A Born-WKBJ Inversion Method for Acoustic Reflection Data. In *Sepreport* (Vol. 46, *1*, pp. 135–158).

Cochrane, G. R. & Cooper, A. K. (1991). Sonobuoy seismic studies at ODP drill sites in Prydz Bay, Antarctica. *Proceedings of the Ocean Drilling Program, Scientific Results*, *119*, 27–43.

Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, *19*(90), 297–301.

Deng, L. & Yu, D. (2013). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, *7*(3–4), 197–387.

Elshafiey, I. M. (1991). *Neural network approach for solving inverse problems* (Master's thesis, Iowa State University, Ames, Iowa).

Falsaperla, S., Graziani, S., Nunnari, G. & Spampinato, S. (1996). Automatic classification of volcanic earthquakes by using multi-layered neural networks. *Natural Hazards*, *13*(3), 205–228.

Galliani, S., Lanaras, C., Marmanis, D., Baltsavias, E. & Schindler, K. (2017). Learned Spectral Super-Resolution. *arXiv preprint arXiv:1703.09470*.

Gardner, G. H. F., Gardner, L. W. & Gregory, A. R. (1974). Formation Velocity and Density - The diagnostic basics for Stratigraphic Traps. *GEOPHYSICS*, *39*(6), 770–780.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. Adaptive Computation and Machine Learning series. MIT Press.

Hastie, T., Friedman, J. & Tibshirani, R. (2001). The Elements of Statistical Learning. *Springer Series in Statistics*.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., ... Sainath, T. N. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, *29*(6), 82–97.

Hornik, K., Stinchcombe, M. & White, H. (1989). Multilayer feedforward networks are universal approximator. *Neural Networks*, *2*, 359–366.

Igel, H. (2016). *Computational Seismology: A Practical Introduction* (1st ed.). Oxford University Press.

Innanen, K. (2014). Quantifying the incompleteness of the physics model in seismic inversion. *CREWES Research Report*.

Kelly, B., Matthews, T. P. & Anastasio, M. A. (2017). Deep Learning-Guided Image Reconstruction from Incomplete Data. *arXiv preprint arXiv:1709.00584*, (Nips).

Kingma, D. P. & Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., Hinton, G. E., Tasci, T. & Kim, K. (2015). ImageNet Classification with Deep Convolutional Neural Networks. *Stanford cs231b*.

Kumar, J., Ramrez, A. & Butt, S. (2012). Preparing Data for Full Waveform Inversion: A Workflow for Free-surface Multiple Attenuation. In *74th EAGE Conference and Exhibition - Workshops*. Copenhagen, Denmark: EAGE.

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In J. Bednar, R. Redner, E. Robinson & A. Weglein (Eds.), *Conference on Inverse Scattering, Theory and Application* (pp. 206–220). Tulsa, Oklahoma: Philadelphia: Society of Industrial and Applied Mathematics.

Larsson, G., Maire, M. & Shakhnarovich, G. (2016). Learning representations for automatic colorization. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 9908 LNCS, pp. 577–593).

Lee, M., Hutchinson, D., Collett, T. & Dillon, W. P. (1996). Seismic velocities for hydrate-bearing sediments using weighted equation. *Journal of Geophysical Research: Solid Earth*, *101*(B9), 20347–20358.

Lewis, W. & Vigh, D. (2017). Deep learning prior models from seismic images for full-waveform inversion. In *Seg technical program expanded abstracts 2017* (pp. 1512–1517). Society of Exploration Geophysicists.

Lines, L. (2014). FWI and the "Noise" Quandary. *CREWES Research Report.*

Lucas, A., Iliadis, M., Molina, R. & Katsaggelos, A. K. (2018). Using Deep Neural Networks for Inverse Problems in Imaging: Beyond Analytical Methods. *IEEE Signal Processing Magazine*, *35*(1), 20–36.

Mispel, J., Furre, A., Sollid, A. & Maaø, F. A. (2019). High Frequency 3D FWI at Sleipner: A Closer Look at the CO2 Plume. In *81st EAGE Conference and Exhibition 2019.* London, UK: EAGE.

Morgan, J., Warner, M., Arnoux, G., Hooft, E., Toomey, D., VanderBeek, B. & Wilcock, W. (2016). Next-generation seismic experiments - II: Wide-angle, multi-azimuth, 3-D, full-waveform inversion of sparse field data. *Geophysical Journal International*, *204*(2), 1342–1363.

Morgan, J., Warner, M., Bell, R., Ashley, J., Barnes, D., Little, R., . . . Jones, C. (2013). Next-generation seismic experiments: wide-angle, multi-azimuth, three-dimensional, full-waveform inversion. *Geophysical Journal International*, *195*(3), 1657–1678.

Mosser, L., Kimman, W., Dramsch, J., Purves, S., De la Fuente Briceño, A. & Ganssle, G. (2018). Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks. In *80th EAGE Conference and Exhibition 2018.* Copenhagen, Denmark: EAGE.

Mothi, S., Schwarz, K. & Zhu, H. (2013). Impact of full-azimuth and long-offset acquisition on Full Waveform Inversion in deep water Gulf of Mexico. *SEG Houston 2013 Annual Meeting*, (June 2013), 924–928.

Peng, C., Wang, M., Chazalnoel, N. & Gomes, A. (2018). Subsalt imaging improvement possibilities through a combination of FWI and reflection FWI. *The Leading Edge*, *37*(1), 52–57.

Petersen, P. C., Bölcskei, H., Grohs, P. & Kutyniok, G. (2017). Optimal approximation with sparse deep neural networks. *arXiv preprint arXiv:1705.01714.*

Raschka, S. & Mirjalili, V. (2017). *Python machine learning* (2nd ed.). Packt Publishing Ltd.

Ricker, N. (1943). Further developments in the wavelet theory of seismogram structure. *Bulletin of the Seismological Society of America*, *33*(3), 197–228.

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *18th International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer. Munich, Germany.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1985). *Learning internal representations by error propagation.* California Univ San Diego La Jolla Inst for Cognitive Science.

Shin, C., Koo, N.-H., Cha, Y. H. & Park, K.-P. (2010). Sequentially ordered single-frequency 2-D acoustic waveform inversion in the Laplace-Fourier domain. *Geophysical Journal International*, *181*(2), 935–950.

Šíma, J. & Orponen, P. (2003). General-Purpose Computation with Neural Networks: A Survey of Complexity Theoretic Results. *Neural Comput. 15*(12), 2727–2778.

Sun, J., Niu, Z., Innanen, K. A., Li, J. & Trad, D. O. (2019). A theory-guided deep learning formulation of seismic waveform inversion. In *Seg technical program expanded abstracts 2019* (pp. 2343–2347). Society of Exploration Geophysicists.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, *49*(8), 1259–1266.

Tarantola, A. (1987). *Inverse Problems Theory, Methods for Data Fitting and Model Parameter Estimation.* Amsterdam, The Netherlands: Elsevier.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation.* SIAM.

Tikhonov, A. N. (1963). On the Solution of Incorrectly Stated Problems and a Method of Regularization. *Dokl. Acad. Nauk SSSR*, *151*, 501–504.

Tikhonov, A. N. & Arsenin, V. Y. (1977). *Methods for solving ill-posed problems.* John Wiley and Sons, Inc.

Virieux, J. & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26.

Wang, W., Yang, F. & Ma, J. (2018). Automatic salt detection with machine learning. In *80th EAGE Conference and Exhibition 2018.* Copenhagen, Denmark: EAGE.

Wang, Y. (2015). Frequencies of the ricker wavelet. *Geophysics*, *80*(2), A31–A37.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., ... Bertrand, A. (2013). Anisotropic 3D full-waveform inversion. *GEOPHYSICS*, *78*(2), R59–R80.

Wei, Q., Fai, K. & Carin, L. (2017). An Inner-loop Free Solution to Inverse Problems using Deep Neural Networks. *Advances in Neural Information Processing Systems*, 2370–2380.

Yang, F. & Ma, J. (2019). Deep-learning inversion: a next generation seismic velocity-model building method. *Geophysics*, *84*(4), 1–133.

Zhang, C., Frogner, C., Araya-Polo, M. & Hohl, D. (2014). Machine-learning based automated fault detection in seismic traces. In *76th EAGE Conference and Exhibition 2014*. Amsterdam, The Netherlands: EAGE.

Research Article

A Preliminary Assessment of the Efficiency of Using Drones in Land Cover Mapping

Andrea Francesca Bellia[1] **and Sandro Lanfranco**[*1]
[1]*Department of Biology, University of Malta, Msida, Malta*

Abstract. This study represents a preliminary assessment of the efficiency of drones in surveying land cover at both large (*c.* 10 ha) and smaller (1 m^2) spatial scales. A DJI Mavic 2 drone was used to image the entire area of study and an orthomosaic was produced. This was converted into a land cover map through k-means clustering, with $k = 3$, where 'Vegetation', 'Bedrock' and 'Bare soil' corresponded to the land cover categories. Regions of interest (ROIs) were selected and subsequently surveyed from close range. The correspondence between predicted land cover (pLC) and observed land cover (oLC) was then assessed. On a large spatial scale, absolute correspondence was present between pLC and oLC. In terms of relative representation of land cover categories, 'Vegetation' was the only significantly correlated category across pLC and oLC, whilst the analogous correlations for 'Bedrock' and 'Bare soil' were weaker. The lower correspondence between pLC and oLC for 'Bedrock' and 'Bare soil' was due to the low value of $k = 3$ in the k-means clustering algorithm. This constrains a mixture of land covers into just one land cover category, with consequent reduction of the correlation between pLC and oLC. The method's accuracy and cost-effectiveness were compared to that of standard methods for land cover surveying. The entire process, including verification and orthomosaic land cover map processing times, approximated 32 hours. Consequently, this method is much shorter than standard surveys, which take days or weeks, and also requires less manpower.

Keywords: drone imagery, land cover mapping, vegetation mapping, image analysis, k-means clustering

1 Introduction

Amongst other uses, accurate land cover maps are a fundamental prerequisite for vegetation studies, ecological monitoring, geographical mapping, and land use planning. The spatial resolution of such maps depends on the purpose for which the map is intended, and on the size of the area under study. For coarse-grained mapping of large areas with relatively few land cover types, photographs taken from satellites or aircraft are usually useful (Anderson & Gaston, 2013; David & Ballado, 2016). Some satellites are also capable of much higher resolutions. The Sentinel-2 satellite is equipped with an opto-electronic multispectral sensor for surveying with a resolution of 10 to 60 m in the visible, near infrared (VNIR), and short-wave infrared (SWIR) spectral zones (https://eos.com/sentinel-2). However, if fine-grained mapping with a resolution of less than 10 m is required, aircraft and satellites would not be as useful, since their flight path would be too high and too fast, and their repeated use too costly (Anderson & Gaston, 2013). In summary, the data returned from these platforms would usually be too general to be relevant to the localised scales at which many ecological processes operate (Wulder, Hall, Coops & Franklin, 2004).

As such, the compilation of detailed land cover maps for smaller areas is often based on a bottom-up approach. For example, widely applied methods involve sampling using belt transects traversing the area (if an evident ecologically-relevant gradient is observed or suspected), in addition to using quadrat plots positioned at specific points. Such strategies enable direct characterisation of land cover within the sampling footprint, and interpolation of probable land cover in intervening areas that were not sampled directly. The design of a survey programme depends on the trade-off between the coverage required and the time and funds available. The use

Correspondence to: Sandro Lanfranco (sandro.lanfranco@um.edu.mt)

of more transects or quadrats would yield greater coverage, but at the expense of requiring more funds. Data collection is generally carried out on the ground by a team of expert field workers, with the effort required increasing with the size and ecological complexity of the area under study. As a result, the process is commonly time-consuming, effort-intensive and costly.

A possible solution to address the issues of cost, time, manpower and map detail, is the use of low altitude (< 500 m) aerial imagery to visualise an entire area of study, as well as specific regions of interest (ROIs) within it. The only affordable camera platform that can offer this versatility is a drone, a relatively small aircraft system that is remotely piloted through radio waves. Consumer-level drones with high quality photographic capability have recently increased in affordability and availability. This suggests that drones could represent a versatile and viable tool for mapping, since they can be used for surveying land-cover at the landscape scale, as well as at scales comparable to the dimensions of individual shrubs or trees.

2 Aims

This study represents a preliminary assessment of the efficiency of using a drone to survey land cover in an area of study, at both large ($c.$ 10 ha) and smaller spatial scales (1 m^2). The large scale surveying will be used to generate an orthomosaic of the area of study, the accuracy of which would subsequently be validated against imagery captured from very low altitudes. The accuracy and cost-effectiveness of the method will be compared to that of standard methods for land cover surveying.

3 Materials and Method

3.1 Apparatus Used

All aerial imagery was captured using a DJI Mavic Pro 2 drone equipped with a Hasselblad L1D-20C camera, with 35 mm-equivalent focal length of 28 mm, a maximum lens aperture of f/2.8, and a Field of View (FOV) of approximately 77°. The CCD sensor was 13.2×8.8 mm in size and comprised of 20 million effective pixels. The size of the images produced was 5472×3648 pixels. The drone was flown using the proprietary RC unit connected to a Samsung Galaxy S9 smartphone, on which the Litchi app (flylitchi.com; VC Technology Ltd.) was running. This enabled live camera feed from the drone to be visible on the smartphone. A virtual grid with crosshairs intersecting at the centre of the camera's field of view was overlaid on the live camera feed in the Litchi environment, in order to facilitate positioning of the drone relative to specific targets.

3.2 The Area of Study

The study was carried out in an area of study (AoS) at Qasam Barrani, in the north-western part of Malta (Figs. 1 and 2). The primary and secondary axes of the AoS measured 330 m and 190 m respectively, covering an area of approximately 42 000 m^2. Land cover at the time of survey mainly consisted of a mosaic of perennial shrubs and grasses, tracts of exposed bedrock and patches of bare soil. A small number of anthropogenic constructions, including disused trapping hides and dry stone walls, were also present in the area.
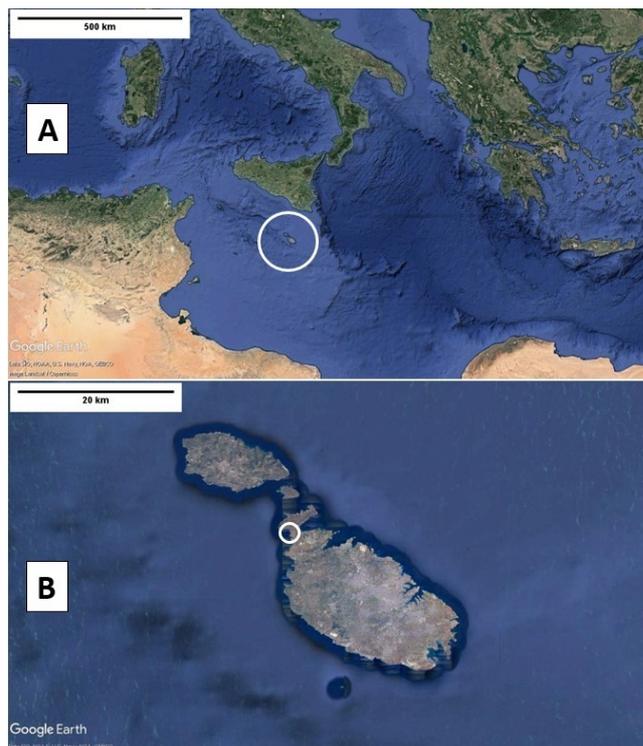

Figure 1: (A) General location, indicated by a white circle, of the Maltese Islands in the Central Mediterranean region. (B) Approximate location, indicated by a white circle, of the area of study in Malta. North is towards the top of the images. Base image: Google Earth.

3.3 Capture of Aerial Imagery

3.3.1 General Method

The mapping process was subdivided into three separate phases that were carried out sequentially. Phase 1 involved flying the drone at an elevation of approximately 30 m above ground in a number of parallel passes (henceforth referred to as 'transects') above the AoS, in order to capture the imagery required. These will subsequently be referred to as the 'high-altitude' images. Phase 2 was the processing stage, where the 'high-altitude' imagery was used to generate an orthomosaic. Phase 3 was the verification stage, where regions of in-

Figure 2: The AoS and its environs. The area that was surveyed is indicated by the red polygon. The blue rectangle indicates the boundaries of the actual area in which land cover was mapped. North is towards the top of the image. The red area covers *c.* 71 000 m^2 and the blue area covers *c.* 42 000 m^2.

terest (ROIs) noted in the orthomosaic were surveyed from a much lower altitude (5–10 m), in order to compare the predicted land cover with the observed land cover. The images captured during Phase 3 will subsequently be referred to as 'low-altitude' images. It should be emphasised that the land cover model produced is restricted to a local coordinate system, where only the distances between features within the model are considered. This model cannot be placed in an absolute geographic coordinate system as no ground control points (GCPs) were utilised. This decision was taken as there was no requirement to assess the area of study in the context of its surroundings, since the principal objective was to generate a map of the land cover in a localised area.

3.3.2 Calibration of FOV Area with Drone Altitude

Determination of the optimum survey altitude necessitated the calculation of the area covered by the camera's field of view (FOV) at different altitudes above ground level. The calibration process was carried out over a number of sessions in the grounds of the University of Malta.

The drone was flown above a car park in which the ground was marked with parking bays of known dimensions. The drone was piloted to the target altitude and photographs of the parking bays were taken in Digital

Negative (DNG) RAW format. This format was selected in order to minimise any artefacts that may be introduced during compression to other formats such as JPEG. The target altitudes selected were 3 m, 5 m, 10 m, 15 m, 20 m, 25 m, 30 m, 35 m, 40 m, 45 m, 50 m, 75 m and 100 m above ground level.

Exchangeable image file format (EXIF) data was subsequently extracted from the metadata and used to record the relative altitude (calculated from the drone's barometric pressure sensor) at which each photograph was taken. The relative altitude sometimes deviated from the target altitude due to air turbulence experienced by the drone. The deviation between the relative altitude and target altitude was never larger than 0.1 m, equivalent to 0% to 0.67% of the target altitude. The images were subsequently processed in Image J v.1.52n (Schneider, Rasband & Eliceiri, 2012). The parking bay markings were used to calibrate the dimensions of each image, and the area covered by the FOV at each altitude was subsequently calculated using the area measurement tool of Image J. The calibration process was carried out independently on three occasions, with almost identical results (Fig. 3).
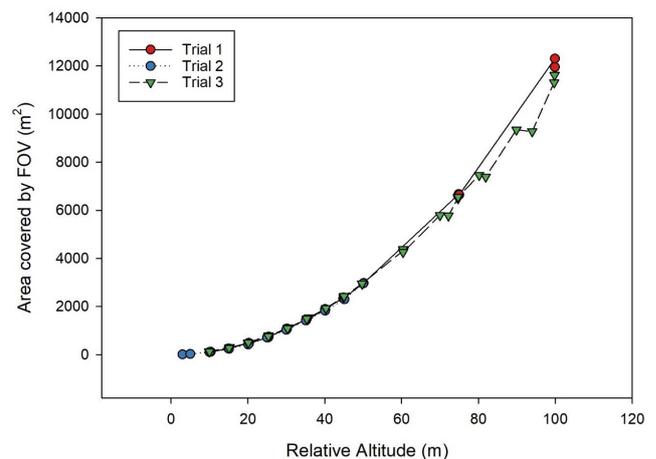
Figure 3: Calibration of the Field of View (FOV) of the drone with altitude above ground. The graph superimposes the results of three independent trials.

The Ground-Sampling Distance (GSD) of the drone camera, defined as the distance between pixel centres measured on the ground, was calculated for various altitudes using simple geometry. The GSD depends on the camera's focal length, on the width of the camera sensor, and on the altitude at which the drone is being flown. The variation in GSD, expressed in cm pixel^{-1}, with camera altitude is shown in Fig. 4.
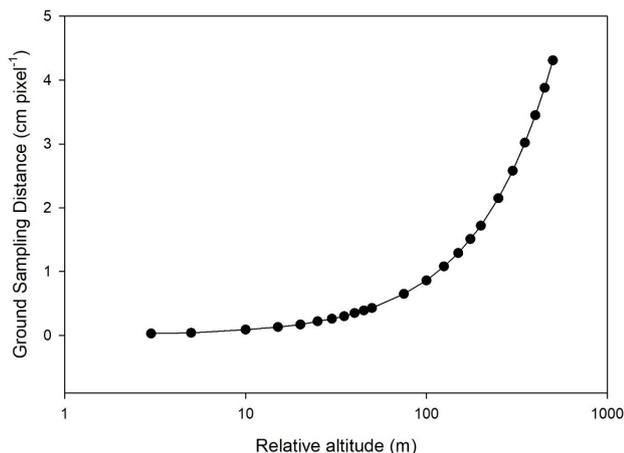
Figure 4: Variation in Ground Sampling Distance with altitude of drone camera above ground level.

### 3.3.3 Determination of Optimum Altitude for Survey

The optimum altitude for survey is dependent on the primary objectives of the survey. Identification of individual species would necessitate low-altitude surveys (3–10 m above ground level), whilst mapping of plant assemblages requires a larger FOV area and would therefore be carried out at higher altitudes (Putch, 2017). The altitude of survey also determines the minimum number of evenly-spaced transects that would be required to obtain sufficient overlap between transects, in order to allow orthomosaic maps to be compiled. After a number of trials, during which the drone was flown at different altitudes, a survey altitude of 30 m was noted to return images with sufficient detail for vegetation mapping, whilst remaining clear of all obstacles that were encountered, particularly trees. The spacing between transects was set to approximately 25–30 m, as this gave 49% overlap between images from neighbouring transects.

3.3.4 'High-Altitude' Image Capture

The flight paths along which the drone was flown during the survey were prepared on the online version of the Litchi app (flylitchi.com/hub). They were then uploaded to the drone at the time of survey. In the terminology used by Litchi, each survey flight constituted a 'Mission'. Missions were constructed using the 'Waypoints' tool in Litchi. These waypoints that the drone was set to visit were superimposed over a georeferenced base map showing the area of study (Fig. 5). The speed and height at which the drone would be flying were also predetermined and set for each waypoint. The drone was programmed to fly at an altitude of 30 m above the launching site at a constant speed of 5 km h^{-1}, taking

photos in DNG format of the ground surface at 5 s intervals. Camera aperture and shutter speed were set automatically, depending on ambient conditions. This relatively slow speed of travel minimised the image blur sometimes observed when the drone travelled at higher speeds, or when the camera shutter was set to a relatively slow speed. The speed and height of flight were kept constant for each of the seven parallel transect belts that were required to cover the whole AoS.

The drone survey was carried out on 28 June 2019 between 0800 and 0845. The whole AoS was imaged over two sequential missions lasting 22 minutes and 15 minutes, respectively. The segmentation of the survey into two separate Missions was necessary as the drone's battery life was limited to approximately 30 minutes. This necessitated a change of battery between Missions. The area imaged by the drone was larger than the AoS in order to ensure that no parts of the AoS were omitted from the imaging process.

The distances between a number of distinctive ground markers, including large rocks and anthropogenic structures, were measured in the field in order to enable calibration of the drone image during the processing phase.

3.4 Processing of Aerial Imagery

The 378 'high-altitude' images obtained from the drone were initially processed in Affinity Photo v.1.7 to remove lens distortion (Kim, Lee & Choi, 2015), and were subsequently imported into Agisoft Metashape Professional version 1.5.4 in order to generate an orthomosaic of the whole AoS. The process involved alignment of the photos, setting of a scale based on the relative positions of fixed ground markers, building of the dense cloud of points, building of the mesh, rendering of the texture, and exportation of the orthomosaic. The workflow is summarised in Fig. 6. The entire processing step required approximately 24 hours on the computer hardware available (Dell G5 5587, hexacore Intel Core i7-8750H processor, 16 GB RAM).

3.5 Segmentation of Orthomosaic

The orthomosaic was processed in order to generate a pseudocolour land cover map. Preliminary surveys of the orthomosaic had suggested that an initial segmentation of the image into three land cover categories would have been reasonable. The three categories selected were 'Vegetation', 'Bedrock' and 'Bare Soil'. In this preliminary study, the 'Vegetation' land cover category was not subdivided into different species or lifeforms, as this would have required multi-seasonal studies over several years. However, the initial results suggest that this would be feasible. The image was subsequently segmented using the k-means clustering plugin in Image J, with the value of $k = 3$, since three land cover categories were defined. This algorithm assigned each pixel in
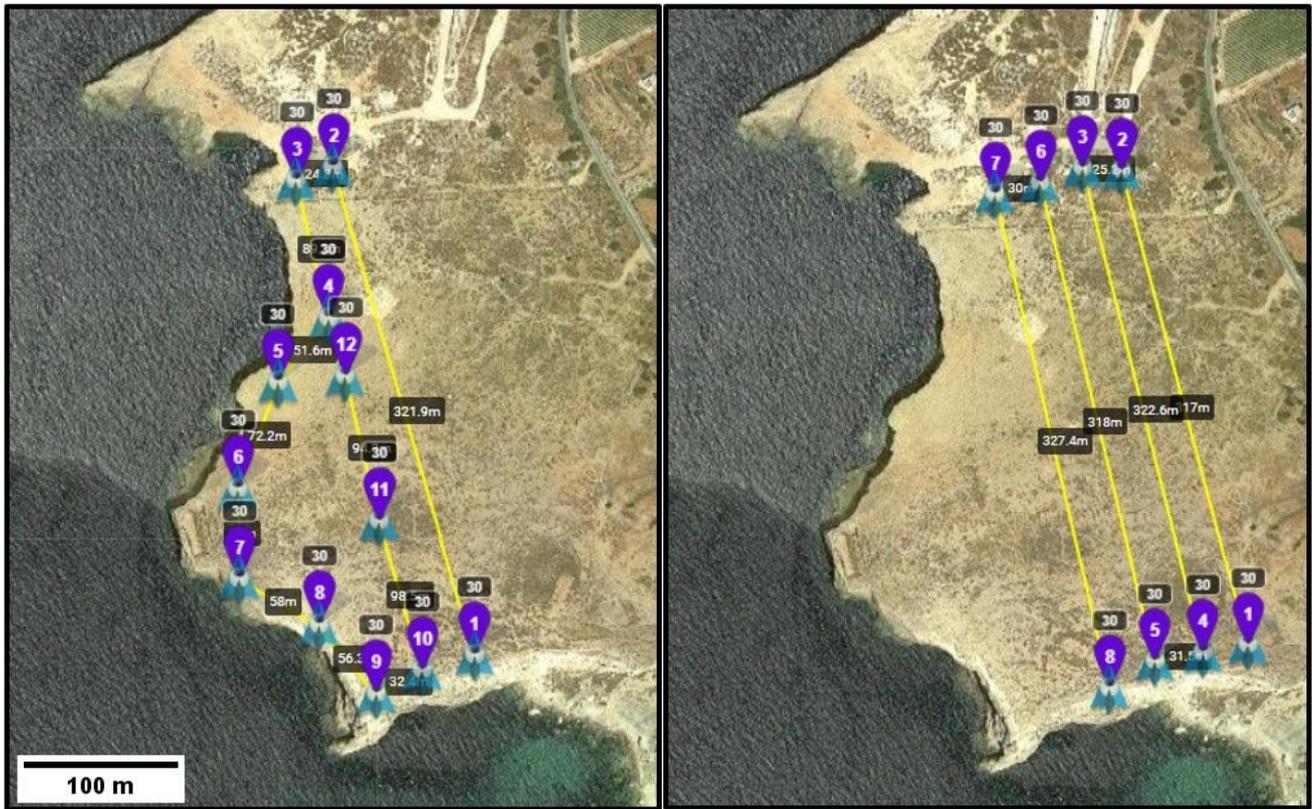
Figure 5: Path of drone transects over the area of study, as constructed in the Litchi app. The survey was split into two missions since a battery change was required after 25 minutes.

the image to an intensity value of 0, 1, or 2, depending on the intensity of its neighbouring pixels. This created an image with three distinct regions, each region presumably corresponding to one of the three land cover categories. For better visibility and to facilitate interpretation, this image was converted into a pseudocolour image by editing its lookup table and changing each of the three coverages into a primary colour (blue, green and red).

3.6   Verification

The effectiveness of the pseudocolour land cover map in predicting actual land cover categories in the field was then assessed. Inspection of the land cover map indicated 14 ROIs of varying areas in which there was either a preponderance of one of the three land cover categories, or in which an approximately symmetrical mixture of land cover was present. These ROIs were selected as areas in which the correspondence between the predicted land cover and observed land cover would be assessed. Each ROI was cropped out of from the land cover map. The relative proportion of 'Bedrock', 'Bare Soil' and 'Vegetation' cover in each ROI was determined by saving the cropped section as a text image. The num-

ber of pixels in each of the three land cover categories were enumerated and expressed as a proportion of the number of pixels in the cropped image. The land cover values obtained for each ROI are henceforth referred to as the 'predicted land cover' (pLC).

In order to measure the observed land cover, a Mission was created in which the drone was programmed to fly to each ROI, hover at an altitude ranging from 5–10 m above its central point, depending on the terrain, and record a 'low-altitude' ground photo. The altitude of the drone allowed the area of the FOV to be read off from the graph in Fig. 3. Therefore, on 2 July 2019 each of the 14 ROIs were visited and the programmed Mission implemented. The photo resolution of each ROI was such that dominant plant species identification was straightforward. The relative proportion of 'Bedrock', 'Bare Soil' and 'Vegetation' cover in each ROI photo was determined through image segmentation using k means clustering ($k = 3$) and verified through visual inspection. These values are referred to as the 'observed land cover' (oLC).

Align Photos

Accuracy: High

Pair preselection: Disabled

Key point limit: 400,000

Tie point limit: 0

↓

Build dense cloud

Quality: High

Depth filtering: Mild

↓

Build Mesh

Surface Type: Height field

Source: Dense cloud

Face count: Medium

Interpolation: Enabled
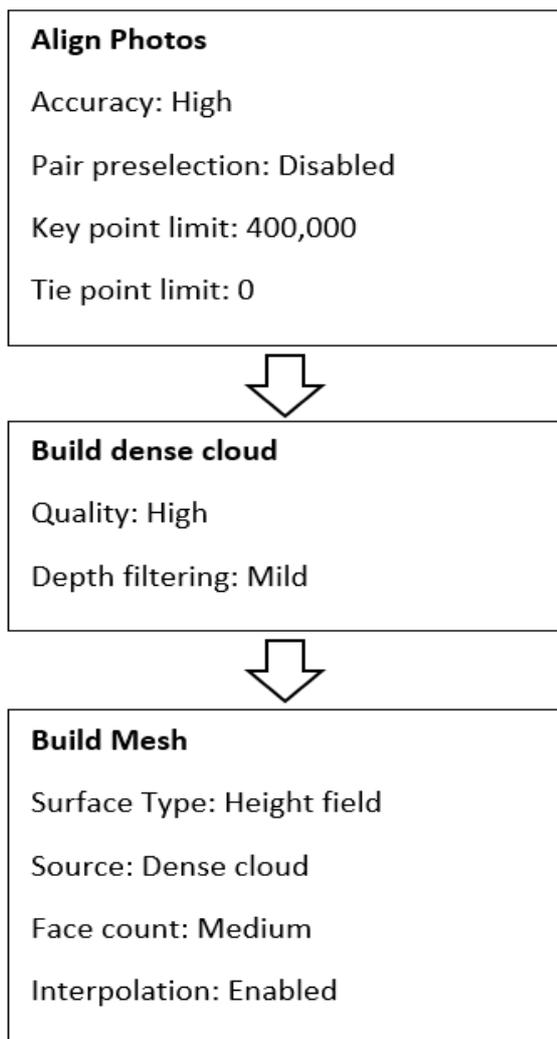
Figure 6: Summary of workflow followed during image processing phase.

3.7 Analysis of Data

The correspondence between the pLC and oLC values was tested by calculating the Pearson Product Moment Correlation Coefficient, after ensuring that the data was parametric. The relative contribution of the three land cover categories to differences within and between the pLC and oLC vales was assessed through Principal Component Analysis in Canoco v.5.12 (ter Braak & Šmilauer, 2018).

4 Results

4.1 Orthomosaic Map

The colour orthomosaic map of the AoS and its environs that was stitched from the 378 drone photos is shown in Fig. 7.

Figure 7: Orthomosaic map of the area of study and its environs. This image was stitched from 378 photos. The scale is given at the top right.

4.2 Land Cover Map

The Red-Green-Blue pseudocolour land cover map of the AoS is shown in Fig. 8. Every pixel in the orthomosaic was constrained to one of three intensity values that were subsequently converted into primary colours. The blue pixels represent 'Vegetation', red pixels are 'Bare soil' and green pixels are 'Bedrock'. The approximate positions of the 14 ROIs used for verification are indicated.

4.3 Correlation Between pLC and oLC

4.3.1 Presence-Absence of Land Cover Categories

When considering the binary presence or absence of land cover categories in each ROI, the overlap between the paired pLC and oLC in each ROI was complete. The oLC of every ROI matched the corresponding pLC.

4.3.2 Relative Coverage of Land Cover Categories

When the relative coverage of each land cover category in the ROIs was taken into account (Table 1), the correlations between the paired pLC and oLC were less pronounced (Fig. 9). Pearson product-moment correlation coefficients for the land cover categories were not statistically significant for any of the land cover categories (Table 2). When potential outlier ROI-7 was removed from the analysis, the correlation between the pLC and oLC for 'Vegetation' was statistically significant ($r = 0.568$, $p = 0.04$, $n = 13$). The corresponding correlations for 'Bedrock' and 'Bare soil' remained below the threshold of statistical significance.
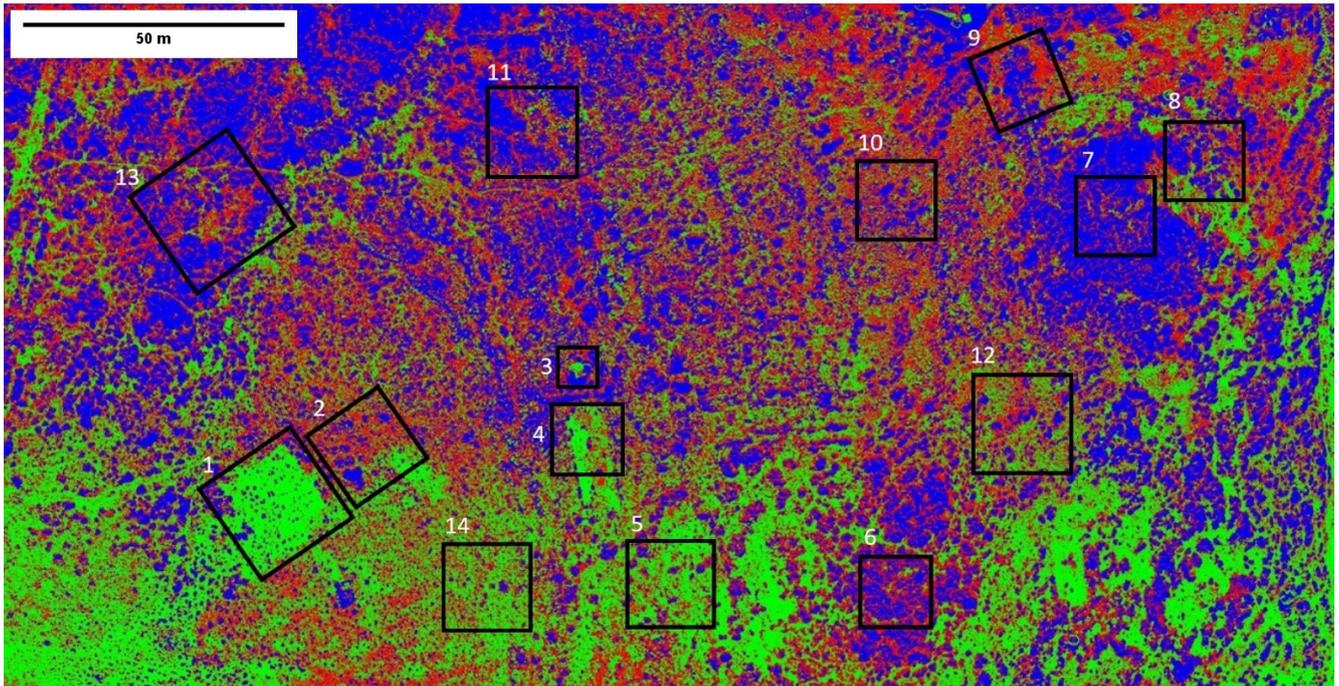
Figure 8: Pseudocolour map of the area of study segmented into three regions: 'Vegetation' (blue), 'Bare soil' (red) and 'Bedrock' (green).

Table 1: Predicted and observed land cover in the 14 ROIs.

ROI	Vegetation (predicted)	Vegetation (observed)	Bare soil (predicted)	Bare soil (observed)	Bedrock (predicted)	Bedrock (observed)
ROI_01	0.157	0.230	0.248	0.420	0.595	0.350
ROI_02	0.127	0.340	0.386	0.390	0.488	0.280
ROI_03	0.344	0.300	0.365	0.340	0.291	0.360
ROI_04	0.163	0.250	0.358	0.580	0.478	0.170
ROI_05	0.093	0.290	0.285	0.350	0.622	0.360
ROI_06	0.414	0.430	0.458	0.230	0.128	0.340
ROI_07	0.729	0.350	0.250	0.260	0.021	0.400
ROI_08	0.384	0.430	0.355	0.300	0.261	0.270
ROI_09	0.486	0.320	0.451	0.400	0.063	0.280
ROI_10	0.424	0.300	0.500	0.280	0.076	0.420
ROI_11	0.504	0.430	0.419	0.260	0.077	0.310
ROI_12	0.239	0.370	0.434	0.320	0.327	0.310
ROI_13	0.504	0.450	0.436	0.270	0.060	0.280
ROI_14	0.089	0.360	0.385	0.370	0.527	0.270

The data were further analysed using Principal Component Analysis (PCA), with potential outlier ROI-7 omitted (Fig. 10). The relative arrangement of ROIs on the PCA plot suggested that there was no particular correspondence in relative land cover composition between the paired pLC and oLC from each ROI. Moreover, the variation in relative land cover composition was much higher for the pLC than it was for the oLC. The maximum axis of variation of the convex hull, representing the pLC ROIs, coincided with the maximum axis of variation of the PCA plot (Axis I). Axis I of the PCA plot was most strongly correlated with the 'Bedrock' land cover category, suggesting considerable variability in assessment of this land cover category from the main orthophoto mosaic. Conversely, the oLC ROIs were mainly dispersed parallel to Axis II, which accounted for a much lower proportion of the total variation in the dataset. The 'Bare soil' and 'Vegetation' vectors were strongly correlated with Axis II.

Table 2: Pearson correlation (r) between predicted and observed land cover in each category. Data derived from the 14 ROIs. The statistical significance (p) and sample size (n) are given.

| LC Category	r	p	n
Vegetation	0.467	0.090	14
Rock	−0.276	0.339	14
Soil	−0.270	0.350	14
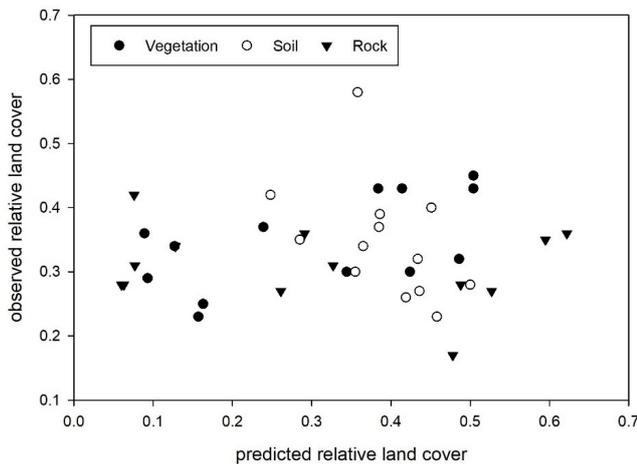
Figure 9: Correlation between the predicted and observed cover of the three land cover categories in each of the 14 ROIs. The ROI labels have not been included as these would have diminished the readability of the figure.
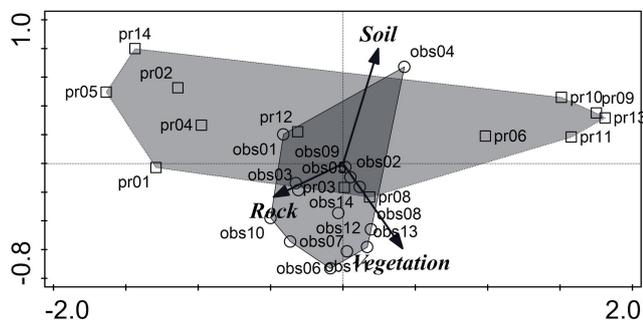

Figure 10: PCA ordination plot of the predicted and observed land cover in the 14 ROIs. Axis I and Axis II account for 83.62% and 12.59% of the variability of the data respectively.

5 Discussion

5.1 Correspondence Between Predicted and Observed Land Cover

Correspondence between the pLC and oLC was absolute when considered on a presence-absence basis. Although this may initially be perceived as a positive result, it should be emphasised that the land cover in the AoS was distributed in a mosaic pattern, and that this correspondence was not significantly different from a random pattern.

In terms of relative representation of the three land cover categories in each ROI, the only category that was significantly correlated across pLC and oLC was 'Vegetation'. This is a consequence of the optical reflectivity properties of plant foliage, which, in the AoS, reflected less light than other types of land cover and was therefore very distinct. Conversely, the optical distinction between 'Bedrock' and 'Bare soil' was much weaker as the soil in the AoS is similar to the parent rock. Furthermore, much of the soil is armoured with cobbles that are derived from the bedrock and therefore, in terms of reflectivity, are indistinguishable from it. Some regions of the land cover map that were interpreted as being covered by 'Bare soil' were actually characterised by 'Bedrock' mantled with a very shallow layer of gravel.

This raises a question regarding the choice of land cover categories. The sharp distinction between land cover categories that was evident in the land cover map was not observed in the field, where there was no clear boundary between 'Bedrock' and 'Bare soil'. The distinction in the land cover map was a consequence of the value of k in the clustering algorithm, as each pixel was constrained into one of only three discrete values. This therefore created abrupt boundaries that muted the range of intermediate land covers which were actually observed in nature.

This may also account for the much higher variability in the pLC data, relative to the oLC data. ROIs that may have been characterised by a high diversity of land cover categories at a large scale, turned out to be much less varied when examined at closer range. Once again, this is attributable to the optical distinction between 'Bedrock' and 'Bare soil'. In many cases, ROIs that were predicted to have relatively high 'Bare soil' cover were in fact observed to be characterised by relatively high 'Bedrock' cover, when examined from close range at low altitudes.

The 'Vegetation' category was less equivocal, although different species could obviously not be distinguished from each other when the value of k was too low for that to be feasible.

The optical characteristics of the 'Vegetation' category are also likely to vary considerably with season. The onset of the wet season would be correlated with the development of new foliage in perennial plants and the emergence of annual plants. This implies that much of the area categorised as 'Bare soil' would transition to 'Vegetation' during the wet season, reducing the land cover diversity of individual ROIs. In this situation, a value of $k = 3$ might no longer represent the optimum number of land cover categories. As such, this suggests that in order to give a more complete assessment of the efficiency of drone-based surveys, the study should be

repeated in different seasons, and preferably over several years.

5.2 Efficiency of the Method

5.2.1 Duration

The duration of the entire process, including verification and processing time for production of the orthomosaic and land cover map, was approximately 32 hours. This duration is much shorter than that required for a standard survey that produces the same output, where the comparable figure is several days or weeks.

5.2.2 Manpower and Expertise

The manpower requirements were also significantly lower, since the whole study was carried out by two persons (the authors), and could certainly be carried out by one. The comparable manpower requirements for a standard survey and map production would usually be higher, with one or two persons carrying out the field survey and at least one person producing the land cover maps. The level of expertise required to pilot a drone is probably not a major limiting factor. Although both authors had each flown well over a hundred Missions prior to conducting the present study, the level of expertise required to implement the Missions for the present study did not require that level of experience.

5.2.3 Cost

If the initial capital costs for the drone and computer hardware and software are excluded, the method tested during this study is more cost effective, less labour intensive and much more rapid than standard methods of field survey. It also generated a land cover map with a spatial resolution that would have required several weeks or months for a standard survey team to produce. However, it should be stressed that the present study did not attempt to place the land cover model in an absolute coordinate system and was only based on relative local coordinates. Locating the land cover model in an absolute frame of reference would have increased the cost considerably, as this would have necessitated the use of ground control points (GCPs) and determination of their precise geographic position. This process would require input from more personnel and would also necessitate the use of very costly equipment. Although the surveying equipment would represent a (considerable) one-time cost, the use of trained personnel to set the GCPs and determine their position would be recurring. As such, the cost-effectiveness of the method should be seen in this light.

5.3 Conclusions

The proposed method, as tested during the present study, highlighted several limitations of using a drone to survey land cover. However, almost all of these limitations were attributable to choices made at the processing stage. The drone imagery was detailed enough to permit identification of individual plant species from a low altitude, and of distinct assemblages from higher altitudes.

The relatively low correspondence between pLC and oLC, particularly for 'Bedrock' and 'Bare soil', was a consequence of the low value of $k = 3$ in the k-means clustering algorithm. This implies that patches of land comprising a mixture of land covers would be constrained into just one land cover category, with consequent reduction of the correlation between pLC and oLC. The low resolution returned by the clustering algorithm was a choice made by the authors and does not represent an inherent shortcoming of the proposed method. A higher value of k would have increased the resolution of the land cover map, but would have increased the probability of generating spurious land cover categories.

All in all, the utility of the land cover map depends on the purpose for which it was created. A survey of plant communities would be likely to require $k > 3$ to permit the identification of different assemblages, whilst for a binary survey of agricultural land use, $k = 3$ would probably suffice.

The preliminary study presented here suggests that $k = 3$ gives good correspondence at a large spatial scale, but is less accurate at smaller spatial scales. Nevertheless, for many purposes, it is the larger spatial scales that are of interest, and the performance of the land cover map at this larger scale was certainly satisfactory, compared to maps produced by standard survey methods. If information at smaller spatial scales is required, then using $k > 3$ will be necessary to highlight transitional land covers that may not be relevant at larger scales.

Acknowledgements

References

Anderson, K. & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Front. Ecol. Environ. 11*(3), 138–146.

David, L. C. G. & Ballado, A. H. (2016). Vegetation indices and textures in object-based weed detection from UAV imagery. In *Proceedings of the 2016 6ᵗʰ IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 669–674). Batu Ferringhi, Malaysia: IEEE.

Kim, B.-J., Lee, Y.-K. & Choi, J.-K. (2015). Investigating applicability of unmanned aerial vehicle to the

tidal flat zone. *Korean J. Remote Sens. 31* (5), 461–471.

Putch, A. (2017). *Linear Measurement Accuracy of DJI Drone Platforms and Photogrammetry.* San Francesco: DroneDeploy.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods, 9* (7), 671–675.

ter Braak, C. J. F. & Šmilauer, P. (2018). *Canoco reference manual and user's guide: software for ordination (version 5.10).* Wageningen: Biometris, Wageningen University & Research.

Wulder, M. A., Hall, R. J., Coops, N. C. & Franklin, S. E. (2004). High Spatial Resolution Remotely Sensed Data for Ecosystem Characterization. *Bioscience, 54* (6), 511–521.

Research Article

Axial Flux Permanent Magnet Motor Design and Optimisation by Using Artificial Neural Networks

Tuğçe Talay[*1] **and Kadir Erkan**[1]

[1] *Yildiz Technical University, Graduate School of Science and Engineering, Mechatronics Engineering, Istanbul, Turkey*

Abstract. In this study, the necessary steps for the design of axial flux permanent magnet motors are shown. The design and analysis of the engine were carried out based on ANSYS Maxwell program. The design parameters of the ANSYS Maxwell program and the artificial neural network system were established in MATLAB, and the most efficient design parameters were found with the trained neural network. The results of the Maxwell program were compared with the results of the artificial neural networks, and optimal working design parameters were found. The most efficient design parameters were submitted to the ANSYS Maxwell 3D design, the cogging torque was subsequently examined and design studies were carried out to reduce the cogging torque.

Keywords: AFPM, ANSYS Maxwell, cogging torque, design optimisation, efficiency, NNTOOL

1 Introduction

Nowadays, electric machines cover 65% of the world's energy consumption. Electric motors are widely used in home appliances, automotive, transportation vehicles, space and aerospace applications.

The increase in demand of these motors has driven many researchers to aim to reach the maximum energy efficiency during their operation.

In particular, there has been an increasing interest in axial flux permanent magnet synchronous motors due to their many advantages, such as compact machine construction with a short frame, high power density and high efficiency. Michael Faraday was the first to design the axial flux engine. In particular, the presence of magnets that provide high magnetic flux, such as neodium (NdFeB) and SmCo, and the reduction in production costs has contributed to the development of permanent magnet motors since the 1980s (Engin & Caner, 2009).

Nowadays, permanent magnet machines are used in many fields, from electric vehicles to space industries. In recent days, the world's leading automobile brands have been operating in the field of electric cars. Due to the high fuel costs and intensive R&D, activities are being carried out to increase the efficiency of the engines of these vehicles.

Firstly, in order to start the design process in electrical machines, design feasibility should be taken into consideration. Hence, presentation of a theoretical design for the perfect electric machine is not sufficient. Therefore, mass production should be thoroughly planned from the first stages of design. In this study, the design of the axial flux permanent magnet motor was created using the ANSYS Maxwell program. For design optimisation, variable ranges were assigned and the performance values were examined, so that an artificial neural network model could be established based on these data. Artificial neural networks can be modelled and learned, and are founded on the neuron structure of the human brain. Artificial neural networks are first trained using existing data and then tested with data not used during training. Although the training process takes time, after this training, the network will have developed a quick decision-making mechanism (Aydin, Zhu, Lipo & Howe, 2007).

The database was developed with various design parameters using the ANSYS Maxwell program, and these data were used in the training and control of artificial neural networks. Detailed design parameter ranges were inputted to the artificial neural network and subsequently compared with the results obtained from Maxwell, in order to determine optimal design parameters.

Correspondence to: Tuğçe Talay (tugcetalay2@gmail.com)

In line with the successful results obtained from the artificial neural network, the ANSYS Maxwell program was used to form a three-dimensional machine design, using the best efficient machine parameters. In addition, studies were carried out in order to reduce the impact moment of this machine, thus optimum design parameters were determined and a low-impact, high-efficiency axial intelligent permanent magnet motor design was created.

2 Modelling and Analysis of AFPM by ANSYS Maxwell

The ANSYS Maxwell 3D software package program was used for modelling and analysis of the axial flow permanent magnet motor.

The parametric values used for the starting model are given in Table 1. Here, the values given are the average initial values obtained as a result of a general literature review (Bouaziz, Jaafar & Ammar, 2016; Pahlavani & Gholinejad Omran, 2015; Darmani & Hooshyar, 2015).

The characteristics of the stator used for the analyses in this study consist of 18 slots, eight poles, one rotor, one stator, axial flux, rotor with a disc and permanent magnets placed on the surface. The following table shows the stator, windings, rotor, shaft, permanent magnets and operating conditions in the first selected condition.

Fig. 1 gives the output torque graph of the motor, depending on the position of the rotor in accordance with the selected initial parameters.

An efficiency ratio graph of the motor, depending on the position of the rotor in accordance with the selected initial parameters is given in Fig. 2.

Table 1: Parameters of reference AFPM Machine.

	Parameter	Values
Machine	Source Type	DC
	Structure	Axial-Flux Rotor
	Stator Type	AXIAL_AC
	Rotor Type	AXIAL_PM
	Air Gap Length (mm)	1

	Parameter	Values
Stator	Number of Poles	8
	Number of Slots	18
	Outer Diameter (mm)	120
	Inner Diameter (mm)	70
	Length (mm)	25
	Stacking Factor	0.95
	Steel Type	D23_50

Hs0 = 0.1 (mm)
Hs2 = 4.5 (mm)
Bs0 = 3.5 (mm)
Bs1 = 6.0 (mm)
Bs2 = 6.0 (mm)

	Parameter	Values
Winding	Winding Layers	2
	Parallel Branches	1
	Conductor per Slots	170
	Coil Pitch	1
	Number of Strands	2
	Wire Wrap	0
	Wire Size (mm)	0.269

	Parameter	Values
Rotor	Number of Poles	8
	Skew Width (deg)	0
	Outer Diameter (mm)	120
	Inner Diameter(mm)	70
	Length (mm)	25
	Stacking Factor	0.95
	Steel Type	D23_50

	Parameter	Values
PM	Embrace	0.7
	Magnet Type	XG96/40
	Magnet Length (mm)	25
	Magnet Thickness (mm)	8

	Parameter	Values
Shaft	Frictional Loss (W)	12
	Windage Loss (W)	12
	Reference Speed (rpm)	3000

	Parameter	Values
Working Conditions	Rated Output Power (W)	500
	Rated Voltage (V)	220
	Rated Speed (rpm)	1500
	Operating Temp. (cel)	75

Figure 1: Output torque graph according to rotor position.

Figure 2: Efficiency ratio graph according to rotor position.

3 Determination and Optimisation of Variable Parameters of AFPM

Inspired by a genetic algorithm method to increase the efficiency of AFPM, design variables and the results of crossing these variable design parameters will be evaluated.

The main design parameters affecting the efficiency are air gap (g), stator slot opening (Bs0), inner diameter of the rotor (Rin), and permanent magnet thickness (Lpm). One of the most important design parameters is λ, which is the result of dividing the motor inner diameter by the motor outer diameter (Di / Do). By choosing rotor initial diameter (Rin) as a design variable, λ will also be affected proportionally. Design values are assigned as Fig. 3 in Maxwell's "optimectrics" tab.

Figure 3: Variable design parameter.

As a result of the analysis, which takes about two days, 2401 data were obtained. The 2401 design parameters obtained were recorded in order to be used in the training and control of artificial neural networks.

4 Structure of the Artificial Neural Network

In this study, input data were generated by randomly selecting 427 of the design data produced in the ANSYS Maxwell program. Selected parameters were as follows: air gap (g), stator slot opening (Bs0), inner diameter of the rotor (Rin), permanent magnet thickness (Lpm), and their combinations.

The output data refers to the efficiency ratio of the values that are formed as a result of the design combination of these parameters.

In the training of the network, the back-propagation algorithm was used, thus the fault will be spread from outer layer to lower layers. Tangent-Sigmoid function was used for transfer function. The structure of the Artificial Neural Network is given in Fig. 4.
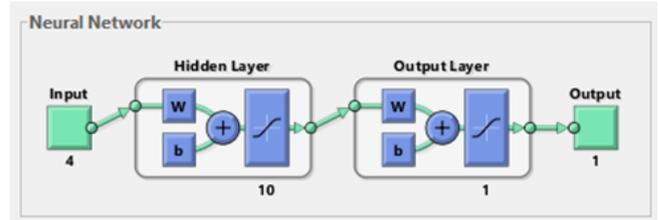

Figure 4: Structure of Artificial Neural Network.

4.1 Performance of the Artificial Neural Network

The neural network reached best validation performance at 35 epochs. In Fig. 5, the mean squared error histogram for training, test, and validation data is shown.

4.2 Regression Results of the Artificial Neural Network

In the study, the artificial neural network (ANN) method was used to find the regression value (correlation factor) between the target values and the calculated values, which was recorded between 0.9995 and 0.9996.
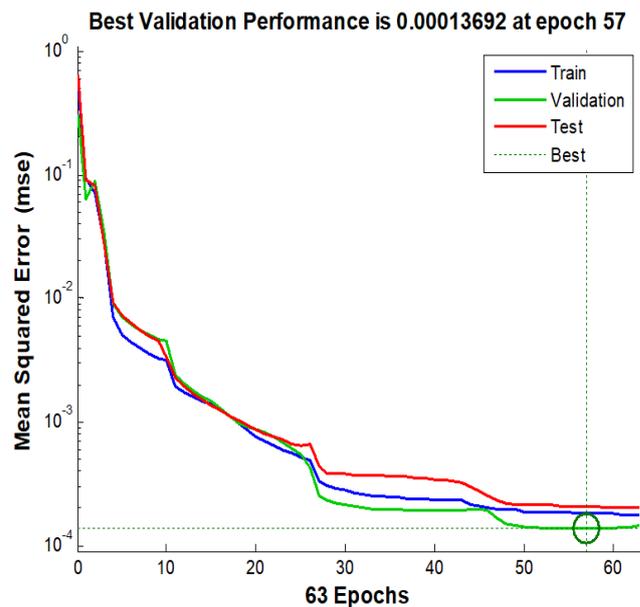

Figure 5: Performance of artificial neural network (main squared error).

The results obtained from the study show that the ANN method can be used successfully. The regression Values for Training, Test, Validation and All data are given in Fig. 6.
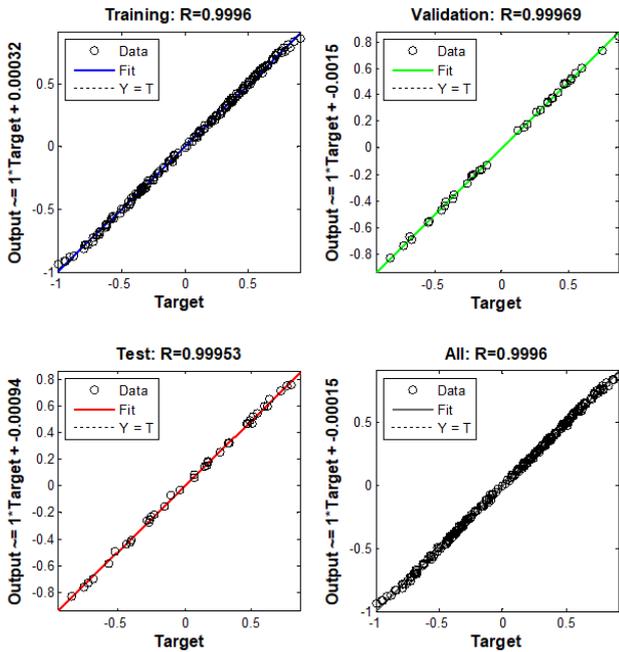

Figure 6: Regression Values for Training, Test, Validation and All.

4.3 Application of the Artificial Neural Network

According to ANN performance results, the training is completed without reaching the number of iterations entered. Our verification and test data are in parallel with the training data, proving that the network has learnt the process successfully. Therefore, network1 is ready to be exported into NNTOOL and simulated.

After proving that the artificial neural network was successfully working (tested and approved 427 pcs data), the aim was to obtain the highest efficiency motor by gathering efficiency ratio results for the different parameter values.

We used 427 of the 2041 data obtained from ANSY Maxwell for the training of the network. Subsequently, 133 different parameter combinations were chosen from the remaining data, so that the performance of the trained network and the amount of error could be investigated.

The selected design parameters were run in the trained network1 and the efficiency ratio values were thus predicted by the artificial neural network. The estimated efficiency ratio values recorded from the artificial neural network were compared to target values which were obtained from ANSYS Maxwell. Referring
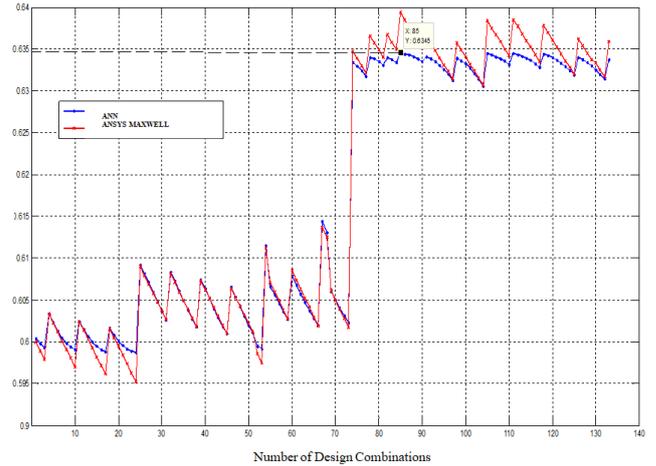

Figure 7: Target efficiency ratio vs. predicted ANN results (ANN predicted maximum efficiency ratio is marked).
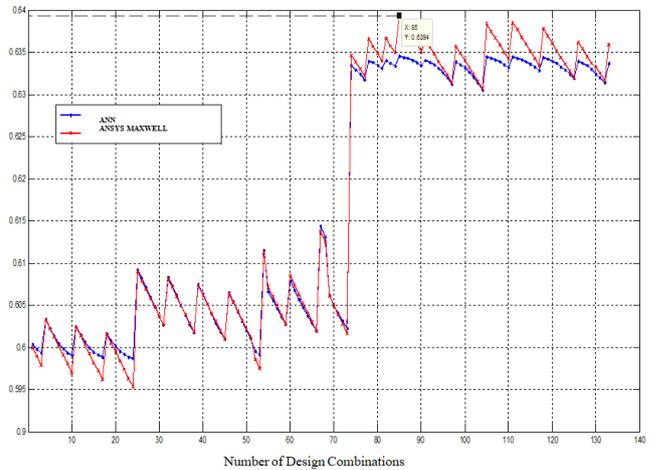

Figure 8: Target efficiency ratio vs. predicted ANN results (Target of maximum efficiency ratio is marked).

to Figs. 7 and 8, the target values from ANSYS Maxwell are parallel with the predicted values from the neural network. This proves that the trained network can be used for the design optimisations, as it confirms that the training process was completed successfully for different design values.

The results obtained from Maxwell in the artificial neural network reached the highest efficiency in the combination of parameters in design 85. According to these results, the 85th design parameters provide the highest efficiency. The 85th design parameters are shown in Table 2.

Table 2: Optimised design parameters.

Bs0 (mm)	G (mm)	Lpm (mm)	Rin (mm)
3	0.98	12	73

The efficiency, torque and losses of ANSYS Maxwell, which are optimised by artificial neural networks, were compared, and the design results of the initial stage were also compared with that of the optimised design results.

The efficiency increased by 2.05% compared to the initial stage. Total losses decreased by 28.66 Watts. Maximum output torque increased by 0.1024 N m. Initial value and optimized comparison are given in Table 3.

Table 3: Initial value and optimised design comparison.

	Initial Value	Optimised Value
Efficiency (%):	61.7592	63.8116
Maximum Torque (N m):	3.9364	4.0388
Total Loss (W):	340.552	311.892

Initial torque value and optimized torque comparison is given in Fig. 9. Initial state efficiency and optimized efficiency ratio comparison is given in Fig. 10.

The optimisation of the axial flow of permanent magnet motor in ANSYS Maxwell are the parameters that give the highest efficiency in the selected range. Using these parameters, the design of the engine was submitted to ANSYS Maxwell three-dimensional machine design.

5   3-D Design Modelling of Axial Flux Permanent Magnet Machine by ANSYS Maxwell

After analysing and modelling in Maxwell rmexport, the design was submitted to Maxwell3DDesign. In Fig. 11, a half-model three-dimensional axial intelligent permanent magnet machine is shown.
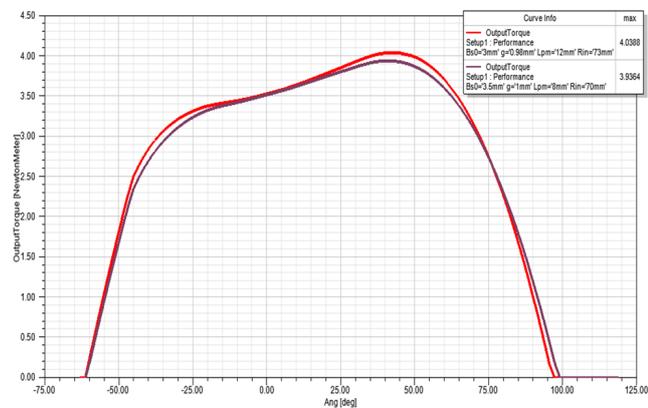

**Figure 9:** According to rotor position, initial state and optimised torque comparison.
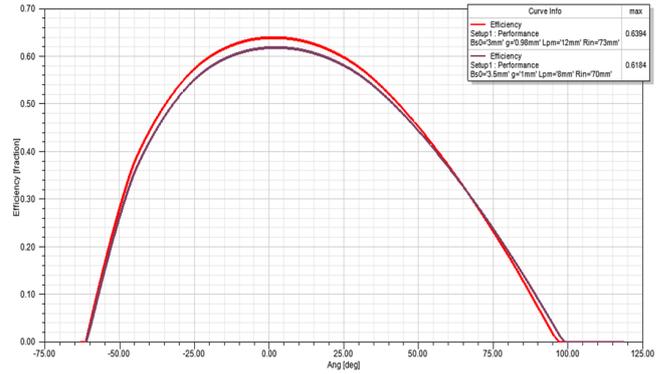
**Figure 10:** According to rotor position, initial state and optimised efficiency ratio comparison.
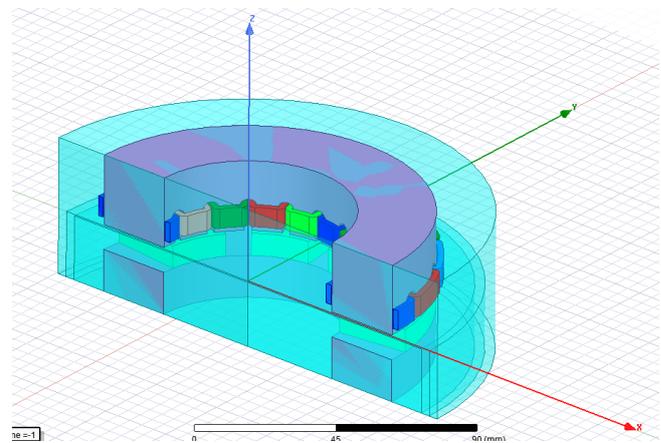
Figure 11: Half-model of permanent magnet motor with axial flow.

6   Calculation of Cogging Torque

Cogging torque is the reluctance moment of the motor due to its magnetic flux and reluctance. This moment is examined structurally and has a negative effect on nominal torque in the operation of the motor. In the simulation of the permanent magnet axial flow motor with ANSYS Maxwell program, different settings must be made in each step. In the calculation of the moment of torque, the motor needs to be moved with mechanical steps (Ayçiçek, Bekiroğlu, Senol & Oner, 2015). For a higher resolution, this setting can be arranged optionally. Cogging Torque in the initial stage values are given in Fig. 12.

7   Cogging Torque Reducing Techniques of AFPM

Minimising cogging torque in designing axial-flux permanent-magnet (AFPM) motors is one of the important issues which must be considered during the design process, as it leads to output torque ripple, vibrations and noise (Kumar & Srivastava, 2018a). There are a variety of techniques that exist for minimising the
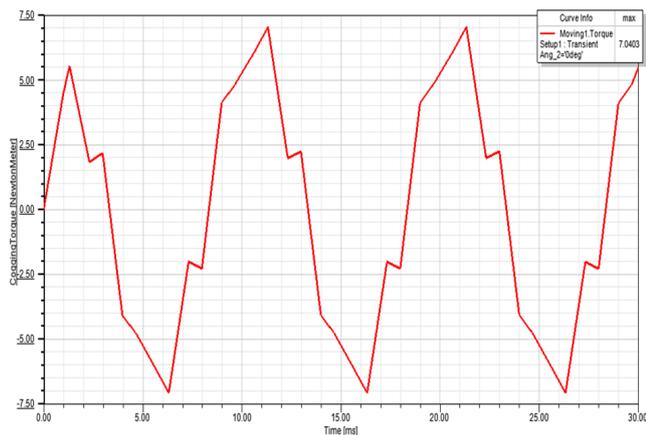
**Figure 12:** Cogging Torque in the initial stage.
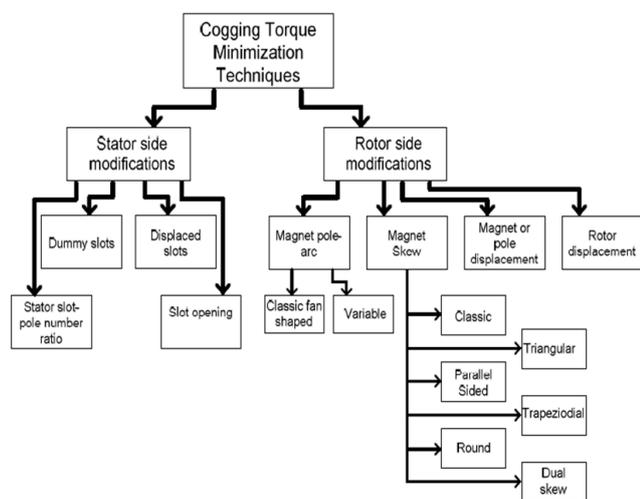
Figure 13: Summary of cogging torque minimisation techniques for AFPM machines (Aydin, 2008).

cogging torque (Bianchini, Immovilli, Bellini & Davoli, 2012). Out of these, the various rotor modification techniques are directly applicable to axial flux machines. However, the stator modification techniques cannot be directly applied, due to difficulty in punching slots in axial stator laminations, leading to high cost (Kumar & Srivastava, 2018a). In general, cogging torque minimisation in AFPM machines can be considered in two manners: modifications from the stator side and rotor side. Summary of cogging torque minimization techniques for AFPM machines are given in Fig. 13.

7.1 Stator Side Modifications

There are some minimisation techniques which can be applied for the stator side, such as dummy slots in stator teeth, ratio of stator slot number to rotor pole number, slot openings and displaced slots (Aydin, 2008).

It is known that cogging torque is produced by the interaction between the edges of the magnet poles and the stator slots. By introducing dummy slots in the stator teeth, the interaction between magnets and stator slots will be increased, thus increasing the main frequency, removing some of the harmonic orders, and reducing the maximum amount of cogging torque. This method can even cause saturation on teeth surface by changing the distribution of flux lines (Bianchini et al., 2012). The dummy slots are not as deep as the slots where windings are positioned, but if feasible, for cogging torque mitigation, small notches in the teeth are sufficient (Ferreira, Leite & Costa, 2015). Due to the high number of poles and consequently, slots required, the width of the teeth would not be enough to accommodate the notches in low speed applications (Ferreira et al., 2015). Furthermore, these methods are not practical and they are high cost solutions for AFPM machines.

One of the cogging torque minimisation methods used in the stator, was to arrange the ratio of stator slot number to pole number. However, in fractional slot machines, the rotor magnets have different positions according to the stator slots, hence generating the cogging torque components which are out of phase refer to each other, are partially cancelled (Aydin, 2008). In axial flux machines, the stator laminations are spirally wound and thus, cutting of slots is difficult. Therefore, its customised design for cogging reduction is far more complex and not practical, as the design cost is high (Kumar & Srivastava, 2018b).

Due to reasons previously mentioned, the axial fluxes we designed were analysed with a focus on rotor side modifications, in order to reduce the impact of the cogging torque.

7.2 Rotor Side Modifications

For AFPM motors, it is more important to work on techniques that will reduce the cogging torque on the rotor side, rather than stator side techniques. This is due to the fact that the motor structure is axial and thus, the modifications required on the rotor are more practical, cost less and are easier to design.

These methods include: magnet shaping, optimising magnet pole arc to pole pitch ratio, pole arc offset, narrowing rotor pole embrace, pole-slot combination, magnet skewing, etc. (Kumar & Srivastava, 2018b). As a result of rotor shifting technology, the rotor torque is reduced at the appropriate angle. Magnet skewing technique is a method in which one side of the magnet is attached to the other side of the magnet; results are based on the production of the magnet according to the resulting new shape. The magnet or pole shift (grouping technique) is the shifting of the adjacent magnets in relation to each other.

7.2.1 Cogging Torque Minimisation by Magnet Pole-Arc to Pole-Pitch Ratio

Cogging torque is being triggering by the interaction between the edges of the magnet poles and the stator slots. Magnet pole-arc is an important topic because both the magnitude and shape of the cogging torque waveform depend on the magnet pole-arc (Aydin, 2008). The magnet leakage flux can be reduced by decreasing the magnet pole-arc to pole pitch ratio, however, this also reduces the magnetic flux and consequently, the average torque (Aydin, 2008). An AFPM rotor with two different magnet steps is shown in Fig. 14. Cogging Torque optimized values after Magnet Pole-Arc to Pole-Pitch Ratio technique are given in Fig. 15.

7.2.2 Cogging Torque Minimisation by Magnet Skew

One of the most efficient methods to reduce cogging torque is Magnet skewing for AFPM machines, because this modification can be implemented quickly with ease. Magnet skewing is generally used for conventional permanent magnet motors, as well as AFPM motors (Güleç, Yolaçan, Demir, Ocak & Aydin, 2016). As shown in Fig. 16, the location of the direct axis of the magnet shifts gradually in peripheral direction, as it is traced from inner radius to outer radius of the rotor core (Reza & Srivastava, 2018). The machine is divided into a certain number of annular slices in the radial direction. The direct axis of each slice is displaced by a certain angle (θs) (Reza & Srivastava, 2018).
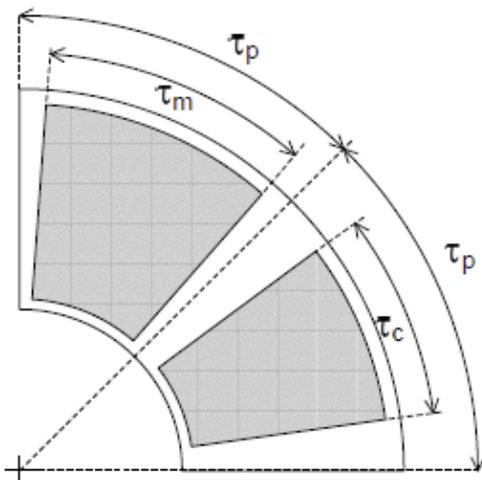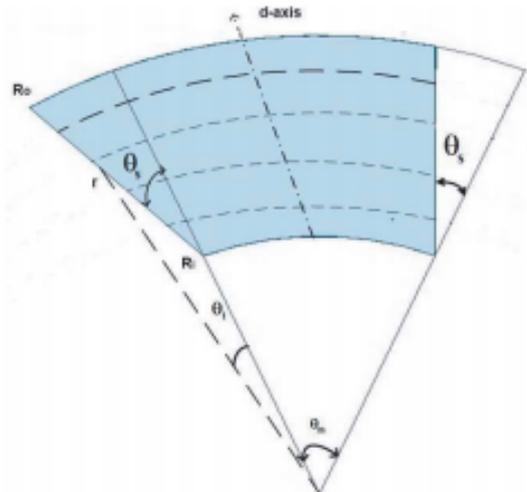
Figure 16: Rotor structure with skewed magnet (Reza & Srivastava, 2018).

Different registration techniques are available for AFPM motors. These techniques include conventional magnet skew, triangular magnet skew, parallel-sided magnets, trapezoidal magnet skew, circular magnets and dual skewed magnets. Such methods are just some of the low-cost techniques that can be applied to the rotor side of the permanent magnet disc motors. Cogging Torque minimized values following application of the skewing technique, are shown in Fig. 17.

7.2.3 Cogging Torque Minimisation by Rotor Displacement

In dual air-gap AFPM motors, the cogging torque is equal to the vector sum of the cogging torques in each air gap. Therefore, if one of the rotors is rotated relative to the other, the peak value of the total cogging torque can be pulled down, since the vector sum of the cogging torque in each air gap will give the total cogging torque (Ayçiçek, 2012). Fig. 18 shows the comparison between two different AFPM motor's rotor displacement. Cogging Torque minimized values are shown in Fig. 19 after rotor displacement technique.

Figure 14: An AFPM rotor with two different magnet steps.

Figure 15: Cogging Torque minimisation after Magnet Pole-Arc to Pole-Pitch Ratio technique.
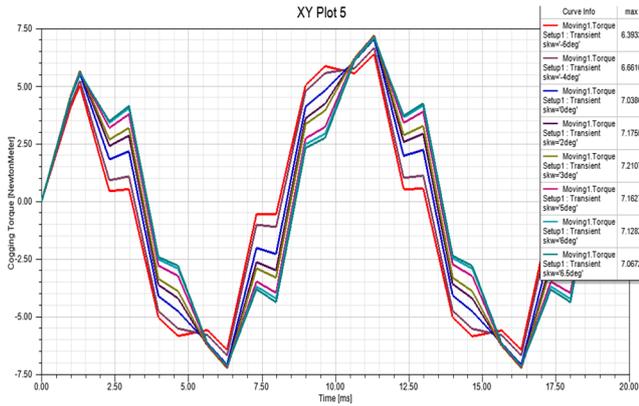
Figure 17: Cogging Torque minimisation after skewing technique.
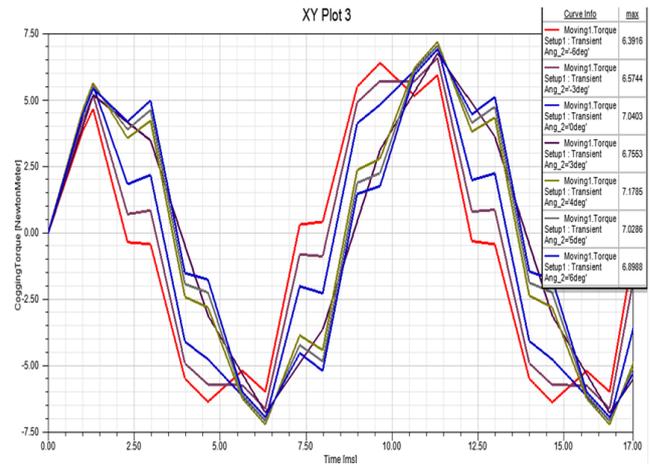
Figure 19: Cogging Torque minimisation after rotor displacement technique.
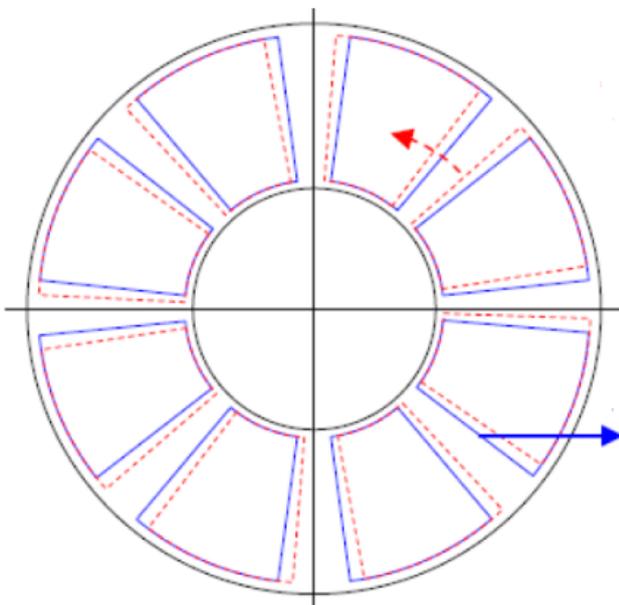
Figure 18: Comparison of two different AFPM motor's rotor displacement (Aydin, Zhu, Lipo & Howe, 2007).
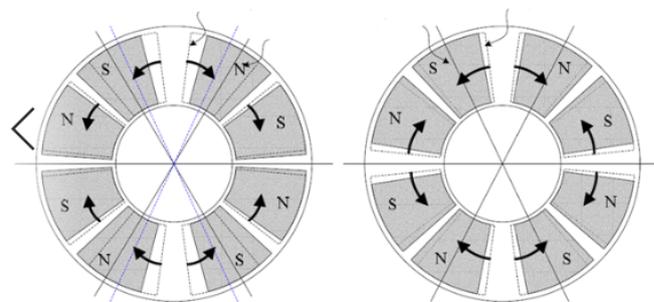
Figure 20: Magnet placement (pole displacement) options of AFPM motors (Aydin, Zhu, Lipo & Howe, 2007).

7.2.4 Magnet or Pole Displacement (Grouping Technique)

Another effective method for cogging torque minimisation is to slide adjacent magnets relative to each other, as shown in Fig. 20. However, when the magnets are shifted from their symmetrical position, the leakage flux on one side of the magnet will increase, whilst the leakage flux on the other side will decrease. Another disadvantage of this method is that it causes distortion of the counter electromotive force waveform and the moment fluctuation (Aydin et al., 2007).

8 Conclusions

In this paper, an artificial neural network-based design optimisation algorithm for an AFPM motor was proposed. The design variables were varied, and the subsequent effects on performance indices were studied by performing ANSYS Maxwell. Having obtained discrete data from ANSYS Maxwell, NNTOOL was used for training the data and further optimisation. The results of the network disclosed an optimised AFPM motor design.

The trained network was used to validate the artificial neural network-based design solution. Results from the artificial neural network prove the precision of the design variables by getting maximum effective design solution at the same design variables with ANSYS Maxwell data by using artificial neural network. We analysed so many design data in a very short time and the design variables which gives maximum efficiency, has been found. With the artificial neural network design solution, we were able to increase efficiency by 2.05% and the maximum output torque by 0.1024 N m compared to the initial stage design.

For AFPM, cogging torque is crucial as it needs to be reduced to improve the design optimisation. In this study, various cogging torque minimisation techniques for AFPM machines were reviewed. It is evident that there are a range of different techniques which are applicable for AFPM machines and can be used to reduce the cogging torque. Due to the complexity and cost of stator manufacturing, this study focussed on rotor side minimisation techniques. Results showed that adjustment of Magnet Pole-Arc to Pole-Pitch Ratio resulted in reduction of cogging torque by 10.1%, that the magnet skewing technique produced a 9.17% decrease and that following rotor displacement, cogging torque decreased by 9.2%, compared to initial structure.

References

Ayçiçek, E. (2012). *Design of Axial Flux Permanent Magnet Synchronous Motor with Reduced Cogging Torque* (Doctoral dissertation, Yildiz Technical University, Institute of Science, Technology, Department of Electrical and Electronics Engineering, Istanbul, Turkey).

Ayçiçek, E., Bekiroğlu, N., Senol, I. & Oner, Y. (2015). Rotor Configuration for Cogging Torque Minimization of the Open- Slot Structured Axial Flux Permanent Magnet Synchronous Motors. *Appl. Comput. Electromagn. Soc. J. 30*(4), 396–408.

Aydin, M. (2008). Magnet Skew in Cogging Torque Minimization of Axial Gap Permanent Magnet. In *2008 18th International Conference on Electrical Machines*. Vilamoura, Portugal: IEEE.

Aydin, M., Zhu, Z. Q., Lipo, T. A. & Howe, D. (2007). Minimization of cogging torque in axial-flux permanent-magnet machines: Design concepts. *IEEE Trans. Magn. 43*(9), 3614–3622.

Bianchini, C., Immovilli, F., Bellini, A. & Davoli, M. (2012). Review of design solutions for internal permanent-magnet machines cogging torque reduction. *IEEE Trans. Magn. 48*(10), 2685–2693.

Bouaziz, O., Jaafar, I. & Ammar, F. B. (2016). 3D Finite Element Modelling of Single-Sided Axial-Flux Permanent-Magnet Synchronous Machine. In *3rd International Conference on Automation, Control,*

Engineering and Computer Science (ACECS'16) (Vol. Proceedings of Engineering & Technology (PET), pp. 669–674). Hammamet, Tunisia.

Darmani, M. A. & Hooshyar, H. (2015). Optimal Design of Axial Flux Permanent Magnet Synchronous Motor for Electric Vehicle Applications Using GA and FEM. *J. Electr. Comput. Eng. Innov. 3*(2), 89–97.

Engin, H. & Caner, A. (2009). An Overview Control Methode of Axial Flux Motors. In *5. Uluslararasi ileri Teknolojiler Sempozyumu (IATS'09), 13–15 Mayis 2009* (Vol. IATS'09). Karabük, Turkey.

Ferreira, A. P., Leite, A. V. & Costa, A. F. (2015). Comprehensive Analysis and Evaluation of Cogging Torque in Axial Flux Permanent Magnet Machines. In *2015 IEEE 10th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*. Guarda, Portugal: IEEE.

Güleç, M., Yolaçan, E., Demir, Y., Ocak, O. & Aydin, M. (2016). Modeling based on 3D finite element analysis and experimental study of a 24-slot 8-pole axial-flux permanent-magnet synchronous motor for no cogging torque and sinusoidal back-EMF. *Turkish J. Electr. Eng. Comput. Sci. 24*(1), 262–275.

Kumar, P. & Srivastava, R. K. (2018a). Cost-Effective Stator Modification Techniques for Cogging Torque Reduction in Axial Flux Permanent Magnet Machines. In *2018 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific)*. Bangkok, Thailand: IEEE.

Kumar, P. & Srivastava, R. K. (2018b). Influence of Rotor Magnet Shapes on Performance of Axial Flux Permanent Magnet Machines. *Prog. Electromagn. Res. C, 85*, 155–165.

Pahlavani, M. A. & Gholinejad Omran, H. R. (2015). A New Analytical Description and FEA Validation of an Effective Method to Reduce the Cogging Torque in SM-AFPM Motors. *Prog. Electromagn. Res. M, 42*, 189–197.

Reza, M. M. & Srivastava, R. K. (2018). Semi-Analytical Model for Skewed Magnet Axial Flux Machine. *Prog. Electromagn. Res. M, 68*, 109–117.

Research Article

The Evolution of Malta's Tourism Sector

Silvio Attard[*1]
[1]*Central Bank of Malta, Pjazza Kastilja, Valletta, Malta*

Abstract. The study aims to analyse historic developments in Malta's tourism industry, focussing on the changing characteristics of demand and supply. The recent surge in inbound tourism appears to be largely driven by increased air connectivity to and from Malta. The advent of low-cost carriers is considered an important positive supply shock on the local sector. At the same time, the sustained shift towards stays in private accommodation can be partly explained by changing preferences, but also by capacity constraints in collective accommodation establishments. Moreover, the paper compares a number of indicators which shed light on the economic importance of tourism for the Maltese economy. It also examines the issue concerning sustainable growth of the sector and seeks to draw some policy inferences.

Keywords: tourism economics, demand factor, supply factors, over-tourism, sustainability

JEL Classifications: R41, Z31, Z32, Z38

1 Malta's Tourism Development: A Tourism Destination Life Cycle Case-Study

The development of tourism in Malta started in the late 1950s, as the Maltese authorities sought to identify new areas of economic development in view of the diminishing role of the British military base in Malta. The Government initiated a programme of capital spending aimed at improving access to beaches, together with increased promotion and advertising. The ensuing increase in passengers and aircraft movements necessitated the construction of an air terminal for civilian activity in 1956, with the first passenger terminal inaugurated in 1958.

In the beginning, initiatives such as the setting-up of the Malta Government Tourist Board in 1958, had led to moderate increases in the number of tourist arrivals, partly due to the limited funds made available. However, the adoption of a grants programme to encourage business in the hotels sector in the mid-1960s, even though more costly, was very successful, with the sector expanding further through the seventies. This was followed by a period of decline partly driven by the ensuing global recession, especially in the UK in the early 1980s, which had resulted in a sharp drop in arrivals from the UK. Since then, arrivals have increased almost uninterruptedly.

The evolution of tourism in Malta can be analysed in the light of Butler's tourism destination life cycle model (Butler, 1980). According to this model, visitors will come to an area in small numbers initially, restricted by lack of destination access, facilities, and limited awareness of what the country offers. As facilities are provided and awareness grows, visitor numbers will increase. With marketing, information dissemination, and further facility provision, the area's popularity will grow rapidly. Eventually, however, the rate of increase in visitor numbers will decline as levels of carrying capacity are reached. These may be identified in terms of environmental factors (e.g. land scarcity, water quality, air quality), physical plant (e.g. transportation, accommodation, other services), or social factors (e.g. crowding, resentment by the local population). As the attractiveness of the area declines relative to other areas, because of overuse and the impacts of visitors, the actual number of visitors may also eventually decline unless steps are taken to rejuvenate the tourist product. These stages are depicted in Fig. 1.

*Correspondence to*: Silvio Attard (attardsl@centralbankmalta.org)
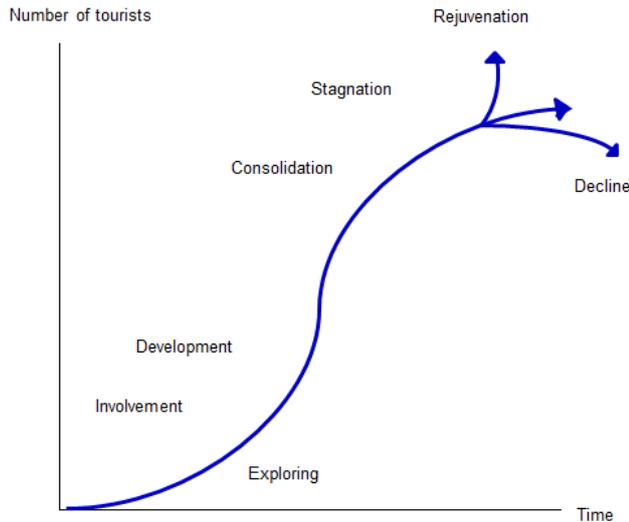
Figure 1: Butler's lifecycle curve.

Figure 2: Total inbound tourists. Source: Central Bank of Malta.

Fig. 2, which shows the number of tourists in Malta since 1960, conforms to Butler's model. The development of the tourism industry involved tackling two main challenges, namely: the insufficient skills and education of the local population, and the lack of infrastructure. In fact, initially Malta was predominantly visited by the British to meet friends and family members stationed on the islands. As the numbers of visitors increased, Malta moved on to the involvement stage, where policymakers, having considered the potential of this industry, took steps to support tourism-friendly facilities. In Malta, this stage was characterised by notable private investment in the tourism industry; international hotels chains such as the Hilton and Sheraton groups started operating locally. As a result, different forms of employment such as travel agencies, real-estate, banking, insurance, leisure, and retail outlets were generated.

As the number of visitors continued to rise, the country moved to the development stage where heavy advertising takes place and the number of tourists in peak periods is likely to be equal or exceed the permanent local population. At this stage, mass tourism had developed with structures of hotels built in a way to accommodate the peaks being reached during summers. The idea of Malta as a sun, sea and sand destination was already creating the problem of seasonality, especially where the labour force was concerned. The problems exacerbated to a point where Malta started to experience a shortage of water supply, a problem of sewage disposal in the peak months of summer, and, high dependence on British visitors.

The first half of the 1980s was characterized by a sharp decline in the number of arrivals. By 1984, the total number of tourists declined by about a third when compared with the high point reached in 1980. This slowdown was not exclusive to Malta, as other Southern European countries also experienced a lacklustre performance in their respective tourism markets in the wake of the second oil shock and deep recession in the UK (Mifsud, 2006). However, in the case of Malta, a strong exchange rate vis-á-vis competing 'sun currencies' (Spanish peseta, Portuguese escudo, Greek Drachma, and Italian lira) also led to a loss in cost competitiveness, particularly for UK package tourists. The latter constituted the largest segment of all visiting tourists.

By the end of the decade, the losses in tourism were reversed. The market had returned to growth, mainly due to the authorities' implementation of a diversification strategy in source markets for tourists. This included more competitive packages offered by industry operators, as well as a dual exchange rate in respect of the British pound, which was introduced to reduce the impact of the Maltese Lira appreciation. As the country reached the mass tourism market, it moved to the consolidation stage where visitors rose above the one million mark in 1992. This momentum was maintained until 1995 when the total number of tourist arrivals fell for two consecutive years, mainly due to a reduced demand from some British tour operators.

From the mid-1990s through the mid-2000s the number of visitors reached a plateau and this can be considered Malta's stagnation stage according to Butler's model. The authorities worked hard to avoid going into the decline stage shown in Fig. 1. With the setup of the

Malta Tourism Authority in 1999, Government aimed to improve the Maltese tourism product by shifting attention from a practically exclusive sun and sea destination to a more segmented industry niches based on improvements in the quality of tourism services. The advent of low cost airlines in 2006 increased route connectivity with mainland Europe. The effects of the air transport reform were complemented in subsequent years by the woes of competing destinations in the wake of the Arab Spring. As a result of this during the past decade, the Maltese tourism product experienced the rejuvenation stage, with the industry returning to very strong growth.

Assessing the evolution of tourism trends in Malta is necessary to understand the overall economic development of the country and towards the development and well-being of Malta's society. Thus, the inter-winding of sociocultural, demographic, economic and environmental factors connecting the tourism sector are of relevance to any policymaker.

2 The Changing Characteristics of Inbound Tourism in Malta

2.1 Source Markets and Profile of Visitors

Historically, reflecting the ties between Great Britain and Malta, the local tourism industry was practically exclusively reliant on the UK. Indeed, during the 1960s, when the first tourism statistics started to be compiled, almost 75% of all visitors were UK nationals (see Fig. 3). Arrivals from other European markets were mainly from Italy, and to a lesser extent from Germany. This situation more or less prevailed until the late 1980s.
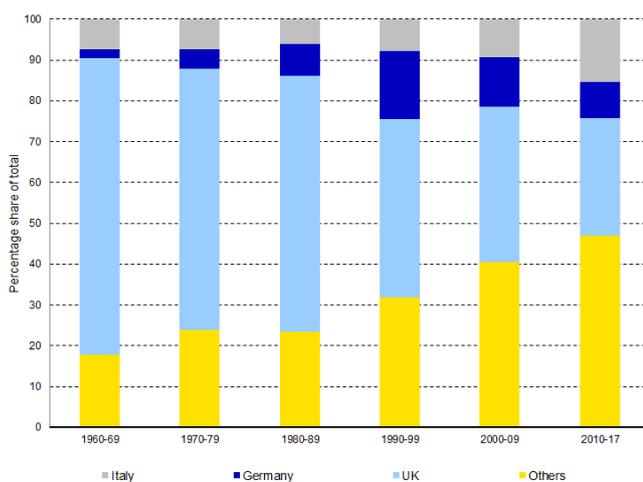
Figure 3: Source market of inbound tourists (figures for Germany pre-1990 represent figures for West Germany). Source: NSO, authors' own calculations.

The composition of tourists began to change significantly during the 1990s as the authorities responded with diversification strategies aimed at reversing the decline in tourism activity recorded in the previous decade.

By 2017, when a record of 2.3 million tourists visited Malta, the UK's market share, albeit still the largest, had fallen to close to 25%. Meanwhile, for the first time, arrivals from Italy and Germany collectively exceeded UK visitors, and tourists coming from the remaining source markets represented more than half of total visitors. The decreased reliance on few source markets is largely credited to increasing airline connectivity, but also the authorities' ability to diversify the Maltese tourism product.

Leisure tourism remains the main purpose of visit for the vast majority of arrivals, with its share increasing from 83.7% in 2010 to 85.3% of total inbound tourists in 2017, largely reflecting the expansion in self-catering accommodation facilities and internet marketing. Moreover, over the past decades, in order to reduce the strong seasonality factor, which is a major issue throughout the Mediterranean, Maltese policymakers adopted smart business diversification approaches to target non-leisure segments. From 2010 to 2017 the number of visitors for business purposes increased by over 62,000 persons, accounting for 7.9% of total inbound tourism. Meanwhile, the share of the "other" tourist segment, which includes tourism for educational, religious and health-related purposes, stood at 6.8%, accounting for more than 154,000 visitors in 2017.

Despite the increase in leisure tourism, there has also been some diversification within the Maltese holiday product itself. Indeed, based on the 2017 Malta Tourism Authority's (MTA) Market Profile Survey (Malta Tourism Authority, 2018), only 15.7% chose Malta as their destination solely for the traditional 'sun and sea' factor. The largest share of tourists, standing at 42.9% chose Malta for its culture and heritage. Moreover, important tourism niches such as, health, wellness, scuba diving and other sports segments were mentioned by a further 12.7% of the tourists surveyed as their main motivational factor in choosing Malta as their destination (see Fig. 4).

2.2 Nights Spent in Malta – Private Versus Collective Establishments

Over the ten year period to 2010, total nights spent by inbound tourists remained broadly stable (see Fig. 5). From 2001 to 2008, total nights stayed averaged 10.9 million nights per year, before declining to 9.9 million nights in 2009 on the back of dwindling demand spurred by the global financial crisis. The effect of the crisis was limited to a single year and the losses of more than one million nights in 2009 was more than recouped in the following year. Subsequently visitor nights spent in
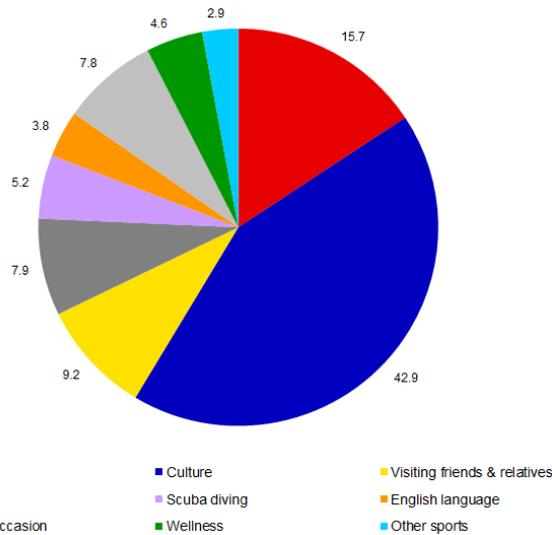
Figure 4: Main motivations for choosing Malta (percentage share of total). Source: MTA's Market Profile Survey, 2017 (Malta Tourism Authority, 2018) and NSO (Inbound Tourism, 2017).
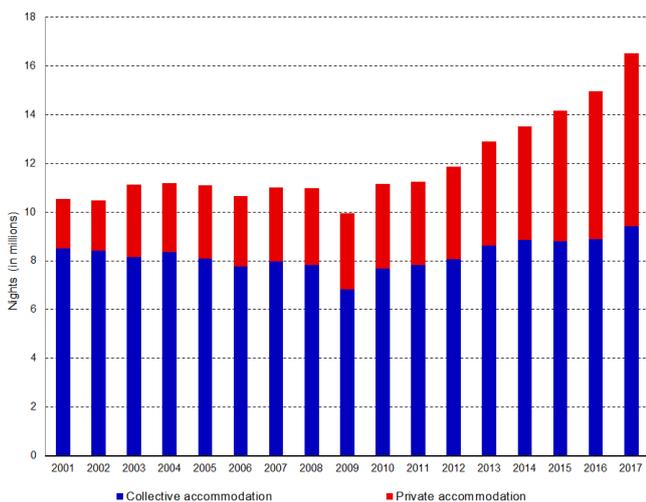

Figure 5: Nights stayed in collective and private accommodation. Source: NSO.

Malta continued to increase steadily and exceeded 16.5 million nights in 2017, almost 50% more than in 2010.

For several years, the vast majority of tourists preferred to spend their stays abroad in collective accommodation establishments.[1] However, since the turn of the century there has been a pronounced shift in preferences from collective accommodation towards stays in private accommodation establishments.[2,3] This change in preference is mainly driven by technological advances and the ease in booking alternative accommodation on online platforms. The switch towards private accommodation also reflects increasing attention to the concept of value for money and a search for more experiential tourism, where visitors seek authentic local experiences.

In 2001, out of the 10.5 million nights spent by tourists in Malta, nights in private accommodation amounted to only about 2 million nights, or slightly less than 20% of total nights. Over the following eight years this share continued to trend upwards, and in 2009 it surpassed the 30% mark, as the negative impact of the global financial crisis was almost entirely absorbed by collective accommodation establishments. By 2017 the nights spent in collective accommodation establishments reached a high of 9.4 million nights. Nevertheless, its share in total nights spent by tourists in Malta declined to 57.0%, as nights stayed in private accommodation establishments rose at a faster rate.

As the popularity of private accommodation increased, the tourism industry in Malta has gradually also shifted from package to non-package holiday makers, with the traditional tour operator business losing market. The share of spending on package holidays almost halved from 2005 to 2017, down from 47.8% to 26.8% of total expenditure. At the same time, the share of spending on accommodation rose by ten percentage points to stand at 17.7% of total expenditure by 2017 (see Fig. 6).

2.3 Average Length of Stay and Tourism Expenditure

During the same period, the average length of nights spent in Malta, but also worldwide, have generally declined (see Fig. 7). In 2001, on average, inbound tourists' spent 9.2 nights in Malta, rising to 10.2 nights by 2003.[4] However, since 2004, the average length of stay has been on a downward trend, standing at 7.3 nights in 2017. The drop in the average length of stay was more pronounced post-2007, partly reflecting the advent of low cost carriers which led to more frequent airline connectivity of the islands, increasing Malta's viability as a short holiday destination.

---

[1] As per NSO definition, these include hotels, guesthouses, hostels, tourist villages, holiday complexes, bed & breakfast, and campsites.

[2] The changing trend in tourists' preference to stay in private accommodation is a global phenomenon and not exclusive to Malta.

[3] Private accommodation includes rented accommodation, own private residence, staying with friends and other private accommodations.

[4] In the 1980s and 90s Malta was a favoured long winter stay for the 'senior citizen market' segment, particularly the British, who used to spend the whole winter in Malta. Thus, prior to the 2000s average length of stay tended to be higher.

Figure 6: Tourist spending. Source: NSO.
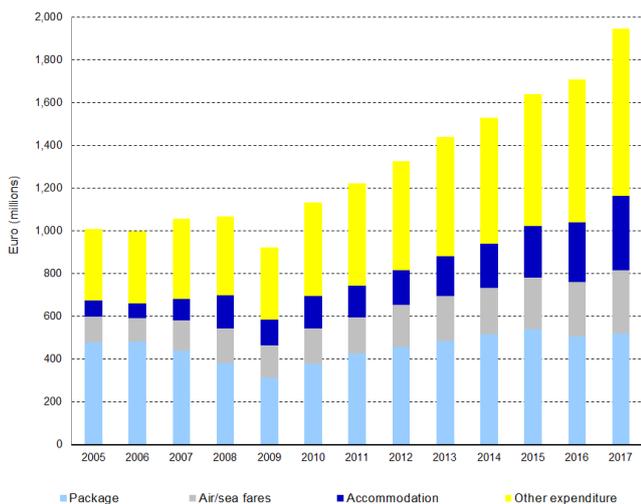
Figure 8: Expenditure per capita and per night. Source: NSO.

Figure 7: Average length of stay. Source: NSO.

The relative stability in the average length of stay and spending per night masks notable heterogeneity across different source markets. Fig. 9 and 10 compare the various markets in 2001 and 2017, respectively. In both years, Libyan tourists recorded the largest expenditure per night spent in Malta, but on average also had the shortest stay. In contrast, in 2001, the average British visitor had the lowest expenditure per night amongst the different source markets (see Fig. 9). At the same time, together with visitors from the USA and Russia, British tourists spent the longest stays in Malta. Meanwhile, German tourists spent only marginally more than British visitors and stayed for a shorter period of time.

The situation changed considerably by 2017 - the average length of stay fell to 7.3 nights while the average tourist expenditure per night edged up to €118 (see Fig. 10). While visitors from Libya continued to lead in terms of expenditure per night, tourists from the USA, Switzerland, Austria and Scandinavia also spent substantially more than the average spend. At €104 per night, Italian tourists were the lowest spenders, followed by the French and British tourists. The practically unchanged spend per night by the Italian tourists between 2001 and 2017 can be partly attributed to increased connectivity and lower airfares between the two countries, which has changed the profile of the average Italian visitor. Meanwhile, only Russian tourists continued to prefer spending around 10 days in Malta. By contrast, recent years have seen the average Scandinavian and Swiss tourist lowering their average length of stay but increasing their spending considerably.

Reflecting the rise in inbound tourists and total nights stayed in Malta, visitors' spending also surged markedly. In 2001, total tourist expenditure stood at €960.4 million. It surpassed the €1 billion mark in 2003. Spending remained broadly stable until 2008 before declining by 13.5% in 2009. From 2010 onwards, expenditure grew at an average annual rate of 9.9%, reaching €1.9 billion in 2017, more than double the level recorded in 2009 (see Fig. 8).

Given the decline in the average length of stay, the average expenditure per visiting tourist did not change much from 2001 onwards. On a per capita expenditure basis, tourists spent €838 per stay in 2001, rising only to €856 in 2017 (see Fig. 8). However, when taking into account the impact of shorter stays, the average expenditure per night increased on a nominal basis from €91 in 2001 to €118 in 2017.

**Figure 9:** Expense per night and length of stay (2001). Source: NSO.
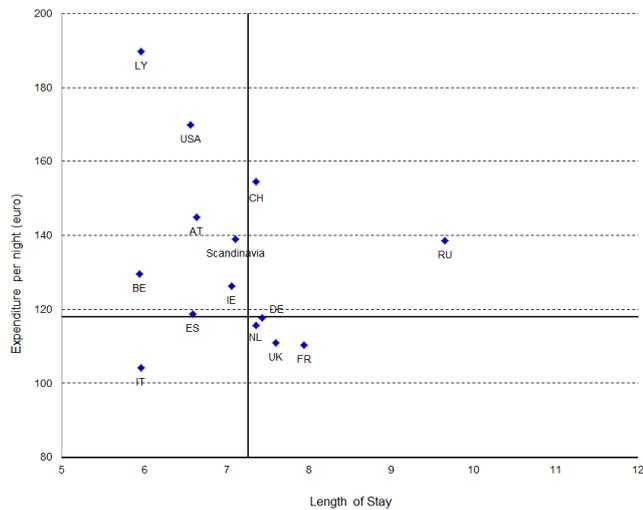
Figure 10: Expense per night and length of stay (2017). Source: NSO.

3 Adapting Supply-Side Factors

Over the years, tourism-related supply factors have adjusted to more dynamic demand. In this study, we limit our consideration of supply to capacity in collective accommodation establishments, aircraft movements, seat capacity and seat load factor, as well as cruise liner calls. A more comprehensive analysis would also warrant an assessment on infrastructural needs and utilisation, and the increasingly importance of stays in private accommodation. However, data on the latter indicators are unavailable or fragmented at best.

3.1 Availability and Occupancy Rates in Collective Accommodation Establishments

Possibly reflecting the requirement of relatively large tracts of land, which is a scarce resource, for the ho-

tels sector, the number of collective accommodation establishments rose at a more muted pace than inbound tourism. Over the last decade, the number of collective establishments on the island rose from 161 in the peak month of August in 2008 to 188 in the same month of 2017. The majority of the new establishments were three-star hotels, followed by two-star establishments. Over this period, the number of five-star hotels remained unchanged at fifteen and only three new four-star establishments were opened.

Since 2008, bed capacity in collective accommodation establishments rose by 8.9%, or close to 3,500 more bed-places (see Fig. 11). Most of the increase emanated from the three and four-star segments. At the same time, capacity in the 'other collective accommodation' establishments declined notably. This drop can be partly attributed to the increased preference towards transforming guesthouses and hostels into boutique hotels which tend to have a smaller number of rooms but offer higher quality standards and facilities.
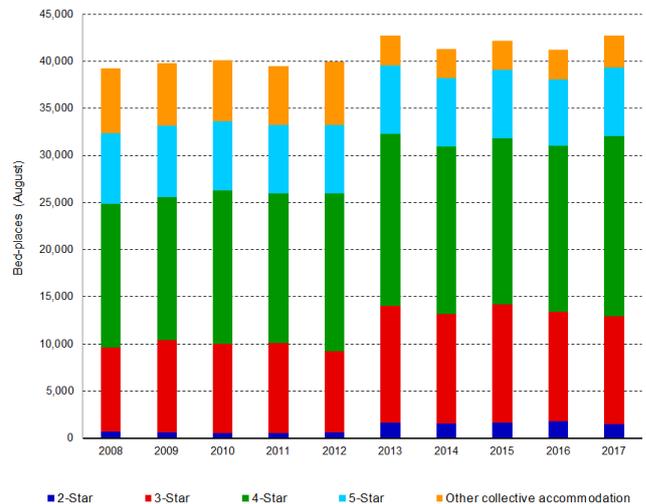
Figure 11: Bed-places by type of accommodation. Source: NSO.

Despite losing market share to private accommodation establishments and increasing the number of bed-places, hotels still managed to increase their occupancy rates. Indeed, apart from a dip in 2009, where occupancy rates bottomed out at 50.7%, occupancy rates in collective accommodation establishments increased steadily, reaching 65.6% in 2017 (see Fig. 12). This rise was a result of higher occupancy rates across all hotel categories. Nonetheless, the most notable increases were recorded in the five and three-star sectors, where occupancy rates rose by 12.5 and 12.6 percentage points respectively since 2008. Throughout the period under review, four-star hotels continued to enjoy the highest average annual occupancy rates. Occupancy rates in this category edged up to 71.4% in 2017, from 67.5% in 2008.
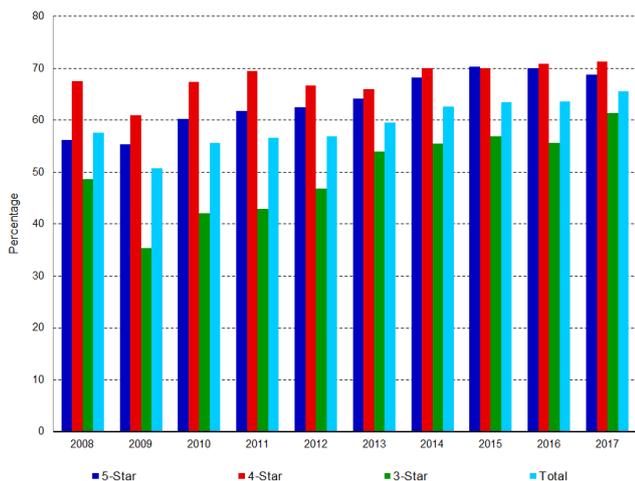
Figure 12: Collective accommodation – occupancy rates. Source: NSO.
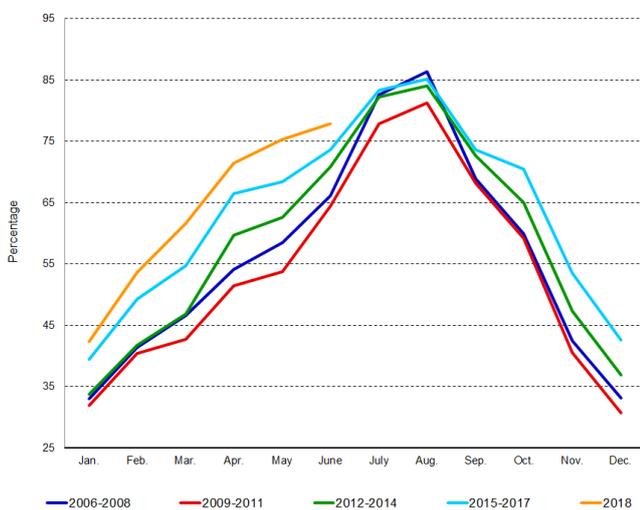
Figure 13: Seasonality in occupancy rates. Source: NSO.

Figure 14: Average daily rates (average based on rates charged between January and June of each year). Source: BOV-Deloitte MHRA Reports.

Higher occupancy rates can be largely attributed to a concerted effort, by various stakeholders, to attract visitors during the winter and the shoulder months of the peak season. In fact, the occupancy rates of total collective accommodations during their peak experienced little change. As can be seen in Fig. 13, apart from the period characterized by the global financial crisis, the occupancy rates in July and August remained broadly stable around the 85% mark when taken as three-year averages. However, occupancy rates have clearly risen in the shoulder months, and the increase is more pronounced in recent years. Data for the first half of 2018 show that occupancy rates averaged 63.7%, up from 50.0% in the 2006 to 2008 period.

Interestingly, higher occupancy rates were not achieved at the expense of lower room rates but through non-cost factors and an improved quality product. As can be seen in Fig. 14, hoteliers were able to increase daily room rates markedly from 2006 onwards.[5] Average rates in the first half of every year suggest that between 2006 and 2018 five-star hotels increased their rates by around 80%, whereas the four-star establishments doubled the rates, and three-star hotels charged 115% more. However, over time, while wage bill and operational costs rose, hoteliers managed to increase their gross operating profit margin per available room. The increased revenue per room was also positively impacted by lower utility tariffs and an influx of foreign workers who have muted increases in wages.

3.2 Airline Connectivity and Seat Capacity

The island's tourism sector is very dependent on the operations of the country's sole airport. Route connectivity and increased aircraft movements are major supply-side factors in the country's tourism market. Over the last decade route connectivity increased from 61 in 2009 to 90 in 2017, and MIA claims that in summer 2018 over 100 destinations were being served (see Fig. 15). Consequently, aircraft movements and the resulting seat capacity increased markedly.

After peaking at close to 28,400 in the year 2000, aircraft movements dropped to almost 24,300 in 2002, and remained close to 25,000 until 2006. The advent of low cost carriers in 2006 led to a rise in the number of aircraft movements but the high point reached in 2000 was only surpassed in 2010 (see Fig. 16). In the following years, aircraft movements continued to grow steadily, reaching close to 43,000 scheduled and chartered flights

---

[5]BOV-Deloitte MHRA Reports.
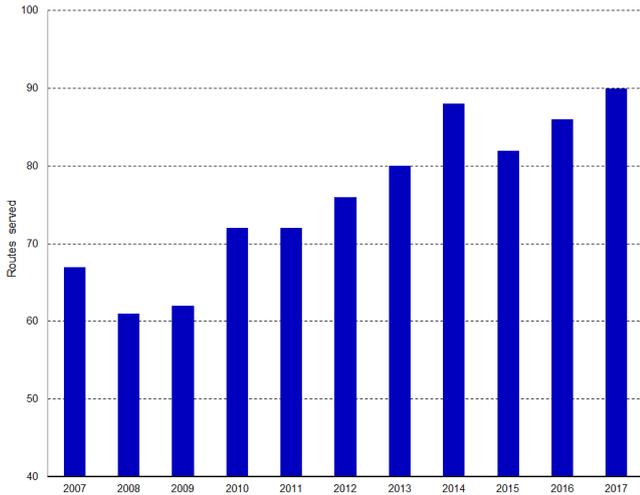
Figure 15: Routes Served. Source: MIA.
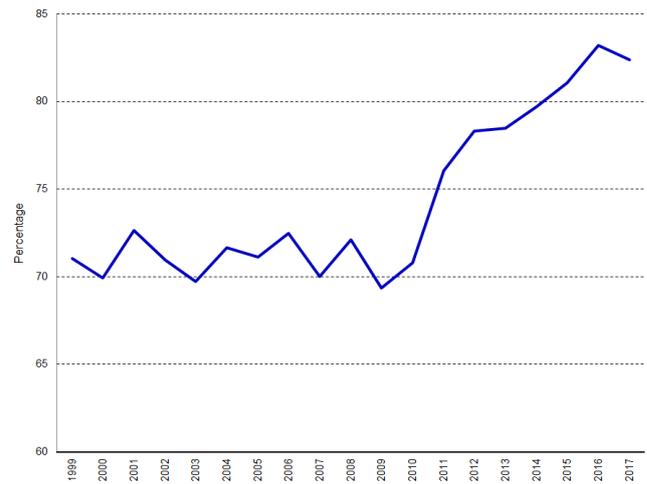
Figure 17: Seat load factor. Source: Authors' own calculations based on MIA releases.
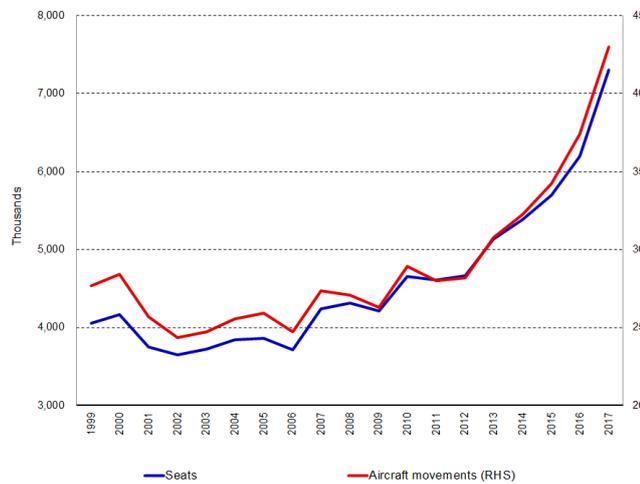
Figure 16: Aircraft movements and available seats. Source: MIA.
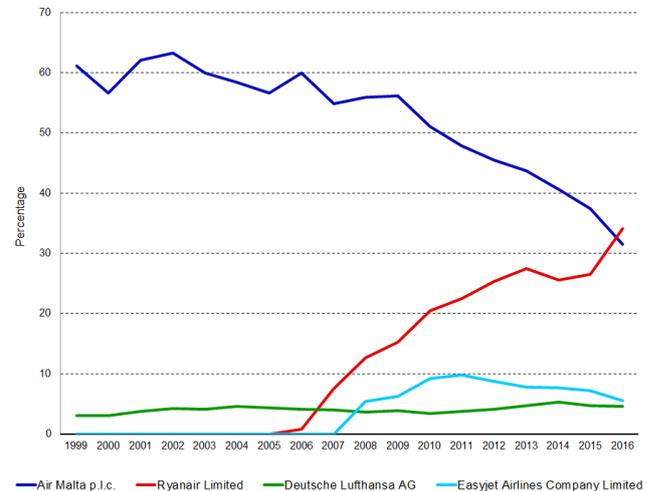
Figure 18: Share of passenger movements by major airlines. Source: MIA.

in 2017. In line with the rise in aircraft movements, the number of seats available to passengers surged to 7.3 million in 2017.

The increase in seat capacity was also met by a rise in passenger, or seat load factor, which measures the capacity utilisation of aircraft. Fig. 17 shows that from 1999 to 2010 the seat load factor of aircraft to and from Malta hovered around 71%; meaning that on average, close to 30% of the aircraft seats were not being sold. Over the succeeding years, the seat load factor climbed gradually to reach 83.2% in 2016, before receding marginally in 2017.

The rise in the seat load factor of aircraft can be partly attributable to the emergence of low cost carriers which are more flexible in altering routes to more profitable destinations. Moreover, as opposed to legacy airlines,

low cost carriers tend to have less, or no, first class seating – which are more difficult to sell. Data show that improvements in the seat load factor coincided with the increased intensity of Ryanair. As can be seen in Fig. 18, in 2006, the year during which Ryanair commenced its operations in Malta, Air Malta had a market share of 59.9% of all passengers travelling through the MIA. Over time, mainly through increases in the number of routes by Ryanair, but also due to financial challenges faced by the national airline, the share of Air Malta's total passenger movements dipped to 31.5% in 2016. During the same year, Ryanair's market share rose to 34.1%, becoming, for the first time, the largest carrier of passengers travelling to and from Malta.

3.3 Activity in Cruise Liner Tourism

Malta's unique geographical location has for long attracted the interest of major cruise liner companies operating in the Mediterranean Sea. The islands' central position makes it possible to cater for both West and East bound cruises. In more recent years, improvements in port infrastructure and ancillary services, combined with challenging economic and political conditions in neighbouring competing destinations, have increased the country's attractiveness.

Between 2008 and 2014 the number of cruise liner passengers arriving in Malta averaged slightly less than half a million passengers per year (see Fig. 19). From 2015 onwards cruise liner passengers rose markedly, reaching 670,000 in 2017. Since 2008, EU nationals accounted for around three-quarters of all arrivals, with the share dipping marginally in the last few years. In 2017, Germany and the UK were the most important source markets, accounting for half of all EU arrivals, whereas the US was the largest source of passengers from outside the EU.

Interestingly, the increase in passenger arrivals was not matched by a similar rise in cruise liner calls (see Fig. 18). In 2017, 342 cruise liner calls were made, significantly more than in the previous years, but still below the 397 calls made in 2008. This shows that larger vessels, with increased accommodation capacity are calling into the Maltese ports.
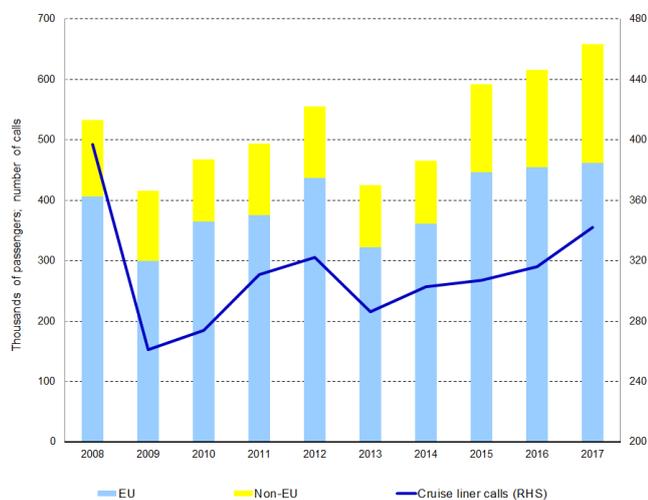
Figure 19: Cruise liner passengers and calls. Source: NSO.

4 Economic Importance of Tourism

This section will shed some light on a number of comparative indicators with neighbouring destinations, as well as comparing other previously conducted studies on the economic importance of tourism for the Malta. This study is not intended to comprehensively assess all the economic factors which constitute this sector.

Data from the World Travel and Tourism Council (WTTC) show that the travel and tourism industry's *total* contribution to Malta's GDP stood at 27.1% in 2017 (see Fig. 20).[6] This was the highest share recorded within the Mediterranean region by a notable margin. The second highest share pertained to Cyprus, at 22.3% of GDP, followed by Greece and Lebanon. Malta's share was also well above the Mediterranean, EU and World averages, which ranged between the 10% and 12% (World Travel and Tourism Council, 2018a, 2018b).

The contribution of the travel and tourism sector towards employment is correspondingly also very high in Malta when compared with other tourism destinations within the region. In 2016, it stood at 28.3%, once again higher than other countries surrounding the Mediterranean, as well as the Mediterranean, EU and World averages (see Fig. 20). When compared with other countries in the region, the second highest total share was recorded by Greece at 24.8%, followed by Cyprus. Such a high contribution of tourism towards employment serves to once again highlight the importance of tourism, both in the Maltese economic and social contexts.
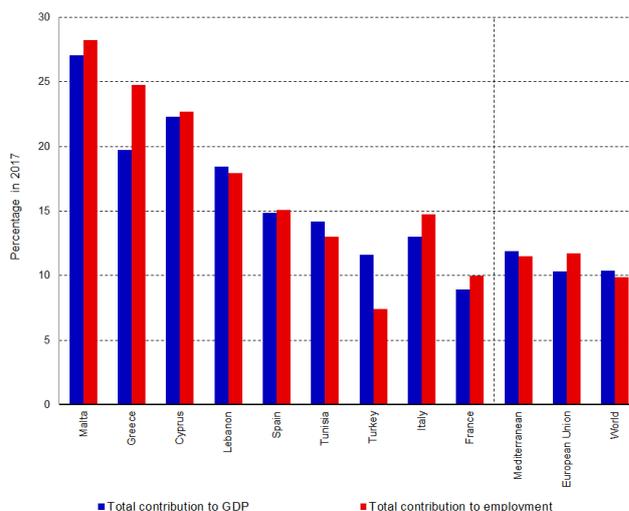
Figure 20: Contribution of travel and tourism to GDP and employment. Source: WTTC.

The WTTC estimated that the direct contribution of travel and tourism to GDP in Malta during 2017 was equal to 14.2%. This primarily reflects the economic activity generated by industries such as hotels, travel agents, airlines and other passenger transportation services (excluding commuter services), but it also includes,

---

[6]Total contributions to GDP and employment include all activity and jobs generated directly, indirectly, as well as the induced impacts of the travel and tourism sectors.

Table 1: Input-Output based estimates of the contribution of tourism to the Maltese Economy. Sources: Blake, Sinclair, Sugiyarto and DeHann (2003), Cassar (2015), National Statistics Office of Malta (NSO) (2016).

	Blake, Sinclair, Sugiyarto and DeHann (2003) (% of GNP for 2001)	Cassar (2015) (% of total GVA for 2008)	National Statistics Office of Malta (NSO) (2016) (% of total GVA for 2010)
Direct and indirect (Based on Type 1 multipliers)	12.3	12.5	11.5
Direct, indirect and induced (Based on Type 2 multipliers)	17.1	17.7	N/A

for example, the activities of the restaurant and leisure industries directly supported by tourists. Estimates of the same Council show that in 2017 the travel and tourism sectors generated 31,000 jobs directly in Malta (15.7% of total employment). This includes employment by hotels, travel agents, airlines and other passenger transportation services, as well as jobs in restaurant and leisure industries directly supported by tourists. The total contribution of these sectors to employment including indirect and induced impacts was estimated to reach 55,000 jobs in 2017 (28.3% of total employment).

The WTTC estimates are the most recently available impacts of the contribution of the tourism industry on the Maltese economy and are one of the most widely referenced studies. Nonetheless, the input-output methodology used to estimate travel and tourism contributions to GDP and employment by the WTTC are based on assumptions which contrast with other studies conducted by other researchers and academics. Furthermore, it should be noted that the WTTC study did not utilize a fully specified input-output table for Malta for the derivation of the multiplier estimates, but applied technical coefficients obtained from the input-output tables of other countries where data was incomplete. It is important to point out that there is no indication of the country source of the input-output matrices which were applied to Malta.

In the research by Cassar, Vella and Buttigieg (2016) it is argued that the estimates of WTTC are relatively higher than what has been estimated separately by Blake, Sinclair, Sugiyarto and DeHann (2003) for 2001, Cassar (2015) for 2008 and National Statistics Office of Malta (NSO) (2016) for 2010 (see Table 1).

Table 1, sourced from Cassar et al. (2016), shows that, notwithstanding the divergences in the three methodologies and time elapsed between 2001 and 2010, the results obtained from the three input-output based studies all seem to corroborate a direct and indirect contribution to the Maltese economy equal to approximately 12% of total gross value added (GVA).

Furthermore, observing the Type II value added multipliers, which in the case of Cassar (2015) are based on the computed Type II weighted average tourism value added multiplier, one can note further similarities in results between the two studies with a contribution of tourism at around 17% of total GVA. Within the context of assessing the true contribution of tourism to the Maltese economy on the basis of the input-output studies discussed, would imply that a realistic contribution of the tourism sector to the Maltese economy would lie in the range of 12% to 17% of total GVA.

5   Sustainable Tourism and Policy Considerations

In devising tourism policies, policymakers need to take into consideration the impact on various stakeholders, including the economy, the environment, entrepreneurs, visitors and residents. Nonetheless, striking this fine balance is by no means an easy feat. This is particularly challenging in small countries like Malta, which saw 5.3 visitors per inhabitant in 2017; the second highest ratio worldwide after Iceland. When one takes into account tourists' concentration on Malta's limited geographical size and high population density, the situation is significantly more severe. As one of the main economic pillars, Malta's tourism sector creates jobs, drives exports, and generates prosperity. However, more needs to be done to ensure that recent growth rates in inbound tourism are sustainable.

The United Nations has designated 2017 as the International Year of Sustainable Tourism for Development. The recent growth in tourism requires the sector to ensure sustainable policies and practices that minimise adverse effects of tourism on the use of natural resources, infrastructure, mobility and congestion, as well as its socio-cultural impact (United Nations World Tourism Organisation, Centre of Expertise Leisure, Tourism & Hospitality, NHTV Breda University of Applied Sciences & NHL Stenden University of Applied Sciences, 2018). Increased reports of negative attitudes among local populations towards visitors, due to perceived overcrowding, noise and other issues, have led to

the spreading of terms such as 'overtourism' and 'tourismphobia' in the media. The first, and most prominent, cases of 'tourismphobia' occurred in Barcelona and Venice, where tensions have been rising for years over the unregulated surge in visitors and the negative impact on the local housing market. In these cities residents have voiced their anger over rising rents and pollution caused by huge cruise ships.

Looking closer at our shores, despite the general welcoming attitude of Maltese nationals towards tourists, the local tourism product faces the challenge of over-utilised infrastructure and the impact on the environment of construction and overcrowding in beaches, historic and cultural sites.

The success of the tourism industry cannot be merely measured by the number of inbound visitors but by a better economic and social return to society. As evidenced by official data, the increase in the number of inbound tourists was not matched by significant increases in nominal expenditure per night stayed. Furthermore, if one had to look at deflated figures, real expenditure per visitor per night is estimated to have declined in recent years, even after allowing for potentially lower airfares. This may suggest that the country's stakeholders' efforts were more successful in attracting higher volumes but lower spending visitors. It could also reflect the scarcity of large scale hotels and capacity constraints in established ones, which may have conditioned tourists to look for private accommodation, which tends to be cheaper.

The tourism product is composite and complex in nature. It encompasses anything that a tourist purchases, sees, experiences and feels from the moment they leave home to the time they return. It is in the interest of the operators within the tourism industry to sustain the destination. While many large operators are adopting better practices which will also bring environmental benefits, these practices are not generally being adopted by the smaller operators. Identifying and improving different aspects of the tourism product is key, but more importantly a holistic approach needs to be adopted, where the whole is greater than the sum of its parts. For example, the excellent provision of one sub-product (such as hotels) may not be worthwhile if the other inter-related services or sub-products (such as physical infrastructure) are not up to the expected standard.

Over the past few decades, a number of tourism master plans and policies have been successfully implemented. However, synergies between key stakeholders - the Government, Malta Tourism Authority, Malta Hotels and Restaurants Association (MHRA) and Malta International Airport, amongst others - were instrumental in reaping the positive economic benefits of the tourism industry. Efforts to increase route connectivity, thereby reducing dependence on few source markets, and creating important niche markets to diversify the Maltese tourism product have helped in significantly tackling the issue of seasonality. Improvements on roads infrastructure and embellishment projects in a number of key locations are also commendable. However, additional effort is required to enhance the quality of the tourism product and improve public transport services. Moreover, hoteliers and entrepreneurs should continue to invest in high-end establishments that can help in attracting more affluent visitors. This would enable better wages to employees in the local industry, which would render it more attractive as a vocational career while ensuring that a larger share of the population would benefit from the buoyancy of the sector. Finally, the education system should also do its part to instil a nurturing mind-set towards the local cultural heritage.

Acknowledgements

References

Blake, A., Sinclair, T. M., Sugiyarto, G. & DeHann, C. (2003). *The Economic Impact of Tourism in Malta: Input-Output Modelling*. The University of Nottingham, Report for the Malta Tourism Authority.

BOV-Deloitte MHRA Reports 2006–2018.

Briguglio, L. (1992). Tourism Multipliers in the Maltese Economy. In P. Johnson & B. Thomas (Eds.), *Tourism Perspectives in Tourism Policy*. U.K.: Mansell Publishing.

Briguglio, L., Archer, B., Jafari, J. & Wall, G. (Eds.). (1996). *Sustainable Tourism in Islands and Small States: Issues and Policies*. Island Studies Series. London: Pinter Publishing.

Butler, R. W. (1980). The concept of a tourist area cycle of evolution: implications for management of resources. *Can. Geogr. 24*, 5–12.

Camilleri, M. A. (2018). *Travel Marketing, Tourism Economics and the Airline Product – An introduction to theory and Practice*. Tourism, Hospitality & Event Management. Springer Publishers.

Cassar, I. P. (2015). *Estimates of Output, Income, Value Added and Employment Multipliers for the Maltese*

Economy (WP-03-2015), Central Bank of Malta. Valletta.

Cassar, I. P., Vella, K. & Buttigieg, S. (2016). Understanding the Economic Contribution of Tourism in Malta: A Literature Review. *Mediterr. J. Soc. Sci. 7*(6), 49–60.

Dwyer, L., Forsyth, P., Spurr, R. & Vanho, T. (2003). Tourism's Contribution to a State Economy: A Multi-Regional General Equilibrium Analysis. *Tourism Econ. 9*(4), 431–448.

Malta International Airport (MIA). (2019). Statistics. Retrieved September 20, 2018, from https://www.maltairport.com/corporate/traffic-development/statistics/

Malta Tourism Authority. (2018). *Tourism in Malta – Facts and Figures 2017*. Kalkara: Malta Tourism Authority.

Mifsud, M. (2006). *Historical Analysis of Tourism Policy and Planning in Malta* (Master's thesis, University of Nottingham, Nottingham, U.K.).

National Statistics Office of Malta (NSO). (2016). *Supply, Use and Input-Output Tables 2010*. Valletta: National Statistics Office.

National Statistics Office of Malta (NSO). (2017). *Tourism Satellite Accounts 2010*. Valletta: National Statistics Office.

National Statistics Office Stat DB database.

Pratt, S. (2015). The economic impact of tourism in SIDS. *Ann. Tour. Res. 52*(May), 148–160.

United Nations World Tourism Organisation, Centre of Expertise Leisure, Tourism & Hospitality, NHTV Breda University of Applied Sciences & NHL Stenden University of Applied Sciences. (2018). *'Overtourism'? – Understanding and Managing Urban Tourism Growth beyond Perceptions*. Madrid: UNWTO.

World Travel and Tourism Council. (2018a). *Travel & Tourism Economic Impact 2018 Malta*. London: WTTC.

World Travel and Tourism Council. (2018b). *Travel & Tourism Economic Impact 2018 World*. London: WTTC.

Review Article

Iron and its alloys for Bone Regeneration Scaffolds – A Review

Christabelle Tonna[*1] **and Luke Saliba**[2]

[1]*Department of Metallurgy and Materials Engineering, University of Malta, Msida, Malta*
[2]*Trauma and Orthopaedic Surgery Department, Mater Dei Hospital, Msida, Malta*

Abstract. Permanent implants and bone grafts have been used successfully to repair bone defects for a number of years. However, there are significant limitations, for example patients requiring revision surgery for implant removal, inadequate mechanical properties leading to stress-shielding and osteoporosis, as well as restricted bone development, particularly in paediatric patients. As a result, those implants with a more active involvement in the healing process than the original inert implants, were favoured. Biodegradable scaffolds are porous implants which are incorporated into sizeable bone defects in order to support the damaged area while the bone regenerates. In response to bone healing, the structure is expected to degrade at a controlled rate *in vivo*. Following the promising research published in relation to magnesium-based alloys for cardiovascular stents, iron and its alloys have recently been proposed for this application. *In vivo* evidence show that pure iron exhibited an inadequately slow degradation rate. Therefore, research efforts have been focused on accelerating the corrosion rate by implementing various material design strategies. This review presents an overview of notable research work treating the tailoring of corrosion, mechanical and cytotoxic response as well as promising processing methods for the production of iron-based foam structures. To conclude, based on current research, the clinical potential for these materials will be analysed.

Keywords: iron alloys, bone scaffolds, biodegradable implants, biomaterials, biometals

1 Introduction

Statistics released in the year 2016 showed that, of the 7.9 million fractures that are reported annually, 5–10% result in delayed or impaired healing, of which a significant amount are eventually classified as non-union fractures (Buza & Einhorn, 2016). To encourage bone healing in such cases, surgeons may opt to use bone scaffolds.

In the pursuit of the ideal scaffold, researchers have established a set of optimal characteristics for the final product. Firstly, the implant's porous structure must allow for cell proliferation and vascularisation (the mechanisms that govern the reconstruction process). Moreover, it must be able to mechanically-support the surrounding tissue during reconstruction (Bose, Roy & Bandyopadhyay, 2012). Over the past decades, research into metallic biomaterials have focused on the study of relatively inert materials, including austenitic stainless steel, titanium alloys and cobalt-chrome alloys (Eliaz, 2019). However, the ideal scaffold necessitates the use of a material that has the ability to degrade at a controlled rate in order to match the rate of bone regeneration, while releasing by-products that do not interfere with regular metabolic activity. The use of such materials would eliminate the need of a second surgery to remove the implant after it serves its temporary function, while also contributing to healthcare institutions by lessening the financial burden associated with revision surgeries (Bose et al., 2012; Heiden, Kustas et al., 2015).

The most common bone grafting products available on the market are either made of allograft, that is, human bone taken from cadavers, including demineralised bone matrix (DBM), or ceramic products, like β-tricalcium phosphate (American Association of Surgeons, 2010). However, significant research has also been carried out on biodegradable polymeric materials

like polylactic acid (PLA), polyglycholic acid (PGA), polylactic co-glycholic acid (PLGA) and polycaprolactone (PCL) (Campana et al., 2014, 10). Despite the fact that results are often positive when using these polymeric and ceramic materials, they generally exhibit inadequate strength, toughness or degradation rates (Heiden, Walker & Stanciu, 2015).

The focus has therefore shifted onto biodegradable metals, which have the potential to offer a more satisfactory mechanical performance, especially for orthopaedic applications. Magnesium and its alloys have been by far the most studied metals. However, their lack of resistance to high amounts of chloride ions results in excessively rapid degradation rates. Apart from this, they produce large quantities of hydrogen gas during degradation, which could result in tissue necrosis (Heiden, Walker & Stanciu, 2015). A more recent development has been centred around the study of zinc alloys. When compared to magnesium alloys, these materials have consistently shown a more adequate corrosion rate *in vitro*, while their low melting point greatly facilitates their processing (Katarivas Levy, Goldman & Aghion, 2017). The main limitation of zinc has been its generally poor mechanical strength, however recent studies have shown that alloying with elements like magnesium and strontium could lead to significant improvements in mechanical strength, coupled with promising biocompatibility in various potential implantation sites, as analysed through acute *in vivo* studies in connective, bone and vascular tissues (Zhu et al., 2019). Despite the positive results, prolonged *in vivo* evaluations are required in order to further evaluate the potential of these alloys.

Over the past couple of decades, iron alloys started being considered for biodegradable orthopaedic applications, as this element is also an essential trace element within the body, although the daily exposure limit is set much lower than that of magnesium (Yuen & Ip, 2010). However, as opposed to the latter, iron does not generate harmful gas during the degradation process, and its mechanical properties are more tailored for load-bearing applications (Cheng, Liu, Wu & Zheng, 2013; Heiden, Walker & Stanciu, 2015). The largest issue with respect to these materials was presented to the scientific community in 2001, when Peuster et al. (2001) indicated that pure Fe stents implanted in New Zealand white rabbits underwent very little degradation after one year, highlighting their slow degradation rate. Consequently, studies have since focused on the acceleration of the degradation rate, while simultaneously treating other issues like cytotoxicity, ferromagnetism, and finding appropriate ways to fabricate implants with the qualities mentioned previously.

2   Degradation of Pure Fe

For researchers in the field to be able to accelerate the corrosion rate of pure Fe, it is essential to first understand the degradation process of pure Fe. This has been discussed in a number of publications (Gorejová, Haverová, Oriňaková, Oriňak & Oriňak, 2019; Hermawan, Purnama, Dube, Couet & Mantovani, 2010; Zheng, Gu & Witte, 2014). The process could essentially be split into four sections, as schematically represented in Fig. 1 (Zheng et al., 2014). In the initial step, the electrochemical processes described in Eqs. (1) and (2), take place at anodic and cathodic areas all over the specimen surface. In pure Fe, a potential difference generally exists between the grains and grain boundaries, as displayed in Fig. 1(a).

$$\text{Fe} \longrightarrow \text{Fe}^{2+} + 2\,\text{e}^- \tag{1}$$

$$\text{O}_2 + 2\,\text{H}_2\text{O} + 4\,\text{e}^- \longrightarrow 4\,\text{OH}^- \tag{2}$$
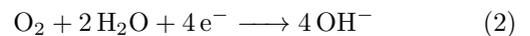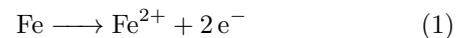
In the second stage, the Fe ions (represented as M^{n+}) react with the OH^- ions to form iron (II) hydroxide and consequently, ferric hydroxide, or hydrated iron oxides, as in Eqs. (3) and (4). The products could react to form a magnetite layer in contact with the specimen surface, and haematite on top, as is often reported (Hermawan, Purnama et al., 2010; Gorejová et al., 2019).

$$2\,\text{Fe}^{2+} + 4\,\text{OH}^- \longrightarrow 2\,\text{Fe(OH)}_2 \tag{3}$$

$$4\,\text{Fe(OH)}_2 + \text{O}_2 + 2\,\text{H}_2\text{O} \longrightarrow 4\,\text{Fe(OH)}_3 \tag{4}$$
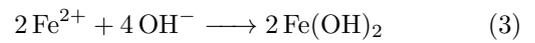
As corrosion progresses, weak spots in the hydroxide layer result in the formation of pits on the surface. Such regions are acidified through metal cation hydrolysis, as exemplified in Eq. (5), while gradually being depleted of oxygen, thus limiting the cathodic reaction (Cramer & Covino Jr, 2003). The combined effect of these aspects results in an increasingly aggressive environment that leads to further acceleration of the corrosion process.

$$\text{FeCl}_2 + \text{H}_2\text{O} \longrightarrow \text{FeOH}^+ + 2\,\text{Cl}^- + \text{H}^+ \tag{5}$$

Fig. 1(c) also shows cell attachment as the surface roughens, as well as the formation of calcium phosphate (or hydroxyapatite-like) compounds, fuelled by the rise in pH as a result of the cathodic reaction. This is particularly positive for bone scaffold applications, as the compounds formed are very similar to the main component of bone itself, encouraging strong attachment. Zhang, Chen and Shen (2010) also describe other reactions that may take place at the metal surface; however, these are not central to the aim of this review.

As suggested in Fig. 1(d), chunks of various sizes from the metal may break off after significant corrosion has taken place, contributing to further material loss.
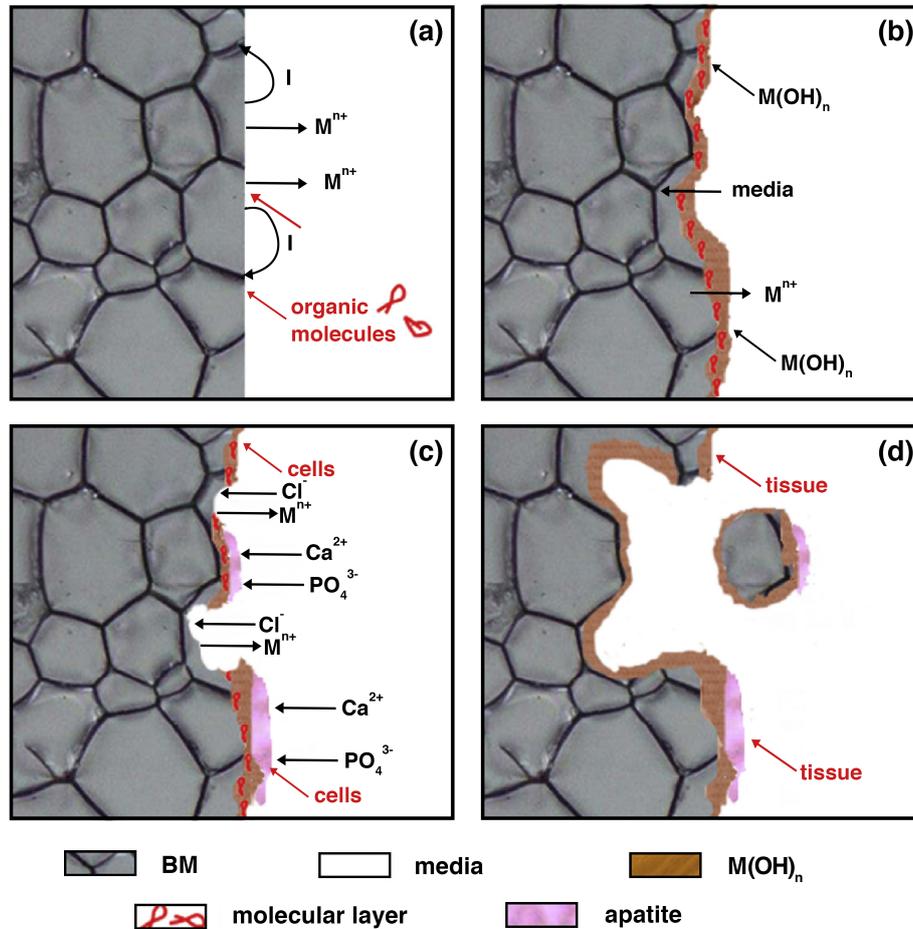
Figure 1: A schematic representation of the degradation process of pure Fe. Reprinted with permission from Elsevier from Zheng, Gu and Witte (2014).

While there have been studies that confirm the described mode of degradation when testing both pure Fe and Fe-alloys *in vitro* (Huang & Zheng, 2016; Huang, Zheng & Han, 2016), both Zhang et al. (2010) and Zhu et al. (2009) have suggested that the dominant degradation mode in pure Fe is uniform corrosion, following analysis of tested samples using Scanning Electron Microscopy (SEM). In fact, Zhu et al. (2009) speculated that a typical vascular stent following the same mean corrosion rate would be completely degraded in 1 month. However, local biological changes and corrosion kinetics *in vivo* are likely to lead to considerably different realities, as discussed in Section 5 of this review.

Another concern that arose when testing pure Fe was related to the corrosion product build-up. When testing pure Fe coupons using static immersion in Hank's solution, Zhang et al. (2010) reported a positive and constant corrosion rate until day 21. The authors attributed the decrease in corrosion rate after day 21 to the accumulation of phosphates on the sample surface. However, that same product did not result in surface passivation when testing the same samples using electrochemical means. On the other hand, in a very similar testing setup, Cheng et al. (2013) reported the build-up of Fe_2O_3 after only 3 days of immersion. It must be noted that this, as well as other Fe corrosion products, are insoluble. Therefore, achieving full degradability without leaving any residuals from the implant is particularly difficult if there are no mechanisms of transport out of the body for the products formed (Cheng et al., 2013). Such products are likely to inhibit further degradation and the remaining implant will stay in the body (Gorejová et al., 2019). However, this should not be an issue, provided that healthy bone is left *in situ* and that the remaining foreign material does not interfere with normal biological activity.

As to the passivating layer, some have also commented on its dependence on the microstructure of the sample. In fact, Obayi et al. (2016) concluded that, in accordance with the study by Ralston, Birbilis and Davies (2010), the corrosion rate actually rises with increase in grain size, contrary to popular belief. As a matter of

fact, the pure Fe samples with the smallest grain size, developed a more dense and stable oxide layer, impeding further release of ions, in the presence of an even slightly alkaline environment (Obayi et al., 2016).

Section 3 will provide an overview of notable works dealing with alloying of pure iron for enhanced degradation rates. Variations in similar corrosion testing procedures make it somewhat difficult to quantitatively compare the performance of different materials, however, general observations could be used to provide an indication of the most promising directions for these innovative materials.

3 Alloying for Improved Performance

3.1 Fe-Mn Alloys

In a series of studies starting in 2007, Hermawan, Dubé and Mantovani (2007) introduced the idea of iron-manganese alloys as biodegradable metals. The value of -1.18 volt corresponds to the electrode potential of Mn not to the electrode potential of the metal after the addition of Mn. The addition of Mn could have the overall effect of lowering the potential of the bulk material from an initial value of $-0.44\,\mathrm{V}$, thus increasing its susceptibility to corrosion (Schinhammer, Hänzi, Löffler & Uggowitzer, 2010). Apart from this, manganese could enhance the strength of the material through the formation of an Fe-Mn solid solution, and also allows for the elimination of the ferromagnetic characteristic of the material, through the formation of the antiferromagnetic γ-austenite or ε-martensite Fe-Mn phases. These two phases are the face-centred cubic (FCC) and hexagonal close packed (HCP) variations of the body-centred cubic (BCC) arrangement of pure Fe at room temperature, brought about by the addition of varying amounts of Mn in conjunction with thermal or mechanical processes, and resulting in specific mechanical, magnetic and electrochemical properties. Originally aimed for application as cardiovascular stents, the authors first investigated an alloy which is produced via the powder metallurgy route, wherein metal particles are uniaxially pressed to form powder compacts. These powder compacts are subsequently subjected to high temperatures below the material melting point, in order to stimulate the particles to bind together. This process is more commonly known as conventional sintering. In the case of Hermawan et al. (2007), the coupons were subjected to various cycles of cold-rolling operations to eliminate the resultant porosity in the specimen from the sintering process. The material could then be laser-cut to produce the typical stent geometry. From their initial *in vitro* degradation studies, the Fe-35Mn coupons tested seemed to exhibit a uniform degradation mechanism, with grain boundaries all over the sample surface acting as anodic sites. No pitting was observed after 144

Table 1: Mechanical properties of Fe35Mn and SS316L (Hermawan, Alamdari, Mantovani & Dubé, 2008).

Material	UTS[a] (MPa)	YS[b] (MPa)	e$_{max}$[c] (%)
Fe35Mn	550 ± 8	235 ± 8	31 ± 5
SS316L[d]	580 ± 2	250 ± 2	56 ± 2

[a]UTS: Ultimate Tensile Strength;
[b]YS: Yield Strength;
[c]e$_{max}$: maximum elongation;
[d]SS316L: annealed foil of 0.5 mm thickness.

hours of immersion testing. Moreover, the magnetic susceptibility of the same material was even less than that of AISI 316L stainless steel; a material that has been widely implemented in various biomedical applications. In their second publication on the same material, Hermawan, Alamdari, Mantovani and Dubé (2008) also showed that Fe-35Mn exhibited comparable mechanical properties to AISI 316L, as shown in Table 1, placing the alloy ahead of both pure Fe and most Mg-alloys, in terms of mechanical performance.

Following the positive results obtained using Fe-35Mn, Hermawan, Dube and Mantovani (2010) studied the effect of varying the amount of Mn between 20 and 35wt.%. It must be noted that the harder martensitic phases are more stable at lower concentrations ($< 27\%$) of Mn in Fe (Zhang & Cao, 2015). Apart from this, the same martensitic phases could also be induced through plastic deformation (Sato, Ichinose, Hirotsu & Inoue, 1989). In fact, lower concentrations of Mn (20–25wt.%) led to a bi-phase composition of austenite and ε-martensite, while subsequent plastic deformation resulted in a fully martensitic structure. On the other hand, higher amounts of Mn (30–35wt.%) led to an initially austenitic structure, and a bi-phase composition following plastic deformation. The dual phase compositions enhanced the micro-galvanic degradation effect and resulted in higher corrosion rates, while simultaneously improving the strength, at the expense of ductility. Although the mechanical, corrosion and magnetic performance was improved for all Fe-Mn alloys tested when compared to pure Fe, the authors proposed Fe-30Mn and Fe-35Mn for further investigation (Hermawan, Dube & Mantovani, 2010; Hermawan, Purnama et al., 2010). This research in particular served as the basis for several other studies, as will be noted throughout this review.

The micro-galvanic effect with Fe-Mn alloys was also observed recently by Dehestani, Trumble, Wang, Wang and Stanciu (2017), who studied the effect of varying powder particle size on the degradation rate of Fe-30Mn. Interestingly, the use of a relatively large particle size led to a corrosion rate that was about five times higher than that achieved when using finer particles. The formation

of a homogeneous composition across the sample volume is dependent on the diffusion of Mn into Fe during the heat treatment. The use of larger particles provided diffusion distances that were significantly greater, so as to prevent the complete diffusion of Mn. With respect to the less noble Mn-rich regions, the resultant Fe-rich centre of the particles therefore acted as a cathodic site, accelerating the corrosion rate. Similar outcomes were also observed when studying the microstructure and degradation characteristics of various Fe-Mn alloys (Kupková, Hrubovčáková, Zeleňák et al., 2015; Kupková, Hrubovčáková, Kupka, Oriňáková & Morovska Turonova, 2015). Moreover, when using larger powders, the packing efficiency is naturally poorer, contributing to a higher amount of porosity in the coupon. This translated into a larger surface area exposed to the corrosive medium, which further contributed to an increase in corrosion rate. The latter effect is commonly observed when dealing with powder processed test samples and is a serious consideration when designing the actual bone scaffold (Zhang, Wang, Cao & Gao, 2012; Zhang & Cao, 2015).

Similar to the issues related to corrosion product build-up when testing pure Fe, the agglomeration of oxides on the surface of Fe-Mn alloys could also act as a barrier, preventing further ion exchange between the metallic sample and the corrosive medium. In fact, Hermawan, Purnama et al. (2010) reported the presence of magnetite and other hydrated products on the surface of degradation tested Fe-Mn, which led to this effect on the corrosion rate. Similarly, Heiden, Walker, Nauman and Stanciu (2015) observed no corrosion on Fe-20Mn during the first 50 days of an immersion test in osteogenic cell culture medium, and attributed this lack of weight loss to the formation of the observed oxide layer. However, the corrosion rate drastically increased following the 50 days, as the loosely bound oxide layer fissured and allowed the corrosive fluid to come into contact with the metallic substrate. This effect was not reported elsewhere, however, it is suspected that the residual strains generated during the machining of the coupons for testing, could have stimulated the eventual oxide cracks. In fact, it is generally accepted that the stable, insoluble oxides formed will significantly diminish the degradation rate with time.

Although alloying with Mn had a positive effect on the corrosion rate in the works cited previously, the high Mn compositions used could, if not carefully controlled, result in severe toxicological effects. Therefore, in an effort to reduce the risks of disease, Drynda, Hassel, Bach and Peuster (2015) investigated Fe-(0.5–6.9wt.%)Mn for its cardiovascular application. However, the coupons lost even less weight than pure Fe when immersed for 84 days in 0.9% NaCl. These results were supported by similar poor degradation rates when testing Fe-3Mn, compared to pure Fe (Liu & Zheng, 2011).

3.2 Alloying with Noble Elements

Leading research groups in the field have approached the main issue of pure Fe in either of two primary ways. Corrosion could be accelerated through the addition of a second phase that lowers the overall standard electrode potential of the alloy, and thus makes the material more susceptible to corrosion, which is in essence the target when adding Mn. The alternative is to add a phase that is nobler than pure Fe, in order to trigger micro-galvanic corrosion (Schinhammer et al., 2010). The latter has proven to be the most popular route adopted.

In the first study using noble metals, Schinhammer et al. (2010) obtained promising results when including palladium (Pd), a material commonly used in dental alloys (Wataha & Shor, 2010). The Pd-rich intermetallic phase that precipitated in the Fe-10Mn-Pd tested, not only exhibited a higher corrosion rate when compared to the reference Fe-Mn alloy, but also demonstrated higher compressive strength. The use of just 10wt.% Mn led to the formation of martensite. In their later work, in order to avoid the presence of this brittle ferromagnetic phase, Schinhammer, Steiger, Moszner, Loffler and Uggowitzer (2013) prepared an Fe-21Mn-Pd alloy which was tested using Electrochemical Impedance Spectroscopy (EIS). The alloy was shown to exhibit a polarisation resistance that was five times lower than that of the pure Fe sample, indicating its significantly faster corrosion rate. As opposed to some of the previously mentioned work, the layered corrosion product formed, similar to that previously described by Hermawan, Purnama et al. (2010), did not necessarily result in reduced corrosion rates. In fact, EIS tests showed a decrease in polarisation resistance with time. Čapek, Msallamová, Jablonská, Lipov and Vojtěch (2017) also used Pd as the alloying element in the preparation of Fe-2Pd by Spark Plasma Sintering (SPS). Similar to the previous findings, this alloy displayed approximately four times the compressive yield strength of pure Fe, as well as a high corrosion current density in both static immersion tests and electrochemical corrosion tests.

To prepare Fe-5Pt, Huang, Cheng and Zheng (2014) introduced platinum (Pt) with a standard electrode potential of 1.2 V in pure Fe. They also compared this to the performance of pure Fe and an Fe-5Pd alloy. Since both noble metals have limited solubility in Fe, the documented increase in yield strength was mainly due to the secondary phase strengthening effect. Moreover, the corrosion of both alloys was visibly much more severe than that of pure Fe, with the areas surrounding Pd- and Pt-rich areas corroding most severely. In a similar study, based on the same principal, Huang, Cheng, Bian and Zheng (2016) added silver (Ag) and gold (Au) to Fe.

Interestingly, in the case of Fe-Au alloys, the mechanical performance generally improved as a result of secondary phase strengthening and solid-solution strengthening. However, Fe-10Ag demonstrated poorer mechanical properties compared to pure Fe, indicating that excessive additions of the softer metal into the Fe-matrix is detrimental to the mechanical performance. Again, corrosion results were positive, with Fe-5Ag and Fe-5Au demonstrating the fastest corrosion rates in static immersion tests.

With regards to the effectiveness of the Fe-5Ag, conflicting results were described by both Čapek, Stehlíková, Michalcová, Msallamová and Vojtěch (2016) and Wegener et al. (2011). Čapek, Stehlíková et al. (2016) prepared Fe-2Pd, Fe-2Ag and Fe-2C using powder metallurgy. They reported that while both Pd and carbon (C) additions were successful in accelerating the corrosion rate in static immersion, there was a notable decrease in corrosion rate for the Fe-2Ag alloy due to the formation of AgCl. In fact, there seemed to be no preferential attack near Ag-rich sites in post-corrosion SEM micrographs. Potentiodynamic tests supported these observations. Besides this, when testing Fe-5Ag, Wegener et al. (2011) noted only a slight increase in corrosion, which they concluded was primarily down to the higher amount of micro-porosity present in the sample, due to an increase in surface area exposed to the corrosive fluid. However, in a study by Wiesener et al. (2017) that specifically targeted the understanding of Fe-Mn-Ag corrosion, the authors assured that no AgCl should form on such alloys during the degradation process, and that attack of the Fe-Mn matrix close-by should progress considerably in the first stages. Nonetheless, it is inevitable for phosphates and oxides to deposit on the corroding surface due to the local increase in pH, as a result of the cathodic reaction occurring at the Ag-rich areas, as in Eq. (2).

In fact, Safaie, Khakbiz, Sheibani and Bagha (2015) and Sotoudehbagha, Sheibani, Khakbiz, Ebrahimi-Barough and Hermawan (2018) prepared Fe-Mn-(1-3)Ag alloys and noted no formation of Ag-chlorides. The authors noted that with increase in wt.% Ag the amount of micro-porosity reduced. This is a result of the liquid phase sintering mechanism that is activated during the processing of the material, wherein the Ag enters the liquid phase at 962 °C and flows into the pores present in the surrounding area. This increase in density contributed to the rise in hardness and strength measurements with increase in wt.% Ag. It was also revealed, that the presence of micro-porosity may have a more pronounced effect on the corrosion rate than the incorporation of the noble metal, due to the higher surface area exposed to the corrosive medium. This was revealed through the decrease in corrosion rate with increase in wt.% Ag. On the other hand, Liu, He, Xu and Guo (2018) again confirmed the effectiveness of Ag when reporting a 12 μm/year corrosion rate for Fe-30Mn-Ag alloys, as opposed to the 7 μm/year obtained by both Fe and Fe-30Mn. In this case, the samples were prepared via the casting route and therefore contained no or minimal micro-porosity.

Other techniques for incorporating the noble phase have emerged over the years. Huang and Zheng (2016) used the ion implantation technique in order to introduce Ag in the first few nanometres of the coupon. The ion-implanted coupons exhibited higher corrosion rates and a more uniform corrosion than the pure Fe coupon in the immersion test. However, the shallow depth of the Ag-containing layer limited the acceleration effect to the first few days.

Micro-patterning of pure-Fe surface using Au and Pt has also given positive results (Cheng, Huang & Zheng, 2015; Huang & Zheng, 2016). The authors reported more than double the increase of corrosion rate when depositing circular patterns of Pt, and four times the corrosion rate when patterning with Au. Similar to the issue with ion implantation, the limitation of such techniques is the effective depth of patterning. Moreover, the applicability on more complex 3D structures, such as in the cardiovascular stent and bone regeneration scaffold applications, may be somewhat limited.

3.3 Other Material Design Studies

Another approach for enhancing the performance of pure iron is to alloy with other alloying elements that, unlike Ag, Au and Pd, do not form strong microgalvanic couples. Liu and Zheng (2011) studied the effect on the mechanical properties and corrosion rate with the addition of 3wt.% of cobalt (Co), aluminium (Al), tungsten (W), tin (Sn), boron (B), carbon(C) and sulfur (S). All these elements are commonly used to strengthen steel, yet their potential for use in biomedical applications is yet unknown. Co, W, C and S were found to be effective in increasing the potentiodynamic corrosion rate, although the increase was rather low. Co, W, B, C, and S were found to improve the tensile and yield strengths of the material. Despite the fact that the study deemed certain elements unsuitable for biodegradability applications, the additions were relatively small and the effect of higher concentrations may yield different results, as was proven with additions of higher wt.% Mn (Hermawan, Purnama et al., 2010).

In a study by Wegener et al. (2011) in which the authors tested compact powder metallurgy processed Fe-0.06B, Fe-0.6P and Fe-1.6P, additions of boron also yielded positive results. The alloys all exhibited a slight increase in corrosion rates when tested in simulated body fluid. This was somewhat unexpected for Fe-P alloys due to the corrosion inhibition effect phosphorus gener-

ally has when added to steels. Moreover, while Wegener et al. (2011) reported only a small increase in corrosion, Oričaková et al. (2016) reported a threefold increase in degradation of Fe-0.5P alloy in Hank's solution. Electrochemical tests carried out on Fe-P foams also indicated an increased corrosion rate, however, changes in the exposed surface area may have had a significant effect (Hrubovčáková, Kupková & Džupon, 2016). Thus, further testing must be carried out to determine the suitability of both Fe-B and Fe-P alloys.

Furthermore, other elemental additions have been made to the promising Fe-Mn alloys. In fact, Hong et al. (2016) tested the electrochemical corrosion rate of Fe-Mn-(1-2)Mg based on the premise that Mg, being a less corrosion resistant material, will induce fast degradation of the scaffold. This was confirmed as Fe-Mn-2Mg exhibited more than nine times the degradation rate of the Fe-Mn matrix in Hank's solution. Encouraging results have also been published with additions of copper to the Fe-Mn matrix (Santanu, Raviteja, Madhuparna, Vamsi & Mangal, 2019). This addition resulted in a highly increased corrosion rate, a pronounced antimicrobial characteristic when tested with *E.coli*, and no problematic cytotoxic effects when tested with MG-63 osteosarcoma. Fe-Mn-Si is an Fe-Mn alloy that has been popularly studied by Xu, Hodgson and Cao (2015). The authors showed that while the forged alloy exhibited a rise in corrosion rate with increase in wt.% silicon (Si), the powder-processed compacts with the same composition demonstrated the opposite effect. In this case, the densification of the compact improved with increase in wt.% Si, and thus the area exposed to the corrosive fluid decreased.

The addition of second phases has also been tested using various other materials. Cheng and Zheng (2013) studied the addition of 0.5wt.% and 1wt.% of carbon nanotubes (CNTs). While CNTs have been used in numerous biomedical applications, they also have a potential of 0.2 V and are thus expected to trigger microgalvanic corrosion. In fact, sintered Fe-CNT composites exhibited significant increases in electrochemical and immersion corrosion rates, potentially due to a finer grain size compared to pure iron, as well as the microgalvanic effect. Fe-CNT also displayed an increase in ultimate compressive stress. The same effect was not observed in the study carried out by Oriňaková et al. (2013), wherein Fe-CNT foams exhibited a lower corrosion rate compared to pure Fe foams in Hank's solution.

Following the conclusion by Liu and Zheng (2011) that the presence of Fe_2O_3 as a corrosion product may lead to enhanced anodic dissolution, Cheng, Huang and Zheng (2014) proposed Fe-Fe_2O_3 for biodegradable implant applications. Results from the study showed that Fe with up to 5wt.% addition of Fe_2O_3 may potentially

be applicable for biodegradable orthopaedic implants, as they showed enhanced strength and increase in corrosion rates, resulting from the fine microstructure formed, coupled with the anodic effect of defects present.

Finally, Fe-ceramic composites have also been investigated by various research groups, including Ulum et al. (2014). Additions of 5wt.% of biodegradable materials including hydroxyapatite (HA), β-tricalcium phosphate (β-TCP) and biphasic calcium phosphate (BCP), were added to create sintered composites. The materials showed slightly poorer compressive properties, potentially due to the brittle nature of the incorporated ceramic materials. Furthermore, the Fe-ceramic showed increased degradation, however this was mainly attributed to the weight loss of the biodegradable ceramics.

4 Processing and Performance of Fe-based Foams

Apart from finding the ideal scaffold composition, researchers have also been concerned with developing an adequate technique to produce the optimal porous structure. The techniques being researched are heavily based on powder metallurgy, most probably due to the high degree of flexibility that this technology allows. Other methods, including melting and deposition, have been mentioned in select works (Alavi, Trenggono, Champagne & Hermawan, 2017; Drynda et al., 2015; Schinhammer, Steiger et al., 2013; Dehestani et al., 2017).

Quadbeck, Stephani, Kümmel, Adler and Standke (2007) first suggested the use of the replication method for this application. This method involves the use of a polymeric porous template, generally reticulated polyurethane foam, that is coated with a slurry containing metallic powders. The coated foam is then subjected to low-temperature de-binding and template removal, followed by sintering at 1000–1200 °C. The technique has been used successfully in the preparation of Fe-P, Fe-B and Fe-Fe_2O_3, with porosities generally varying between 80 and 90% (Hrubovčáková et al., 2016; Oriňaková et al., 2013; Orinak et al., 2014; Feng et al., 2018). While the low structural density suggests rather poor mechanical properties, Quadbeck et al. (2011) and Wang et al. (2017) have reported the effectiveness of varying the slurry coating thickness in achieving the desired mechanical performance.

Another technique that has been employed is the space-holder method, wherein the metallic powders are uniaxially compacted along with organic space-holding particles. The latter particles are subsequently dissolved or de-binded prior to sintering, leaving behind a porous network. Čapek and Vojtěch (2014) and Čapek, Vojtěch and Oborná (2015) used ammonium bicarbonate (NH_4CO_3) as a space-holder to obtain porous Fe. Resulting structures had up to 82% porosity with
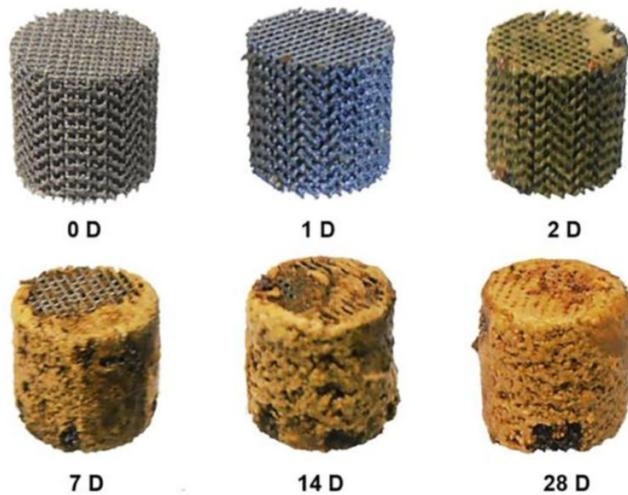
Figure 2: Visual evidence of progressive *in vitro* degradation of iron scaffolds prepared using Direct Metal Printing and tested in revised simulated body fluid for 28 days. Reprinted with permission from Elsevier from (Li et al., 2018).

macro-pore sizes varying between 250–500 μm, as is required for the application (Bose et al., 2012). Moreover, they exhibited mechanical properties comparable to human cancellous bone. Zhang and Cao (2015) used the same space-holder to prepare Fe-35Mn. The authors reported a considerable acceleration in corrosion rate in various physiological media, with increase in percentage porosity, however, the mechanical properties were not deemed suitable for load-bearing applications. Sodium chloride was also used as a space-holder by Feng et al. (2017) to produce porous Fe-Mn-Si-Pd. However, incomplete leaching of this space-holder may lead to premature corrosion of the metallic foam.

A more recent approach has been the use of 3D inkjet printing; Chou et al. (2013) were the first to develop Fe-30Mn foams using this technique. The process involves the printing of powders mixed with a water-based organic material which bind in discrete layers, approximately 100 μm thick, followed by a thermal de-binding step to remove the organic material, and sintering. Due to the absence of compacting pressure during the procedure, the initial powders are generally mechanically milled in order to facilitate the alloying of Fe and Mn. Miniature human femurs with pore sizes of 500 μm and 1 mm were successfully developed using this technique. The process was further developed by Hong et al. (2016) to produce Fe-35Mn and Fe-34Mn-Ca alloys with up to 53% porosity. A similar technique was used by Li et al. (2018) to print scaffolds with an approximate pore size of 749 μm. These scaffolds showed almost twelve times the corrosion rate of cold-rolled iron due to the increase in surface area and the reduced grain sizes, resulting

from the Direct Metal Printing (DMP) technique employed. The mechanical properties of these scaffolds remained within the acceptable range for the application after 28 days of immersion in revised simulated body fluid. The results of the DMP technique and consequent corrosion studies can be better appreciated in Fig. 2. It is interesting to note that in both the study by Chou et al. (2013) and the study by Li et al. (2018), the authors noted very similar corrosion product morphology through SEM analysis. Li et al. (2018) indicated that the products were mainly iron hydroxides, phosphates and carbonates, as analysed through XRD and Fourier-transform Infrared Spectroscopy (FTIR). EDS analysis reported in the study by Chou et al. (2013) indicate very similar compositions, although Mn-based products were also present in this case. These results also concur with those obtained when performing corrosion tests on representative coupons, wherein the typical corrosion products observed are generally iron and other metal oxides, calcium and metal phosphates and metal carbonates (Hermawan, Purnama et al., 2010; Schinhammer, Steiger et al., 2013; Wiesener et al., 2017; Zhang et al., 2012; Čapek, Stehlíková et al., 2016). However, with the most common method of analysis being EDS during SEM imaging, few studies pinpoint the exact composition of the degradation products.

Sharma and Pandey (2018a) proposed an alternative method for creating topologically ordered porous iron foams (TOPIF) that incurred fewer capital costs. The technique includes the development of a polymeric 3D printed template with a specific pore architecture. The template is used to create a phosphate-based investment mould, from which the template is removed thermally in a tube furnace. The mould is ultrasonically filled with Fe powder, which is subsequently sintered in a microwave heating furnace. The mould is removed, and the structure is post-processed. In a later study, the authors proved the flexibility of this technique in the fabrication of foams with various unit cell shapes, aimed at improving compressive performance. In fact, these structures showed a mechanical performance similar to that of human bone, however, further studies are required to analyse the degradation performance and to optimise the control over the final geometry, through the initial CAD design (Sharma & Pandey, 2018b)

On the subject of processing porous structures, an interesting development has seen the preparation of nanoporous topologies on the surface of Fe-based implants, aimed at improving tissue/implant interaction through the increase of surface area. The technique used by Heiden, Johnson and Stanciu (2016) in their studies is known as selective dealloying, wherein the more anodic component of the alloy composition is chemically removed in a four-step process. In their initial study, the

authors chose to prepare an Fe-Mn-Zn alloy, in which the more anodic components were the Mn and Zn. The first step in the creation of the nanoporous surface is to immerse the Fe-30Mn substrate in a molten Zn bath, allowing Zn to diffuse into the first few micrometers of the coupon. An etching step then removes most of the Zn from the surface. In the next annealing heat treatment, the remaining zinc moves via bulk diffusion to the surface and is consequently etched to leave behind nanometric pores. The drawback of this process was that stable oxides were formed on the surface during the annealing process, possibly limiting the biodegradability of the alloy despite the potential better osseointegration. A possible improvement was reported in their second study, when Heiden, Huang, Nauman, Johnson and Stanciu (2016) demonstrated that the use of citric acid as the final etchant led to the removal of the passivating oxide layer. This subsequently exposed the underlying metal and successfully accelerated the potentiodynamic corrosion rate, when compared to polished Fe-30Mn, while simultaneously increasing the attachment and proliferation of mouse bone marrow stromal cells (D1-UVA). The large number of optimisable parameters in this technique leaves much to be studied, however this method holds promise for the enhanced performance of biodegradable implants.

5 Cytotoxicity and *In Vivo* Corrosion of Fe Alloys

Biocompatibility is an important factor to consider whenever foreign material is to be implanted *in vivo*. In the case of corroding materials, the corrosion products must be accepted by the body while causing as little cytotoxicity as possible. For this reason, implants are often based on materials which form part of the body's normal biochemical milieu, many elemental metals falling within this bracket (Wolber, Beck, Conlon & Kruger, 2013). Fe is an essential element needed for the normal physiology of the body and, as explained elsewhere in this paper, it possesses other characteristics which render it an even more attractive option. However, in spite of this, too great an amount of Fe within the body may cause accumulation of Fe deposits within different organs, such as endocrine organs, the liver, kidneys and heart, ultimately leading to negative effects, secondary to cytotoxicity. This is often seen in diseases such as thalassaemia (Ballas, 2001). For this reason, corrosion products which are harmful at low concentrations, especially when forming part of a rapidly corroding material, may cause significant cytotoxicity and the ultimate failure of the implant.

Initially it was pure Fe that was studied as a potential material for such corroding implants, these materials having been studied both *in vivo* and *in vitro*. The first

in vivo studies included leading work by Peuster et al. (2001), where > 99.8% pure Fe was used to construct an arterial stent. This stent was then implanted into the descending aorta of 16 New Zealand White Rabbits over a study period of up to 18 months. The animals were eventually sacrificed in order to retrieve the implant, along with organs such as the heart, lungs and kidneys. This study reported no events of thrombosis within the stent in any of the studied animals. Histological examination was performed and revealed that there were only mild inflammatory reactions surrounding the stent. However, it must be noted that there were only very small amounts of corrosion products seen, suggesting only minimal corrosion of the stent, providing a potential reason for the lack of cytotoxicity. Examination of the other retrieved organs revealed no signs of Fe overload or Fe-related toxicity.

Other *in vivo* studies tend to mirror these results, with both short and long-term studies on pigs and minipigs showing good biocompatibility, some results in fact comparable to those of the widely-accepted cobalt chromium coronary stents (Waksman et al., 2008; Peuster et al., 2006; Wu et al., 2012). In studies where internal organs such as the kidneys, spleen and heart were harvested, cytotoxicity due to iron accumulation was not observed. In one study, Fe deposits were observed in the spleens of minipigs after a one-year period, but once again no signs of toxicity were noted (Peuster et al., 2006). *In vitro* results agree with those obtained from *in vivo* tests. Fe appears to have minimal cytotoxic effects on red blood cells and may in fact produce a lesser inflammatory response than stainless steel AISI 316L (Walker, Nauman, Allain & Stanciu, 2015).

As implied earlier in this review, these trials led to the conclusion that Fe on its own was not producing a sufficiently fast corrosion rate *in vivo*, thus spurring the studies on various other alloys, with Fe-Mn alloys quickly becoming the focus of many research groups. Mn is another elemental metal that is essential to normal human physiology, most frequently implicated as a co-enzyme. It is therefore beneficial in small amounts for normal body processes, but in excess it is known to cause disease, including one similar to Parkinson's disease (Horning, Caito, Tipps, Bowman & Aschner, 2015). As previously discussed, Fe-Mn alloys have been found to possess favourable mechanical and corrosion properties, but questions arose as to the alloys' potential cytotoxicity. Mn is found at much lower concentrations than Fe in the human body and therefore, it followed that Mn could have cytotoxic effects at lower concentrations than Fe. Once again however, this would depend on the rate of corrosion of the alloy.

Studies have shown that pure Mn is significantly more cytotoxic than pure Fe or AISI 316L (Cheng & Zheng,

2013; Hermawan, Purnama et al., 2010). Metabolic rates of 3T3 mouse fibroblast cell lines decrease rapidly when exposed to growing concentrations of pure Mn, especially when compared to similar concentrations of pure Fe and AISI 316L. Pure Fe and AISI 316L only have minimal effects on metabolic activities of cells, this being comparable to control cell lines cultured in standard culture medium. However, when considering the Fe-Mn alloy, its effect on metabolic rates lies somewhere in between those of pure Fe and pure Mn. Cell lines are seemingly able to tolerate significantly higher concentrations of Fe-Mn alloy than pure Mn (Hermawan, Purnama et al., 2010). However, it must also be kept in mind that Fe-Mn has a different corrosion rate to that of pure Fe and therefore, this also plays a part in the cytotoxicity profile of this alloy. With regards to haemocompatibility and thrombogenicity, *in vitro* testing appears to have positive results. Fe-Mn alloys do not seem to cause haemolysis, but instead produce results similar to those exhibited by pure Fe. Thrombogenicity is also not thought to be an issue with such implants, with *in vitro* test results being even more promising than those given for AISI 316L (Liu & Zheng, 2011; Walker et al., 2015).

In vitro cytocompatability studies confirm that Fe-Mn does have an effect on cultured cells, decreasing their metabolic rate. However, this effect does not appear to be too dissimilar from the effect of pure Fe on similarly cultured cell lines. In fact, when taking standards such as ISO 10993-5 into account, Fe-Mn still appears to have less of an effect on the metabolic rate of L929 cell lines than is accepted by this standard (Čapek, Kubásek et al., 2016). A study by Liu and Zheng (2011) documents a decrease in metabolic rate of L929 (mouse fibroblast), VSMC (rodent vascular smooth muscle cells) and ECV304 (human umbilical vein endothelial cells) after 2-day exposure to Fe-Mn corrosion products, but an increase of their metabolic functions after a further 2-day exposure to these same corrosion products (Liu & Zheng, 2011). There are other *in vitro* studies published in the literature on L929 cell lines which corroborate these works and tend towards a good biocompatibility of Fe-Mn alloys (Xu, Hodgson & Cao, 2016). Certain works include the seeding of cells such as MC3T3 (mouse osteoblast) onto alloys such as Fe-30Mn, in order to study cell growth on the alloy itself. These studies also document encouraging results, with only a few dead cells being seen on the scaffolds *in vitro* (Chou et al., 2013).

It is worth mentioning one particular study performed by Caligari Conti et al. (2018), where the effects of the corrosion products of Fe-13Mn-1.3C on hFOB 1.19 human osteoblast cell lines were studied. This study group used this material in potentiostatic tests and gathered the elute for cytocompatibility testing. The obtained solutions were split into 'Complete' and 'Supernatant' groups, the former containing both ionic and particulate products from the corrosion tests. The latter on the other hand contained only ionic products, this group allowing the cell lines to proliferate well, similar to control solutions with no corrosion product. In contrast, the 'Complete' solutions containing particulate corrosion products showed a significantly reduced rate of cell growth, inferring the important effect of particulate material on the rate of cell proliferation and growth.

Thus far, there is little in the way of *in vivo* testing of Fe-Mn alloys, but available results appear to be positive. Wires of Fe-30Mn and AISI 316L have been implanted into the femora of rats for comparison purposes. This study performed by Traverson et al. (2018) explains that once retrieved from the euthanised rats six months after implantation, Fe-30Mn showed very little cytotoxicity. In fact, Fe-30Mn produced a slightly more fibrotic tissue reaction than AISI 316L, leading to it being classified as a mild local irritant. In spite of this, the well-being of the rats did not seem to be affected during the study period. Of note is the differing corrosion results of these two materials. Stainless steel showed no signs of corrosion, while Fe-30Mn showed minimal signs of corrosion with the formation of iron oxide crystals on the surface of the implant (Traverson et al., 2018).

Another study by Drynda et al. (2015) took Fe-Mn alloys with a Mn weight below 10wt.% and implanted discs of these alloys into a subcutaneous pouch created above the gluteal muscles of mice. These animals were studied up to a maximum period of nine months, within which no ill effect was observed. However, none of the alloys exhibited significant signs of corrosion. Of note is that this study was performed with alloys with a relatively low wt.% Mn and with samples of a low surface area to volume ratio. Another reason this study gives for the poor corrosion rate is the formation of passivation layers, including layers of metal hydroxides and phosphates. These form protective layers effectively blocking the corrosion of the alloy *in vivo*, as described earlier in this review (Drynda et al., 2015).

From the discussed results, *in vivo* alloy performance is evidently different from *in vitro* performance. Considerable effort has been made on several fronts to make *in vitro* testing reflect *in vivo* conditions more. Hermawan, Purnama et al. (2010) often used a dynamic test bench for corrosion testing. This allowed for fresh medium to deliver a fresh supply of oxygen to the material surface, while also removing lingering corrosion product. In fact, researchers who made use of such setups, including Huang, Cheng, Bian and Zheng (2016), generally reported higher corrosion rates than those obtained using static setups. Apart from this, Schinhammer, Steiger et al. (2013) also revealed that using gaseous CO_2 to
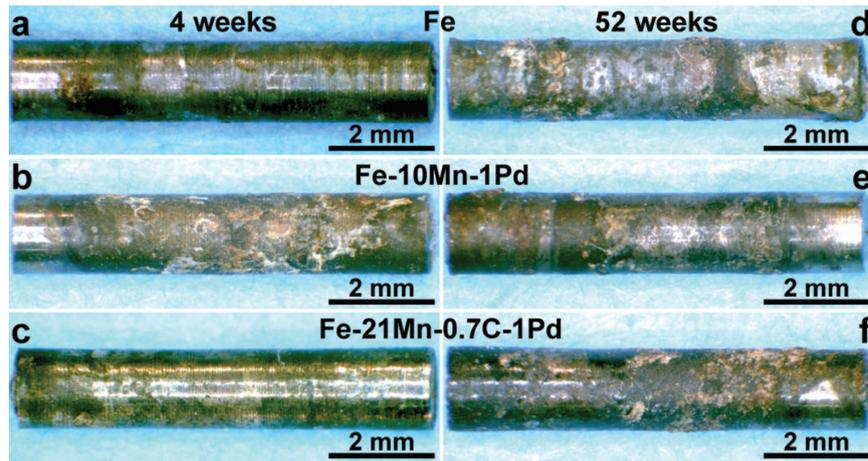
Figure 3: Optical micrographs of pins implanted in the femora of Sprague-Dawley rat, after implantation for (a–c) 4 weeks and (d–f) 52 weeks. Reprinted with permission from Elsevier from (Kraus et al., 2014).

control the pH of the physiological medium produced more realistic results, than when using pH buffers. Another interesting result was obtained by Caligari Conti et al. (2018) when adding protein (bovine serum albumin) to Hank's solution during static immersion testing of Fe-13Mn-1.2C. The corrosion rates calculated for coupons exposed to protein additions were significantly lower than those tested in only Hank's solution. The difference in corrosion rates was attributed to the formation of a biofilm on the specimen surface that limited the dissolution of metal ions. However, the same effect was not observed in potentiodynamic tests with the same solutions. In this case, the anodic potential at the surface interfered with the formation of a stable biofilm, rendering the addition of protein to the solution insignificant, when it came to altering the corrosion rate.

A review by Martinez Sanchez, Luthringer, Feyerabend and Willumeit (2015) attempts to outline the reasons for which *in vitro* and *in vivo* corrosion testing give dissimilar results when testing magnesium-based alloys. The same points seem to be valid when discussing the discrepancies noted when testing iron-based alloys, indicating that there could be a similar trend that applies to the field of biodegradable metals. In this study, the authors proposed multiple reasons for these *in vitro* and *in vivo* discrepancies. For example, the use of several different media during *in vitro* studies which differ from the *in vivo* environment in various ways, including higher concentrations of chloride, the absence of certain buffering mechanisms, or the difference in biochemical milieu, such as in the proteins present. In response to these difficulties, more physiological solutions, such as simulated body fluid (SBF), which mimic the *in vivo* environment to which such scaffolds will be exposed, now exist. These solutions appear to give more similar results *in vitro* to what is observed *in vivo*, but they are still

imperfect. Another issue raised, is that within the available literature dealing with corrosion rates *in vitro*, different testing methods have been utilized, including for example potentiodynamic testing and immersion testing. These different methods inherently give varying results, in particular, potentiodynamic testing tending to overestimate *in vivo* corrosion rates compared with other methods.

Changes in methodology have also been proposed for cytotoxic testing. Wegener et al. (2011) used a perfusion chamber to better simulate the flow of medium through cancellous bone. While cytotoxicity tests using monolayer cell culture of hFIB fibroblasts indicated strong cytotoxic effects, the dynamic setup led to a gradual increase of fibroblast levels over the 4-day testing period. However, in spite of these advances, *in vitro* testing still leaves a lot to be desired, with results often being poorly transferable to *in vivo* testing.

While Fe-Mn promises a lot in the field of bone graft substitutes, other elements have also been added to Fe-Mn to contribute to an improved property profile. Two such elements include C and Pd. It seems that Fe-Mn-C-Pd alloys offer good mechanical properties, but these too must be balanced against their cytotoxic performance (Schinhammer, Gerber, Hänzi & Uggowitzer, 2013). Both Fe-Mn-Pd and Fe-Mn-C-Pd alloys were observed in this *in vivo* study by Kraus et al. (2014). Cylindrical pins formed of these alloys were implanted into the femora of Sprage-Dawley rats, after which the test animals were observed for a period of 52 weeks. While the rats showed no signs of ill health, very limited corrosion was observed on the pin surfaces, despite the positive results obtained by Schinhammer et al. (2010) *in vitro*, as described in Section 3.2. The authors attributed this effect, displayed in Fig. 3, to the lack of oxygen present near the implantation site. The lack of

oxygen did not seem to present considerable issues when Fe stents were implanted in the coronary and illiac arteries of juvenile domestic pigs by Waksman et al. (2008) and Feng et al. (2013), respectively. Both groups reported the presence of brown-ish corrosion product after 28 days and after 12 months, indicative of the presence of Fe_2O_3. No further characterisation of the corrosion product was done in these studies, however, Fântânariu et al. (2015) published EDS results of the surface of FeMnSi coupons implanted in the tibia crest of Wistar rats and indicated the presence of a high percentage of oxides along with large quantities of calcium, phosphorus and sulphur compounds. Even in *in vivo* studies, the presence of calcium and phosphorus is a very positive outcome, indicating the promise of implant bioactivity following implantation.

Another popular addition to Fe-Mn is Ag. While the effect of Ag on the corrosion rate of the alloys has varied across various studies, Ag also imparts a separate, yet very important characteristic. Increasing the Ag proportion tends to increase the alloy's antibacterial characteristic. Sotoudehbagha et al. (2018) have observed a decrease in the number of colonies of both *E. coli* and *S. areus* after exposure to Fe-Mn alloy. This effect is even more pronounced after exposure to Fe-Mn alloys with increasing amounts of Ag. Such a material response is invaluable, and may in fact help to decrease graft infection rates in practice.

6   Conclusions

Research on the application of iron alloys for bone regeneration scaffolds has clearly made significant leaps in the past decade. With the largely positive outcomes of most studies on structure development, corrosion and cytocompatibility, the potential of these materials in matching the popularity of Mg-alloys in the field of biodegradable metal research has become even more probable. This review has mainly covered advances in the development of iron alloyed with other metals, along with manufacturing methods for these same materials. However, a recent review by Gorejová et al. (2019) highlights the interesting work being done on the application of ceramic and polymer based coatings, in order to deal with the issues of slow degradation and on occasion, cytotoxicity and osseointegration.

With a considerable basis of knowledge, it is now crucial for the focus to be shifted to the study of more realistic test samples and the development of testing procedures that better reflect the conditions *in vivo*. Several alloys have been proven to significantly accelerate the corrosion rate of pure Fe, while an increase in porosity has generally led to the same effect. Combining the two aspects of scaffold design to gain control over the degradation and mechanical performance could pave the way for targeted scaffold design, wherein custom products could be developed for optimal healing dynamics, depending on specific clinical cases. Naturally, with *in vivo* Fe alloy trials being rather limited, clinical application still seems to be a goal for the rather distant future. However, with the available *in vivo* results, as well as consistent *in vitro* mechanical performance, these scaffolds show great promise for application in orthopaedic and trauma surgery. More specifically, they promise an alternative to both autografts, that is, bone removed from another site from the same patient, and allografts, especially when managing injuries involving bone loss from load-bearing bones such as the tibia.

Acknowledgements

References

Alavi, R., Trenggono, A., Champagne, S. & Hermawan, H. (2017). Investigation on Mechanical Behavior of Biodegradable Iron Foams under Different Compression Test Conditions. *Metals*, *7*(6), 202.

American Association of Surgeons. (2010). *Summary of typical bone-graft substitutes that are commercially available – 2010*. American Association of Surgeons.

Ballas, S. K. (2001). Iron overload is a determinant of morbidity and mortality in adult patients with sickle cell disease. *Semin. Hematol. 38*, 30–36.

Bose, S., Roy, M. & Bandyopadhyay, A. (2012). Recent advances in bone tissue engineering scaffolds. *Trends Biotechnol. 30*(10), 546–554.

Buza, J. A. & Einhorn, T. (2016). Bone healing in 2016. *Clin. Cases Miner. Bone Metab. 13*(2), 101–105.

Caligari Conti, M., Aquilina, D., Paternoster, C., Vella, D., Sinagra, E., Mantovani, D., ... Buhagiar, J. (2018). Influence of cold rolling on *in vitro* cytotoxicity and electrochemical behaviour of an Fe-Mn-C biodegradable alloy in physiological solutions. *Heliyon*, *4*(11).

Campana, V., Milano, G., Pagano, E., Barba, M., Cicione, C., Salonna, G., ... Logroscino, G. (2014). Bone substitutes in orthopaedic surgery: from ba-

sic science to clinical practice. *J. Mater. Sci. Mater. Med. 25*(10), 2445–2461.

Čapek, J., Kubásek, J., Vojtěch, D., Jablonská, E., Lipov, J. & Ruml, T. (2016). Microstructural, mechanical, corrosion and cytotoxicity characterization of the hot forged FeMn30(wt.%) alloy. *Mater. Sci. Eng. C Mater. Biol. Appl. 58*, 900–908.

Čapek, J., Msallamová, Š., Jablonská, E., Lipov, J. & Vojtěch, D. (2017). PA novel high-strength and highly corrosive biodegradable Fe-Pd alloy: Structural, mechanical and *in vitro* corrosion and cytotoxicity study. *Mater. Sci. Eng. C, 79*, 550–562.

Čapek, J., Stehlíková, K., Michalcová, A., Msallamová, Š. & Vojtěch, D. (2016). Microstructure, mechanical and corrosion properties of biodegradable powder metallurgical Fe-2 wt% X (X = Pd, Ag and C) alloys. *Mater. Chem. Phys. 181*, 501–511.

Čapek, J. & Vojtěch, D. (2014). Microstructural and mechanical characteristics of porous iron prepared by powder metallurgy. *Mater. Sci. Eng. C Mater. Biol. Appl. 43*, 494–501.

Čapek, J., Vojtěch, D. & Oborná, A. (2015). Microstructural and mechanical properties of biodegradable iron foam prepared by powder metallurgy. *Mater. Des. 83*, 468–482.

Cheng, J., Huang, T. & Zheng, Y. F. (2014). Microstructure, mechanical property, biodegradation behavior, and biocompatibility of biodegradable Fe-Fe$_2$O$_3$ composites. *J. Biomed. Mater. Res. A, 102*(7), 2277–2287.

Cheng, J., Huang, T. & Zheng, Y. F. (2015). Relatively uniform and accelerated degradation of pure iron coated with micro-patterned Au disc arrays. *Mater. Sci. Eng. C Mater. Biol. Appl. 48*, 679–687.

Cheng, J., Liu, B., Wu, Y. H. & Zheng, Y. F. (2013). Comparative *in vitro* Study on Pure Metals (Fe, Mn, Mg, Zn and W) as Biodegradable Metals. *J. Mater. Sci. Technol. 29*(7), 619–627.

Cheng, J. & Zheng, Y. F. (2013). *In vitro* study on newly designed biodegradable Fe-X composites (X = W, CNT) prepared by spark plasma sintering. *J. Biomed. Mater. Res. B Appl. Biomater. 101*(4), 485–497.

Chou, D. T., Wells, D., Hong, D., Lee, B., Kuhn, H. & Kumta, P. N. (2013). Novel processing of iron-manganese alloy-based biomaterials by inkjet 3-D printing. *Acta Biomater. 9*(10), 8593–8603.

Cramer, S. D. & Covino Jr, B. S. (Eds.). (2003). *ASM Handbook Volume 13A: Corrosion: Fundamentals, Testing and Protection*. ASM International.

Dehestani, M., Trumble, K., Wang, H., Wang, H. & Stanciu, L. (2017). Effects of microstructure and heat treatment on mechanical properties and cor-

rosion behavior of powder metallurgy derived Fe-30Mn alloy. *Mater. Sci. Eng. A, 703*, 214–226.

Drynda, A., Hassel, T., Bach, F. W. & Peuster, M. (2015). *In vitro* and *in vivo* corrosion properties of new iron-manganese alloys designed for cardiovascular applications. *J. Biomed. Mater. Res. B Appl. Biomater. 103*(3), 649–660.

Eliaz, N. (2019). Corrosion of Metallic Biomaterials: A Review. *Materials, 12*(3), 407.

Fântânariu, M., Trincă, L., Solcan, C., Trofin, A., Strungaru, Ş., Şindilar, E., . . . Stanciu, S. (2015). A new Fe-Mn-Si alloplastic biomaterial as bone grafting material: *In vivo* study. *Appl. Surf. Sci. 352*, 129–139.

Feng, Q., Zhang, D., Xin, C., Liu, X., Lin, W., Zhang, W., . . . Sun, K. (2013). Characterization and *in vivo* evaluation of a bio-corrodible nitrided iron stent. *J Mater Sci Mater Med, 24*, 713–724.

Feng, Y. P., Gaztelumendi, N., Fornell, J., Zhang, H. Y., Solsona, P., Baró, M. D., . . . Sort, J. (2017). Mechanical properties, corrosion performance and cell viability studies on newly developed porous Fe-Mn-Si-Pd alloys. *J Alloys. Compd, 724*, 1046–1056.

Feng, Y., Fornell, J., Zhang, H., Solsona, P., Baró, M., Suriñach, S., . . . Sort, J. (2018). Synthesis of α-Fe$_2$O$_3$ and Fe-Mn Oxide Foams with Highly Tunable Magnetic Properties by the Replication Method from Polyurethane Templates. *Materials, 11*(2), 280.

Gorejová, R., Haverová, L., Oriňaková, R., Oriňak, A. & Oriňak, M. (2019). Recent advancements in Fe-based biodegradable materials for bone repair. *J. Mater. Sci. 54*(3), 1913–1947.

Heiden, M., Huang, S., Nauman, E., Johnson, D. & Stanciu, L. (2016). Nanoporous metals for biodegradable implants: Initial bone mesenchymal stem cell adhesion and degradation behavior. *J Biomed Mater Res A, 104*, 1747–1758.

Heiden, M., Johnson, D. & Stanciu, L. (2016). Surface modifications through dealloying of Fe-Mn and Fe-Mn-Zn alloys developed to create tailorable, nanoporous, bioresorbable surfaces. *Acta Mater, 103*, 115–127.

Heiden, M., Kustas, A., Chaput, K., Nauman, E., Johnson, D. & Stanciu, L. (2015). Effect of microstructure and strain on the degradation behavior of novel bioresorbable iron-manganese alloy implants. *J. Biomed. Mater. Res. 103*(2), 738–745.

Heiden, M., Walker, E., Nauman, E. & Stanciu, L. (2015). Evolution of novel bioresorbable iron-manganese implant surfaces and their degradation behaviors *in vitro*. *J. Biomed. Mater. Res. A, 103*(1), 185–193.

Heiden, M., Walker, E. & Stanciu, L. (2015). Magnesium, Iron and Zinc Alloys, the Trifecta of Bioresorbable Orthopaedic and Vascular Implantation – A Review. *J Biotechnol. Biomater. 5* (2), 178.

Hermawan, H., Alamdari, H., Mantovani, D. & Dubé, D. (2008). Iron-manganese: new class of metallic degradable biomaterials prepared by powder metallurgy. *Powder Metall. 51* (1), 38–45.

Hermawan, H., Dube, D. & Mantovani, D. (2010). Degradable metallic biomaterials: design and development of Fe-Mn alloys for stents. *J. Biomater. Mater. Res. A, 93* (1), 1–11.

Hermawan, H., Dubé, D. & Mantovani, D. (2007). Development of Degradable Fe-35Mn Alloy for Biomedical Application, 15-17, *107–112.*

Hermawan, H., Purnama, A., Dube, D., Couet, J. & Mantovani, D. (2010). Fe-Mn alloys for metallic biodegradable stents: Degradation and cell viability studies. *Acta Biomater. 6* (5), 1852–1860.

Hong, D., Chou, D. T., Velikokhatnyi, O. I., Roy, A., Lee, B., Swink, I., . . . Kumta, P. N. (2016). Binder-jetting 3D printing and alloy development of new biodegradable Fe-Mn-Ca/Mg alloys. *Acta biomater. 45*, 375–386.

Horning, K. J., Caito, S. W., Tipps, K. G., Bowman, A. B. & Aschner, M. (2015). Manganese Is Essential for Neuronal Health. *Annu. Rev. Nutr. 35*, 71–108.

Hrubovčáková, M., Kupková, M. & Džupon, M. (2016). Fe and Fe-P Foam for Biodegradable Bone Replacement Material: Morphology, Corrosion Behaviour, and Mechanical Properties. *Adv. Mater. Sci. Eng. 2016.*

Huang, T., Cheng, J., Bian, D. & Zheng, Y. F. (2016). Fe-Au and Fe-Ag composites as candidates for biodegradable stent materials. *J. Biomed. Mater. Res. B Appl. Biomater. 104* (2), 225–240.

Huang, T., Cheng, J. & Zheng, Y. F. (2014). *In vitro* degradation and biocompatibility of Fe-Pd and Fe-Pt composites fabricated by spark plasma sintering. *Mater. Sci. Eng. C Mater. Biol. Appl. 35*, 43–53.

Huang, T. & Zheng, Y. F. (2016). Uniform and accelerated degradation of pure iron patterned by Pt disc arrays. *Sci. Rep. 6*, 23627.

Huang, T., Zheng, Y. F. & Han, Y. (2016). Accelerating degradation rate of pure iron by zinc ion implantation. *Regen. Biomater. 3* (4), 205–215.

Katarivas Levy, G., Goldman, J. & Aghion, E. (2017). The Prospects of Zinc as a Structural Material for Biodegradable Implants – A Review Paper. *Metals, 7* (10), 402.

Kraus, T., Moszner, F., Fischerauer, S., Fiedler, M., Martinelli, E., Eichler, J., . . . Weinberg, A. (2014). Biodegradable Fe-based alloys for use in osteosynthesis: Outcome of an *in vivo* study after 52 weeks. *Acta Biomater. 10* (7), 3346–3353.

Kupková, M., Hrubovčáková, M., Kupka, M., Oriňáková, R. & Morovska Turonova, A. (2015). Sintering behaviour, graded microstructure and corrosion performance of sintered Fe-Mn biomaterials. *Int. J. Electrochem. Sci. 10*, 9256–9268.

Kupková, M., Hrubovčáková, M., Zeleňák, A., Sulowski, M., Cias, A., Oriňáková, R., . . . Kupka, M. (2015). Dimensional Changes, Microstructure, Microhardness Distributions And Corrosion Properties Of Iron And Iron-Manganese Sintered Materials. *Arch. Metall. Mater. 60* (2), 639–642.

Li, Y., Jahr, H., Lietaert, K., Pavanram, P., Yilmaz, A., Fockaert, L. I., . . . Zadpoor, A. A. (2018). Additively manufactured biodegradable porous iron. *Acta Biomater. 77*, 380–393.

Liu, B. & Zheng, Y. F. (2011). Effects of alloying elements (Mn, Co, Al, W, Sn, B, C and S) on biodegradability and *in vitro* biocompatibility of pure iron. *Acta Biomater. 7* (3), 1407–1420.

Liu, R. Y., He, R. G., Xu, L. Q. & Guo, S. F. (2018). Design of Fe-Mn-Ag Alloys as Potential Candidates for Biodegradable Metals. *Acta Metall. Sin.-Eng. 31* (6), 584–590.

Martinez Sanchez, A. H., Luthringer, B. J. C., Feyerabend, F. & Willumeit, R. (2015). Mg and Mg alloys: How comparable are *in vitro* and *in vivo* corrosion rates? A review. *Acta Biomater, 13*, 16–31.

Obayi, C. S., Tolouei, R., Mostavan, A., Paternoster, C., Turgeon, S., Okorie, B. A., . . . Mantovani, D. (2016). Effect of grain sizes on mechanical properties and biodegradation behavior of pure iron for cardiovascular stent application. *Biomatter, 6* (1), e959874.

Oričakova, R., Oričak, A., Giretová, M., Medvecký, L., Kupková, M., Hrubovčáková, M., . . . Kal'avský, F. (2016). A study of cytocompatibility and degradation of iron-based biodegradable materials. *J. Biomater. Appl. 30* (7), 1060–1070.

Orinak, A., Oriňáková, R., Orsagova Kralova, Z., Morovska Turonova, A., Kupková, M., Hrubovčáková, M., . . . Džunda, R. (2014). Sintered metallic foams for biodegradable bone replacement materials. *J. Porous Mat. 21* (2), 131–140.

Oriňáková, R., Orinak, A., Bučková, L. M., Giretova, M., Medvecky, L., Labbanczova, E., . . . Koval, K. (2013). Iron Based Degradable Foam Structures for Potential Orthopedic Applications. *Int. K. Electrochem. Sci. 8*, 12451–12465.

Peuster, M., Hesse, C., Schloo, T., Fink, C., Beerbaum, P. & von Schnakenburg, C. (2006). Long-term biocompatibility of a corrodible peripheral iron stent in the porcine descending aorta. *Biomaterials*, *27*(28), 4955–4962.

Peuster, M., Wohlsein, P., Brügmann, M., Ehlerding, M., Seidler, K., Fink, C., ... Hausdorf, G. (2001). A novel approach to temporary stenting: degradable cardiovascular stents produced from corrodible metal-results 6–18 months after implantation into New Zealand white rabbits. *Heart*, *86*(5), 563–569.

Quadbeck, P., Kümmel, K., Hauser, R., Standke, G., Adler, J., Stephani, G. & Kieback, B. (2011). Structural and Material Design of Open-Cell Powder Metallurgical Foams. *Adv. Eng. Mater. 13*(11), 1024–1030.

Quadbeck, P., Stephani, G., Kümmel, K., Adler, J. & Standke, G. (2007). Synthesis and Properties of Open-Celled Metal Foams, 534-536, *1005–1008*.

Ralston, K. D., Birbilis, N. & Davies, C. H. J. (2010). Revealing the relationship between grain size and corrosion rate of metals. *Scr. Mater. 63*(12), 1201–1204.

Safaie, N., Khakbiz, M., Sheibani, S. & Bagha, P. S. (2015). Synthesizing of Nanostructured Fe-Mn Alloys by Mechanical Alloying Process. *Procedia Mater. Sci. 11*, 381–385.

Santanu, M., Raviteja, U., Madhuparna, B., Vamsi, K. B. & Mangal, R. (2019). Fe-Mn-Cu alloy as biodegradable material with enhanced antimicrobial properties. *Mater. Lett. 237*, 323–327.

Sato, K., Ichinose, M., Hirotsu, Y. & Inoue, Y. (1989). Effects of deformation induced phase transformation and twinning on the mechanical properties of austenitic Fe-Mn-Al alloys. *ISIJ Int. 29*(10), 868–877.

Schinhammer, M., Gerber, I., Hänzi, A. C. & Uggowitzer, P. J. (2013). On the cytocompatibility of biodegradable Fe-based alloys. *Mater. Sci. Eng. C Mater. Biol. Appl. 33*(2), 782–789.

Schinhammer, M., Hänzi, A. C., Löffler, J. F. & Uggowitzer, P. J. (2010). Design strategy for biodegradable Fe-based alloys for medical applications. *Acta Biomater. 6*(5), 1705–1713.

Schinhammer, M., Steiger, P., Moszner, F., Loffler, J. F. & Uggowitzer, P. J. (2013). Degradation performance of biodegradable Fe-Mn-C(-Pd) alloys. *Mater. Sci. Eng. C Mater. Biol. Appl. 33*(4), 1882–1893.

Sharma, P. & Pandey, P. M. (2018a). A novel manufacturing route for the fabrication of topologically-ordered open-cell porous iron scaffold. *Mater. Lett. 222*, 160–163.

Sharma, P. & Pandey, P. M. (2018b). Morphological and mechanical characterization of topologically

ordered open cell porous iron foam fabricated using 3D printing and pressureless microwave sintering. *Mater. Des. 160*, 442–454.

Sotoudehbagha, P., Sheibani, S., Khakbiz, M., Ebrahimi-Barough, S. & Hermawan, H. (2018). Novel antibacterial biodegradable Fe-Mn-Ag alloys produced by mechanical alloying. *Mater. Sci. Eng. C*, *88*, 88–94.

Traverson, M., Heiden, M., Stanciu, L. A., Nauman, E. A., Jones-Hall, Y. & Breur, G. J. (2018). *In Vivo* Evaluation of Biodegradability and Biocompatibility of Fe30Mn Alloy. *Vet. Comp. Orthop. Traumatol. 31*(1), 10–16.

Ulum, M. F., Arafat, A., Noviana, D., Yusop, A. H., Nasution, A. K., Abdul Kadir, M. R. & Hermawan, H. (2014). *In vitro* and *in vivo* degradation evaluation of novel iron-bioceramic composites for bone implant applications. *Mater. Sci. Eng. C Mater. Biol. Appl. 36*, 336–344.

Waksman, R., Pakala, R., Baffour, R., Seabron, R., Hellinga, D. & Tio, F. O. (2008). Short-term effects of biocorrodible iron stents in porcine coronary arteries. *J. Interv. Cardiol. 21*(1), 15–20.

Walker, E. K., Nauman, E. A., Allain, J. P. & Stanciu, L. A. (2015). An *in vitro* model for preclinical testing of thrombogenicity of resorbable metallic stents. *J. Biomed. Mater. Res. A*, *103*(6), 2118–2125.

Wang, C., Chen, H., Zhu, X., Xiao, Z., Zhang, K. & Zhang, X. (2017). An improved polymeric sponge replication method for biomedical porous titanium scaffolds. *Mater. Sci. Eng. C Mater. Biol. Appl. 70*(Pt 2), 1192–1199.

Wataha, J. C. & Shor, K. (2010). Palladium alloys for biomedical devices. *Expert Rev. Med. Devices, 7*(4), 489–501.

Wegener, B., Sievers, B., Utzschneider, S., Müller, P., Jansson, V., Rößler, S., ... Quadbeck, P. (2011). Microstructure, cytotoxicity and corrosion of powder-metallurgical iron alloys for biodegradable bone replacement materials. *Mater. Sci. Eng. B*, *176*(20), 1789–1796.

Wiesener, M., Peters, K., Taube, A., Keller, A., Hoyer, K. P., Niendorf, T. & Grundmeier, G. (2017). Corrosion properties of bioresorbable FeMn-Ag alloys prepared by selective laser melting. *Mater. Corros. 68*(10).

Wolber, F. M., Beck, K. L., Conlon, C. A. & Kruger, M. C. (2013). Kiwifruit and mineral nutrition. *Adv. Food Nutr. Res. 68*, 233–256.

Wu, C., Hu, X., Qiu, H., Ruan, Y., Tang, Y., Wu, A., ... Gao, R. L. (2012). TCT-571 A Preliminary Study of Biodegradable Iron Stent in Mini-Swine Coronary Artery. *J. Am. Coll. Cardiol. 60*(17).

Xu, Z., Hodgson, A. M. & Cao, P. (2016). Effect of Immersion in Simulated Body Fluid on the Mechanical Properties and Biocompatibility of Sintered Fe-Mn-Based Alloys. *Metals, 6*(12), 309.

Xu, Z., Hodgson, M. A. & Cao, P. (2015). A comparative study of powder metallurgical (PM) and wrought Fe-Mn-Si alloys. *Mater. Sci. Eng. A, 630,* 116–124.

Yuen, C. K. & Ip, W. Y. (2010). Theoretical risk assessment of magnesium alloys as degradable biomedical implants. *Acta Biomater. 6*(5), 1808–1812.

Zhang, E., Chen, H. & Shen, F. (2010). Biocorrosion properties and blood and cell compatibility of pure iron as a biodegradable biomaterial. *J. Mater. Sci. Mater. Med. 21*(7), 2151–2163.

Zhang, Q. & Cao, P. (2015). Degradable porous Fe-35wt.%Mn produced via powder sintering from NH_4HCO_3 porogen. *Mater. Chem. Phys. 163,* 394–401.

Zhang, Q., Wang, X. G., Cao, P. & Gao, W. (2012). Degradation of Biodegradable Fe-Mn Alloy Produced by Powder Sintering. *Int. J. Mod. Phys.: Conf. Ser. 06,* 774–779.

Zheng, Y. F., Gu, X. N. & Witte, F. (2014). Biodegradable metals. *Mater. Sci. Eng. R Rep. 77,* 1–34.

Zhu, D., Cockerill, I., Su, Y., Zhang, Z., Fu, J., Lee, K., ... Wang, Y. (2019). Mechanical Strength, Biodegradation, and *in Vitro* and *in Vivo* Biocompatibility of Zn Biomaterials. *ACS Appl. Mater. Interfaces, 11,* 6809–6819.

Zhu, S., Huang, N., Xu, L., Zhang, Y., Liu, H., Sun, H. & Leng, Y. (2009). Biocompatibility of pure iron: *In vitro* assessment of degradation kinetics and cytotoxicity on endothelial cells. *Mater. Sci. Eng. C, 29*(5), 1589–1592.

Review Article

The Use of Cannabinoids in Parkinson's Disease

Francesca Borg[*1] **and Giuseppe Di Giovanni**[2]
[1]*Department of Physiology and Biochemistry, Faculty of Medicine and Surgery, University of Malta, Msida, Malta*
[2]*School of Biosciences, Cardiff University, Cardiff, UK*

Abstract. Parkinson's disease (PD) is a very common neurodegenerative disorder in the elderly for which there is no current cure. The neuropathological hallmark is the loss of dopaminergic cells in the substantia nigra pars compacta. Current treatments use L-DOPA and dopamine agonists to replace the lack of dopamine, however such treatments have significant limitations and side effects, thus, the need for more effective therapeutics is critical. Cannabinoids (CBs), which include Δ9-tetrahydrocannabinol, cannabidiol and Δ9-tetrahydrocannabivarin, target the endocannabinoid (ECB) system, which is highly involved in dopaminergic functions. The endocannabinoid system undergoes extensive changes in PD such as up-regulation of the ECB anandamide, in addition to variations in the concentration of CB receptors. These changes can be modified and corrected using CB1 and CB2 receptor ligands and by modulating the levels of the ECB catabolic enzyme fatty acid amide hydrolase (FAAH), in order to increase endogenous anandamide (AEA) levels. Therefore, CBs may represent a valid therapeutic alternative to treat PD. CB drugs may not only treat the symptoms of the disease, but may also help slow down disease progression. Nevertheless, with regards to motor symptoms of PD such as rigidity, bradykinesia, postural instability, resting tremors and levodopa-induced dyskinesia, evidence of the therapeutic effect of CBs is somewhat inconsistent. Although only evidence in the preclinical phase, more promising results have been seen in general regarding the neuroprotective effect of CBs, as well as in relation to sleep, depression and pain.

Keywords: Parkinson's disease, exocannabinoids, the endocannabinoid system, dyskinesia, dopamine

1 Introduction

Cannabis is derived from a plant named Cannabis sativa and has been used as a therapeutic agent since ancient times; it was clinically recognized as a medicinal compound in the mid-19th century (Borgelt, Franson, Nussbaum & Wang, 2013), following research done by Sir William B. O'Shaugnessey (Gowran, Noonan & Campbell, 2010). Medical cannabis has been widely used in medicine to treat an extensive range of clinical problems such as inflammation, oxidative stress, spasticity and rheumatism, as well as pain (Gowran et al., 2010; Babayeva, Assefa, Basu, Chumki & Loewy, 2016; Fitzcharles & Eisenberg, 2018) and Parkinson's disease (PD).

Affecting 1% of the population, PD is the second most common neurodegenerative disorder, preceded by Alzheimer's disease (Xu, Kochanek, Murphy & Tejada-Vera, 2010). It is estimated that by the year 2030 there will be approximately nine million cases of PD across the fifteen highest populated countries in the world (Dorsey et al., 2006). Today Parkinsonian patients are only treated with substitutive therapy to compensate the lack of dopamine in their brain. Therefore, such treatment is purely symptomatic and does not affect the progression of the disease, thus there is currently no cure for PD (Babayeva et al., 2016). Moreover, adverse effects appear several years after initiation of treatment (as shown by the graph in Fig. 1) and can include motor fluctuations and dyskinesia (L-DOPA induced dyskinesia: LID) (Marsden & Parkes, 1977). This review aims to highlight the effects of such CBs on patients suffering from PD.

Correspondence to: Francesca Borg (francesca.m.borg.16@um.edu.mt)

1.1 The Phyto-, Exo- and the Endocannabinoids

The products extracted from the cannabis plant consist of a large number of phytocannabinoids (phyto-CBs), a term used to differentiate cannabinoids derived from the plant, from those found in the body (i.e., endocannabinoids, ECBs) or those chemically synthesized (i.e., exocannabinoids, exo-CBs). Of about 85 phyto-CBs contained in the plant there are two significant ones, namely cannabidiol (CBD) and delta-9-Tetrahydrocannabinol (Δ9-THC) (Russo, 2011; Babayeva et al., 2016). Synthetic cannabinoids were made for cannabinoid research (exo-CBs). Exo-CBs encompass a variety of distinct chemical classes: the classical cannabinoids structurally related to THC, the nonclassical cannabinoids (cannabimimetics) including the aminoalkylindoles, 1,5-diarylpyrazoles, quinolines, and arylsulfonamides, as well as eicosanoids related to endocannabinoids (Lambert & Fowler, 2005). Synthetic cannabinoids were needed partly due to legal restrictions on natural cannabinoids, which make them difficult to obtain for research. At the time of the discovery of the cannabinoid CB1R, there were just two main chemical classes of psychotropic cannabinoids: the 'classical cannabinoids' that consist of tricyclic dibenzopyrans, such as Δ9-THC and its far more potent synthetic analogue (−)-11-hydroxy-Δ8-THC-dimethylheptyl (HU-210), as well as 'nonclassical' cannabinoids of which the bicyclic CP55940 is an important member. Subsequently, other chemical classes of psychotropic cannabinoids made their appearance, including the aminoalkylindole R-(+)-WIN55212, endogenous eicosanoids such as anandamide and 2-arachidonoyl glycerol (see below), in addition to the more recently discovered, Bayer compound, BAY 38-7271. Some early synthetic cannabinoids were also used clinically. Nabilone, a first-generation synthetic THC analog, has been used as an antiemetic to combat vomiting and nausea since 1981. Synthetic THC (Marinol, dronabinol) has been used as an antiemetic since 1985 and an appetite stimulant since 1991 (Pertwee, 2006). The CB1R inverse agonist SR141716 (also known as rimonabant), was discovered and developed by Sanofi-Aventis and temporally approved as an anorectic anti-obesity drug, but was later withdrawn in 2008 due to serious psychiatric side effects including suicide (Sam, Salem & Ghatei, 2011).

In the early 2000s, synthetic cannabinoids began to be used for recreational drug use in an attempt to achieve similar effects to that of cannabis. The likelihood of severe abuse and addiction produced by SCs are of concern for the scientific community who are also interested in the potential therapeutic value of cannabinoids (Le Boisselier, Alexandre, Lelong-Boulouard & Debruyne, 2017).

The ECBs consists of several components: lipid-based molecules called ECBs, CB receptors, transport proteins, as well as enzymes responsible for the synthesis and degradation of the ECBs (Rodríguez de Fonseca et al., 2005). Synthesis enzymes include N- arachidonoyl-phosphatidylethanolamine (NAPE)-specific phospholipase D and diacylglycerol lipase-a (DAGLa), while those that catabolize the ECBs include fatty acid amide hydrolase (FAAH) and monoacylglycerol lipase (MAGL) (Stampanoni Bassi, Sancesario, Morace, Centonze & Iezzi, 2017). The body has several ECBs, the most significant being anandamide (N- arachidonoylethanolamine or AEA) and 2-arachidonoyl-glycerol (2-AG). FAAH is responsible for the degradation of AEA, while MAGL degrades 2-AG (Gowran et al., 2010). The ECB system functions to modulate locomotion, memory, feeding behaviour, analgesia, energy balance and metabolism, stress response, sleep and many other processes (Pacher, 2006).

AEA and 2-AG mainly bind to two types of G-protein coupled receptors (GPCRs): CB1 and CB2 receptors (Gowran et al., 2010). They also bind to a lesser extent to other receptors, such as GPCR 55 (GPR55) (Ross, 2009) and GPCR 18 (GPR18) (McHugh et al., 2010). These are orphan receptors that are involved in the assembly and mobilization of microglia, and the regulation of neuropathic pain (Gowran et al., 2010). CBs can also bind to transient receptor potential cation channel subfamily V member 1 (TRPV1) (Tóth, Blumberg & Boczán, 2009), which plays a role in inflammation and pain (Costa et al., 2010), in addition to the abnormal cannabidiol receptor (McHugh et al., 2010) and the peroxisome proliferator-activated receptor (PPAR) (O'Sullivan, 2009). The function of GPR55, GPR18 and TRVP1 highlights them as potential targets in diseases involving neuropathic pain and neuroinflammation (Gowran et al., 2010). CB receptors are extensively found in the body, but are mainly concentrated in the central nervous system (CNS) and the immune system. CB1 receptors are differentially distributed in the CNS (Herkenham, Lynn, Johnson & Melvin, 1991), whilst CB2 receptors are mostly found in the immune system (Gowran et al., 2010), on cells like T-cells, B-cells and monocytes, and in organs such as the spleen. In the CNS, CB2 receptors are restricted to the brainstem (Van Sickle, 2005) and on microglia (Núñez et al., 2008). CB receptors can also be found to a lesser extent in the peripheral nervous system, reproductive organs, cardiovascular system and the gastrointestinal system (Babayeva et al., 2016).

The ECB 2-AG is 150 times more abundant in the brain than AEA and it binds to CB receptors much more potently (Buczynski & Parsons, 2010). It is the main ECB released in the midbrain in a calcium ion

dependent fashion. ECBs regulate synaptic transmission via "retrograde signaling"; one important feature of these ECBs is that they are synthesized *on demand*. Once ECBs are synthesized on the post-synaptic membrane, they travel across the synaptic cleft to bind to CB1 receptors on the presynaptic elements, where they cause cessation of neurotransmitter release (Gowran et al., 2010). This is possible because, on binding to the receptor (which is associated with a specific Golf protein), ECBs inhibit voltage-gated calcium channels and increase the activity of inward rectifying potassium channels in the presynaptic cell (Lévénès, Daniel, Soubrié & Crépel, 1998; Kreitzer & Regehr, 2001; Reggio, 2010). The calcium channel can be of the L-type, Q/P-type and N-type (Melis & Pistis, 2007). In this way, ECBs may protect against excessive excitation or inhibition (Lovinger, 2008). ECBs prevent the release of GABA on binding to CB1 receptors; this process is thus known as depolarization induced suppression of inhibition (DSI). The same process occurs in glutaminergic neurons, but in this case the process is called depolarization induced suppression of excitation (DSE) (Alger, 2002; Heinbockel, 2005). Serotonin (5-HT), acetylcholine and opioid peptides are other examples of neurotransmitters that can be modulated via CBs (Heifets & Castillo, 2009). Following activation of its receptor, the ECB is removed from the synapse via AEA membrane transporters (Melis & Pistis, 2007).

2 Effect of Some Cannabinoids on the Endocannabinoid System

2.1 Δ9-tetrahydrocannabinol

Δ9-THC is a partial agonist of both CB1 and CB2 receptors. The strength of the response it generates on binding to receptors depends on the concentration of receptors and the efficiency of their signalling. It is also affected by continuous liberation of ECBs (Pertwee, 2008). The chemical structure of Δ9-THC resembles that of AEA, implying that it mimics the effect of ECBs. It has been shown to enhance appetite and cause smooth muscle relaxation (Petrocellis, Cascio & Marzo, 2004). Smoking Δ9-THC results in the immediate onset of its effects, since the drug enters the circulation directly from the lungs. These effects become apparent within minutes and are known as a "high". Since Δ9-THC exerts its effect on cannabinoid receptors, which are highly concentrated in the CNS, the smoker experiences changes in conscious perception, euphoria and feelings of well-being. This occurs because the ECB system is over-stimulated (Basavarajappa, 2007). When used short term, Δ9-THC also influences the dopaminergic system by stimulating dopamine release, however, it has the opposite effect if used chronically (Bloomfield, Ashok, Volkow & Howes, 2016).

2.2 Cannabidiol

Cannabidiol (CBD) has different effects on the CB system as it binds to CB1 and CB2 receptors sparingly, but it is still able to block the action of Δ9-THC indirectly. This would theoretically diminish the effect of Δ9-THC. However, the effect is the opposite, as CBD causes upregulation of CB1 receptors (Devinsky et al., 2014). Its influence on the action of Δ9-THC gives it the ability to control several side effects associated with Δ9-THC, such as tachyarrhythmia and increased appetite (Russo, 2011). CBD also supplements some of Δ9-THC's desired outcomes by helping to reduce its psychotic effects, making it more tolerable to the patient and increasing its therapeutic window (Karniol & Carlini, 1973). CBD has also been shown to help with psychosis by preventing ECBs from being catabolized, thus prolonging their activity (Leweke et al., 2012). CBD can also act as an inverse agonist of the CB2 receptor, which allows it to prevent immune system cells from migrating (Pertwee, 2008). CBD delays the uptake of the neurotransmitter adenosine, thus increasing the levels of adenosine A1 receptor activation in the brain, which neutralizes a portion of the effects of activation of CB1 receptors (Devinsky et al., 2014).

2.3 Δ9-tetrahydrocannabivarin

Δ9-tetrahydrocannabivarin (Δ9-THCV) is another cannabinoid extracted from marijuana which mimics Δ9-THC. Like Δ9-THC, it binds with equal efficacy to both CB1 and CB2 receptors, but its effects are several times less potent. *In vitro*, it is a CB2 receptor partial agonist, while *in vivo* it is a CB1 receptor agonist (it is a CB1 receptor antagonist at lower concentrations). However, in tissues that contain CB1 receptors, it blocks cannabinoid receptor agonists (Pertwee, 2008).

3 Parkinson's Disease

Neurodegenerative disorders, which include PD, have also been extensively studied in the context of finding cannabinoid (CB)–based therapy (Gowran et al., 2010). The neurodegeneration associated with this disease occurs in dopaminergic cells of the nigrostriatal system. The main functional component of the basal ganglia affected is the substantia nigra pars compacta (SNc), located in the midbrain. As a result, there is a lack of dopamine in the striatum, leading to less excitation of the direct striatonigral pathway that facilitates movement, and less inhibition of the indirect pathway that antagonizes movement. Degeneration of the SNc neurons occurs over a prolonged period and can take up to several decades (Galvan & Wichmann, 2008). The cause of this is not well known but has been linked to both environmental factors, such as pesticides and certain antipsychotics, as well as genetic factors, such as mutations

in certain genes that encode proteins, like PINK1 and parkin (Gowran et al., 2010).

PD has both motor and non-motor symptoms. The non-motor symptoms appear before the motor symptoms, with the clinical diagnosis made after there is a significant loss of nigrostriatal dopamine (about 70%) (Bernheimer, Birkmayer, Hornykiewicz, Jellinger & Seitelberger, 1973). Since dopamine facilitates movement, PD patients exhibit hypokinesia, along with other motor symptoms such as bradykinesia and akinesia, muscle stiffness, and resting tremor (Thomas & Beal, 2007; Galvan & Wichmann, 2008; Rodriguez-Oroz et al., 2009). Non-motor symptoms that may be experienced include low blood pressure on pertaining an upright posture, depression, difficulty in defecation, anxiety and sleep disorders. When the disease has progressed significantly, dementia may also occur (Velseboer, de Haan, Wieling, Goldstein & de Bie, 2011; Connolly & Fox, 2013; Stocchi et al., 2014). At present, the common approach towards the treatment of motor symptoms involves dopamine replacement therapy (Goetz & Pal, 2014). This is achieved via several means, the main one being administration of dopamine precursor levodopa (L-DOPA) (Pantcheva, Reyes, Hoover, Kaelber & Borlongan, 2015). Other approaches are: (1) limiting dopamine break down with catechol-O-methyltransferase (COMT) inhibitors, (2) stimulating dopamine receptors with dopamine agonists (i.e., Sinemet), and (3) enhancing dopamine release while preventing reuptake of dopamine at presynaptic terminals (Lees, 2005; Goetz & Pal, 2014). Although effective, the efficacy of L-DOPA treatment tends to decrease when used chronically and may also cause significant side effects. Furthermore, L-DOPA only caters specifically to the motor symptoms of the disease as non-motor symptoms do not respond to this type of treatment (Pantcheva et al., 2015).

Certain drugs, such as monoamine oxidase B (MAO-B) inhibitors and COMT inhibitors, are effective in limiting the adverse effects caused by L-DOPA. In addition to these, dopamine agonists, apart from being used in early stages to treat the disease, are also useful in decreasing the motor fluctuations associated with L-DOPA therapy (Pahwa et al., 2006). The various difficulties associated with the forms of therapy mentioned has initiated research into alternative approaches, among them, medical marijuana. Several studies have shown that different phyto-CBs can be used to treat various symptoms that accompany PD, including both motor and non-motor symptoms (Babayeva et al., 2016), by counteracting oxidative stress, neuroinflammation and excitotoxicity (Fernández-Ruiz et al., 2013).

3.1 The Endocannabinoid System and Dopamine

The ECB system is central in regulating dopaminergic transmission and motor functions, hence it is highly expressed in the basal ganglia (El Manira & Kyriakatos, 2010). The globus pallidus and the SNc are components of the basal ganglia associated with locomotion and cognition, and they express large amounts of CB1 receptors and ECBs (especially AEA) (Babayeva et al., 2016). These CB1 receptor expressing neurons (including GABAergic and glutamatergic cells) are found close to, or in contact with, dopaminergic neurons. Nigrostriatal dopaminergic neurons do not contain CB1 receptors (Fernández-Ruiz, Lastres-Becker, Cabranes, González & Ramos, 2002). However, ECBs can exert their effect on dopaminergic signalling indirectly (van der Stelt & Di Marzo, 2003). Therefore, alterations in the ECB system considerably affect the nigrostriatal dopaminergic function. For example, activation of CB1 receptors on GABAergic neurons results in decreased GABA release onto dopaminergic neurons in the SNc, thus resulting in their disinhibition (Fernández-Ruiz, 2009).

On the contrary to CB1 receptors, CB2 receptors are found on dopaminergic neurons (M. García, Cinquina, Palomo-Garo, Rábano & Fernández-Ruiz, 2015). This indicates an immediate relationship between dopaminergic signalling and the ECB system. Also, some ECBs may be produced by the soma and dendrites of the dopaminergic cells themselves, which would act on the pre-synaptic cell as described previously, hence neurotransmitter release onto dopaminergic cells would be reduced (Riegel, 2004). The major factor promoting ECB release from dopaminergic cells is electrical stimulation of the aforementioned dopaminergic cell. This eventually causes the release of calcium ions from intracellular stores, thus promoting ECB release. Suppression of sn1-specific diacylglycerol lipase, which is responsible for the synthesis of 2-AG, leads to the cessation of synaptic modulation. This implies that dopaminergic cells synthesize 2-AG rather than AEA (Melis, 2004).

There are other connections between the dopaminergic system and ECBs, one of them being that the receptor TRPV1, which can bind to AEA, can be found on dopaminergic neurons (Mezey et al., 2000). The CB1 receptor is capable of forming heteromeric dimers with the dopamine receptors D1 and D2, which are both G protein-coupled. Additionally, CB1 receptors and D2 receptors make use of a common pool of G-proteins, which points to overlap in intracellular signalling (Ferré, Goldberg, Lluis & Franco, 2009). It has also been shown that D2 receptors in the striatum control the production of AEA. These associations prove that there is a link between the motor activity induced by dopamine and the ECB system, with the ECB system controlling

it via a negative feedback response (Giuffrida et al., 1999). The link between the ECB system and dopaminergic, GABAergic and glutaminergic signalling makes it an appropriate focal point in the search for drugs to treat PD.
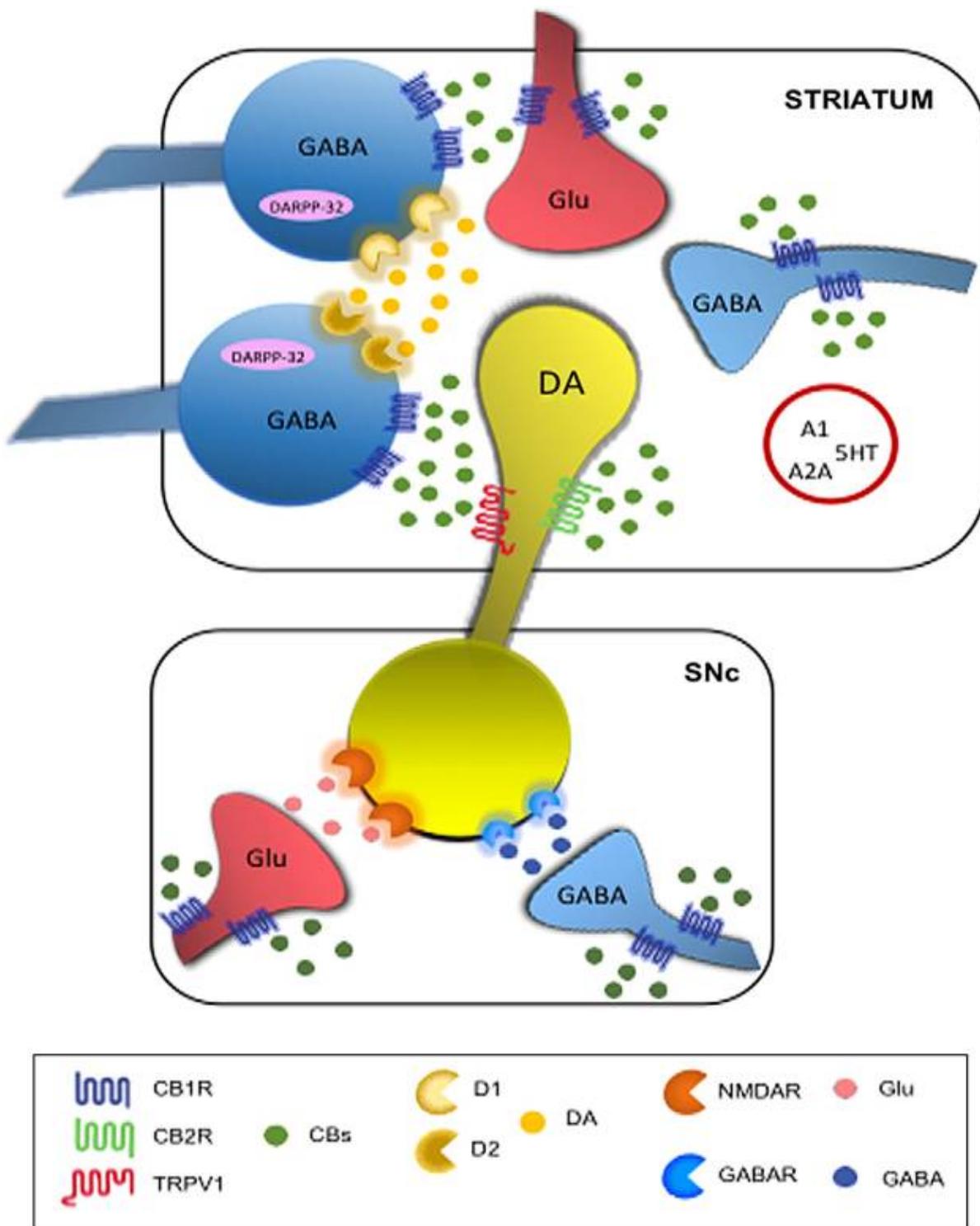

Figure 1: A schematic drawing depicting the way cannabinoids interact with the dopaminergic system in the basal ganglia. Taken from Stampanoni Bassi, Sancesario, Morace, Centonze and Iezzi (2017).

4 Alterations in the Endocannabinoid System in Parkinson's Disease

Various studies have shown that the dopamine depletion occurring in PD leads to changes in the ECB system and a disequilibrium between the direct and indirect pathways of the basal ganglia that regulate movement (V. Pisani et al., 2010).

4.1 Changes in Anandamide Levels

Neurotoxin 6-hydroxydopamine (6-OHDA) is known to mimic the lack of dopamine experienced in PD when infused into the SNc or the striatum of rodent brains; it also induced higher AEA levels (Giuffrida et al., 2004). Similar results were obtained in a group of PD sufferers (with varying degrees of severity), showing double AEA levels than that of the concentration found in control patients. It was also found that AEA levels did not vary with the severity of the disease (A. Pisani et al., 2005).

The rise in AEA might be a compensatory mechanism to make up for the loss of dopamine, as studies have shown that AEA reduces the over-activity of glutamatergic cortico-striatal pathways in rodents injected with 6-OHDA (Giuffrida et al., 2004).

4.2 Changes in CB1 Receptors

CB1 receptors are found on corticostriatal glutamatergic neurons and on GABAergic neurons innervating the medial globus pallidus, lateral globus pallidus and the SNc, as well as in excitatory projections from the subthalamic nucleus to the medial globus pallidus and the SNc (Benarroch, 2007). They are also located in striatum parvalbumin immune-reactive interneurons, cholinergic interneurons, and neurons that contain nitric oxide synthase (Fusco et al., 2004). In PD patients, basal ganglia nuclei exhibit high amounts of mRNA encoding CB1 receptors. This occurs in the intermediate and late stages of the disease when motor symptoms start appearing. In contrast, in the early stages of the disease, CB1 receptors are down-regulated (Gowran et al., 2010). Due to the lack of CB1 receptors, there is a lack of inhibition of glutamate release, which can cause excitotoxicity (van der Stelt et al., 2002). These changes can be applied to medication for PD, which will be discussed in the next paragraphs.

5 Different Approaches to Reversing Endocannabinoid Changes in PD

5.1 CB1 Receptor Agonists

Cannabinoid agonists promote hypokinesia, while cannabinoid antagonists endorse hyperkinesia. Hypokinetic effects are mediated via CB1 receptor activation (Bisogno et al., 1999). In the case of CB agonists, they can be in the form of synthetic cannabinoids, ECBs, or phyto-CB (de Lago, de Miguel, Lastres-Becker, Ramos & Fernández-Ruiz, 2004). Studies have shown that the synthetic cannabinoid HU-210 decreases LID and causes a decline in glutaminergic signalling on activation of CB1 receptors (Gilgun-Sherki, Melamed, Mechoulam & Offen, 2003). When Park-2 knockout (KO) mice (which depict a model of early-stage PD) were treated with Δ9-THC, they were reported to have high tyrosine hydroxylase levels, which is a critical enzyme in the synthesis of dopamine. They also experienced hypokinesia. However, it was of much less significance than that experienced by control mice (González et al., 2005). This is consistent with the fact that CB1 receptors are down-regulated in the early stages of the disease. Treatment with Δ9-THC and two other synthetic cannabinoids (WIN 55,212-2 and CP 55,940) has been shown to increase dopamine levels (Romero et al., 2002). Despite the undesired hypokinetic effects of cannabinoid agonists (except in the case of alleviating dyskinesia), it has been observed that they have beneficial properties when it comes to the neurodegeneration that occurs in PD. A drug named Sativex (which contains both Δ9-THC and CBD) may be used in treatment of PD, since the CBs in it have been found to act as antioxidants and protect nigrostriatal dopaminergic neurons from 6-OHDA-lesion (Lastres-Becker, Molina-Holgado, Ramos, Mechoulam & Fernández-Ruiz, 2005). This neuroprotective property suggests that cannabinoid agonists not only help treat symptoms, but also prevent the disease from progressing further.

5.2 CB1 Receptor Antagonists

In the late stages of the disease, CB1 receptor antagonists can be used to hinder the effects of high CB1 receptor density. This was investigated in a study where rodents were injected with 6-OHDA to mimic the hypokinesia experienced in PD. CB1 receptors were found to be up-regulated in these rats. The specific CB1 receptor antagonist rimonabant was found to be effective against motor inhibition. However, this was only effective when the drug was administered at small doses and was unrelated to dopamine, GABA and glutamate signalling in the striatum (González et al., 2006). In another study, PD-lesioned animals were administered varying concentrations of rimonabant and experienced an increase in glutamatergic function in the striatum, but this occurred only at the highest dose administered (of 1 mg/kg) (García-Arencibia, Ferraro, Tanganelli & Fernández-Ruiz, 2008). The combined use of rimonabant with L-DOPA diminishes the dyskinesia associated with PD (Segovia, Mora, Crossman & Brotchie, 2003). Rimonabant has also been shown to reverse the effects of Δ9-THC (Di Marzo et al., 2001), and prevent the increase in dopamine that is caused by Δ9-THC (Tanda, 1997). Despite the success in other

animal species, rimonabant is not effective in primate species (Meschler, Howlett & Madras, 2001). Other CB1 receptor antagonists, including THCV, are effective against alterations in glutaminergic transmission and prevention of hypokinesia in rodents, thus alleviating symptoms and slowing disease progression (C. García et al., 2011).

5.3 Modifying Endocannabinoid Levels

Since FAAH is the primary enzyme responsible for the degradation of AEA, FAAH inhibitors can be used to manipulate the levels of endogenous cannabinoids. Interestingly, the enzyme is already present at lower than normal concentrations in patients with PD. The FAAH inhibitor [3-(3-carbamoylphenyl)phenyl] N-cyclohexyl carbamate (URB597) has been proven to increase AEA levels, which in turn causes dopamine levels to rise (Solinas, Justinova, Golberg & Tanda, 2006). URB587 is responsible for reducing side effects associated with L-DOPA treatment, such as impulse control disorder and dopamine dysregulation syndrome, while still reaping the anti-Parkinsonian benefits of L-DOPA (Johnston et al., 2010).

5.4 CB2 Receptor Agonists

CB2 receptors have been shown to play a role in the neurological inflammation that occurs in PD. The SNc microglial cells of patients with PD express higher than normal level of CB2 receptors (Gómez-Gálvez, Palomo-Garo, Fernández-Ruiz & García, 2016). Neuroinflammation can be induced in rats by striatal injection of lipopolysaccharide (LPS) to mimic inflammation which occurs in PD. It was shown that CB2 receptor KO mice experienced an increased loss of SNc dopaminergic cells. This suggests that CB2 receptors have a neuroprotective effect. CB2 receptor KO mice also showed a higher rise in CD68 immunostaining (which corresponds to activated microglia and macrophages) than wild type animals. When the inflamed brain was treated with HU-308 (a CB2 agonist), the neuroinflammation induced by LPS was counteracted and there was a fall in CD68 immunostaining. A selective CB2 receptor agonist HU-308 has also been shown to counteract the increase in pro-inflammatory mediators caused by LPS (Gómez-Gálvez et al., 2016).

5.5 Cannabinoids and the Motor Symptoms of PD

As far as the motor symptoms of PD are concerned, studies have shown that cannabinoids are beneficial in preventing dyskinesia induced via dopamine replacement therapy (Sieradzan et al., 2001), as well as helping with PD symptoms such as resting tremors, rigidity, bradykinesia and postural instability (Consroe, Sandyk & Snider, 1986; Venderová, Ružička, Voříšek & Višňovský, 2004).

5.6 Effect of Cannabinoids on Parkinsonian Motor Symptoms

in a survey study, Venderová and colleagues (2004) found that 45.9% of a group of PD patients found cannabis helpful overall, with regards to the alleviation of Parkinsonian motor symptoms. The degree of improvement varied from mild to significant. Moreover, 30.6% experienced fewer resting tremors, 44.7% experienced an improvement with bradykinesia, and 37.7% experienced less muscular rigidity. Almost all of these patients administered cannabis orally, most often with meals and with a frequency of once every single day. Besides from one patient who had only ingested cannabis once, before the day of testing, all patients had been taking cannabis for several months. Only a small percentage of participants (4.7%) felt that cannabis made their condition worse. On average, it took approximately two months from first ingestion for patients to experience any benefits. PD patients on cannabis were tested for 11-nor-9-THCOOH (which is one of the main metabolites of Δ9-THC). It was found that those with a high concentration of this metabolite in their urine ($> 50\,\mathrm{ng/ml}$) experienced improvements in bradykinesia and rigidity, while those with a low concentration of the metabolite in their urine ($< 50\,\mathrm{ng/ml}$) experienced no improvements. Interestingly, the patient who had only taken cannabis on one occasion had very high levels of 11-nor-9-THCOOH, but did not experience any improvements. This suggests that chronic use is essential in order for cannabis to work (Venderová et al., 2004).

In another study, several PD patients were assessed before and after smoking marijuana, using the Unified Parkinson's Disease Rating Scale (UPDRS). Their scores improved significantly after smoking cannabis, both in patients with response fluctuations and in patients who lacked response fluctuations (Lotan, Treves, Roditi & Djaldetti, 2014).

Out of the four main motor symptoms of PD (bradykinesia, postural instability, resting tremor, rigidity), postural instability was the only one that did not improve with smoking marijuana. The effect of smoking marijuana lasted around 3 hours (Lotan et al., 2014).

5.7 Effect on L-DOPA Induced Dyskinesia

There are different hypotheses with regards to the cause of dyskinesia after dopamine replacement therapy. It has been suggested that it occurs because the lateral globus pallidus (GPl) is overexcited (Sieradzan et al., 2001). Interestingly, CB1 receptors are very numerous on the membrane of the pre-synaptic GABAergic neurons of the striatopallidal pathway (Sieradzan et al., 2001). Activation of these CB1 receptors reduces GABA release, thus increasing GPI GABAergic signalling and reversing dyskinesia. PD patients with

dyskinesia were assessed using the Rush Dyskinesia Disability Scale following intake of the CB1 receptor agonist nabilone. These patients' scores were greatly improved compared to those who had taken a placebo (Sieradzan et al., 2001), and this effect is dependent on CB1 receptors (More & Choi, 2015). Nabilone is also effective as an antidyskinetic in 1-methyl-4-phenyl-1,2,5,6 tetrahydropyridine (MPTP)-lesioned monkeys treated with L-DOPA. Furthemore, the pharmacological effects of L-DOPA were enhanced by 76% when given in conjunction with nabilone (Fox, Henry, Hill, Crossman & Brotchie, 2002). In one particular study, 14.1% of a group of PD sufferers experienced improvement of LID after intake of cannabis. It was noted that patients taking cannabis regularly (not less than once per day) were more likely to experience beneficial effects with regards to dyskinesia (Venderová et al., 2004).

5.8   Contraindications of Cannabis Use in Treating Motor Symptoms

Many studies have shown marijuana to be ineffective in treating motor symptoms. For instance, Carroll and colleagues showed that ingested marijuana was ineffective in treating both Parkinsonian symptoms and LID (Carroll et al., 2004). The drug Cannador (which contains unequal parts CBD and Δ9-THC) was proven to be ineffective in the treatment of LID and other motor symptoms (Carroll et al., 2004). Cannabis did not affect PD induced tremors (Stampanoni Bassi et al., 2017). Standardized methods, including the UPDRS score, were used to assess motor fluctuations and LID after administration of the CB1 receptor antagonist rimonabant, which was shown to be ineffective against both (Mesnage et al., 2004). A review by the American Academy of Neurology concluded that cannabis is ineffective in treating dyskinesia (Koppel et al., 2014). The inconsistencies between research findings imply that the use of cannabis to treat the motor symptoms of PD should be investigated more thoroughly.

6   Cannabinoids and the Non-Motor Manifestations of PD

6.1   Effect on Sleep

It has been established that PD sufferers experience problems with sleep (Murillo-Rodriguez, Pastrana-Trejo, Salas-Crisóstomo & De-la-Cruz, 2017). This can be attributed to 2 factors, dopamine-related therapy (especially dopamine agonists), as well as the direct effects of the disease, which result in alterations in the sleep-wake cycle. Patients most commonly complain of fragmented sleep, disproportionate tiredness and parasomnia (Factor, McAlarney, Sanchez-Ramos & Weiner, 1990). Patients with PD experience a disturbance in rapid eye movement (REM) sleep, with lack of reduc-

tion of normal muscle atonia, thus they physically act out their dreams. This disorder is known as REM sleep behaviour disorder (RBD) (Lin & Chen, 2018).

Moreover, PD patients are prone to experience nightmares, resulting in them flailing their limbs and shouting in their sleep (Murillo-Rodriguez et al., 2017). CBD is the main cannabinoid associated with curing sleep disorders in PD. It helps treat behaviours related to RBD without causing any undesired effects. In four unrelated cases of PD with RBD, CBD improved RBD-related behaviours, and these symptoms returned on cessation of treatment (Chagas et al., 2014). These findings suggest that cannabis can improve the quality of life of PD patients by ameliorating their sleep.

6.2   Effect on Depression

Depression is present in 40% of PD patients. However, its cause is unclear (Barrero et al., 2005; Marsh, 2013). It may be a result of neurodegeneration in certain parts of the brain, or it may occur secondary to the symptoms of the disease as they impair the quality of life (Barrero et al., 2005). While there is a correlation between the severity of symptoms and occurrence of depression, depression is more common in PD than any other neurological disorder, which implies that the neurological changes that occur in this disease play a role in the occurrence of depression (Tandberg, 1996). Alterations in neurotransmitter signalling are associated with depression. For example, dopamine depletion plays a role in the development of depression, since it leads to anhedonia, lack of motivation and apathy (Cummings, 1992). Changes in CB1 receptor density may play a role in the development of depression as they interact with the monoamine systems. By activating CB1 receptors, Δ9-THC inhibits the synaptic release of dopamine, serotonin, GABA and norepinephrine, and its influence on dopaminergic and serotonergic function could be the cause of depression in cannabis users (Musty & Kaback, 1995).

Moreover, the limbic system of rats shows high mRNA levels of CB1 receptor genes, which is implicative of the association of CB1 receptors with emotions (Matsuda, Bonner & Lolait, 1993). There is a correlation between the occurrence of depression and expansion (of less than sixteen repeats) on one of the alleles of the gene for CB1. There is a highly polymorphic triplet sequence (AAT) on the CNR1 gene that is repeated varying amounts of times in different patients. If the gene is longer (i.e. the number of AAT repeats is higher), the likelihood of depression in PD patients decreases because it is less likely that the gene will be transcribed (Barrero et al., 2005). Several studies have shown that moderate marijuana use has anti-depressive effects. For example, activating CB1 receptors via Δ9-THC, increased serotonergic function and had anti-depressive effects (Bambico, Hat-

tan, Garant & Gobbi, 2012) in rats. The anti-depressive effect of Δ9-THC is CB1 receptor-mediated, as it ceases when the CB1 receptor antagonist rimonabant is administered (Navarro et al., 1997). Preventing the degradation of AEA also results in anti-depressive effects and enhances the activity of serotonin and norepinephrine in the midbrain (Gobbi et al., 2005). A study on the effects of cannabis on PD symptoms showed that 91.5% of subjects described their emotional state as depressed before cannabis treatment, with most of these patients reporting an improvement in mood following cannabis use (Balash et al., 2017).

6.3 Effect on Pain

The ECB system is expressed in various areas of the brain and spinal cord that modulate pain (Calignano, Rana, Giuffrida & Piomelli, 1998). AEA interacts with CB1 receptors in these areas to modulate pain pathways (Calignano et al., 1998). Pain is very commonly experienced in PD patients (Dworkin et al., 2003). Many analgesics have been used over the years, including gabapentin and opioids, but these have been associated with many adverse effects (Dworkin et al., 2003). Pain in PD has been linked to the presence of a single nucleotide polymorphism (SNP) in the gene encoding the FAAH enzyme (Greenbaum et al., 2012). Patients with PD experienced pain relief 30 minutes after smoking marijuana, which may be attributed to the feelings of well-being induced by the drug, as well as to its psychoactive effects (Lotan et al., 2014). In one study, 114 patients with PD (or Parkinson-plus syndromes) who experienced pain were treated with various analgesics. It was concluded that non-steroidal anti-inflammatory drugs (NSAIDs) were the most effective in treating pain (78% effective), and these were followed closely by medical marijuana (77% effective) (Yust-Katz, Hershkovitz, Gurevich & Djaldetti, 2017).

6.4 Neuroprotective Effect

Phyto-CBs prevented damage to neurons in models of PD due to their antioxidative properties and their ability to alter glial cell signalling (Lastres-Becker et al., 2005). CBD is both an antioxidant and an anti-inflammatory agent, which points to its ability to offer neuroprotection in PD patients. CBD can cross the blood-brain barrier and exert its effects on the brain. It was found that high concentrations of CBD are not toxic to PC12 treated cells, thus demonstrating its low toxicity, in fact CBD improved the viability of MPP+ treated PC12 cells, its effect being most successful at low doses (Santos et al., 2015). CBD prevents the death of dopaminergic neurons, thus aiding dopaminergic function in rat models of PD (Lastres-Becker et al., 2005; C. García et al., 2011). CBD attenuates reactive oxygen species and prevents the increase in expression of NADPH ox-

idase 1 and 4 (NOX1 and NOX4) (Pan et al., 2008). The superoxide produced by NOX enzymes is involved in the demise of dopaminergic cells in PD animal models (Hernandes, Café-Mendes & Britto, 2013). CBD decreases the formation of inducible nitric oxide synthase (iNOS) (Pan et al., 2008) and enhances the production of Cu^{2+}, Zn-superoxide dismutase mRNA, which when transcribed, prevents the formation of reactive oxygen species (García-Arencibia et al., 2007).

Cannabinoids can act as anti-inflammatory agents by preventing activation of microglia (which are found at higher levels in the SNc of patients with PD) and inhibiting the release of toxic cytokines (Sayd et al., 2014). Therefore, cannabinoids indirectly prevent dopaminergic cell death by counteracting neuroinflammation, one of the factors that contributes to their death (Clark & Kadodek, 2016). Cannabinoids also protect against neuroinflammation via reperfusion of the site of injury and by reversing vasoconstriction (Sagredo et al., 2007). CB2 receptor agonists (including cannabinoids) cause up-regulation of these receptors in microglia of the SNc, and also offer neuroprotection to dopaminergic neurons by shielding them from the neuroinflammation caused by the same activated microglia (C. García et al., 2011). Brain slices from Wistar rats were oxygen and glucose deprived, after which they were treated with the cannabinoid CB1/CB2 receptor agonist WIN55212. This drug prevented cell damage, production of nitric oxide and release of glutamate, among other neurotoxic effects (Fernández-López et al., 2006). By activating CB1 receptors, Δ9-THC also acts as an anti-inflammatory agent (Fishbein-Kaminietsky, Gafni & Sarne, 2014). CBD has been combined with Δ9-THC in various drugs, one of which is Sativex, and together, they have been proven to have neuroprotective effects (Iuvone, Esposito, De Filippis, Scuderi & Steardo, 2009). Δ9-THC can augment the survival of dopaminergic cells and enhance the activity of PPARγ receptors (Zeissler, Hanemann, Zajicek & Carroll, 2012). Apart from having neuroprotective properties of its own, CBD helps reduce the side effects associated with Δ9-THC administration. In such drugs, CBD and THC interact with both CB1 and CB2 receptors (Sagredo et al., 2011).

Additionally, Δ9-THCV plays a role in neuroprotection (C. García et al., 2011). In the long term, it counteracts the loss of dopaminergic cells in the SNc, which occurs as a result of 6-OHDA treatment. It also prevents the activation of microglia. It is likely that both of these effects are caused by Δ(9)-THCV's antioxidant properties, rather than its action on CB2 receptors, since they were also observed on treatment with CBD. To further support this hypothesis, when CB2 receptor KO mice were treated with 6-OHDA, their loss of dopaminergic cells was of the same degree as wild type mice, imply-

Table 1: An overview of the action of different cannabinoids on motor and non-motor PD symptoms.

Effect of cannabinoids on motor impairments			
Δ9-THC (> 50 ng/ml of its metabolite 11-nor-Δ-9-THCOOH detected in urine)	Humans	Improvement of bradykinesia and rigidity	Venderová, Ružička, Voříšek and Višňovský (2004)
Smoked cannabis (inhaled amount: 0.5 g)	Humans	Improvement of tremor, rigidity, bradykinesia. Little effect on posture	Lotan, Treves, Roditi and Djaldetti (2014)
Nabilone (dose not specified)	Humans and primate models	L-DOPA induced dyskinesia	More and Choi (2015)
Effect of cannabinoids on sleep			
CBD (75 mg/day and 300 mg/day were both effective)	Humans	REM sleep behaviour disorder related behaviour	Chagas et al. (2014)
Effect of cannabinoids on depression			
Δ9-THC (repeated administration of 1 mg/kg)	male adult Sprague–Dawley rats	Improved serotonergic function	Bambico, Hattan, Garant and Gobbi (2012)
Smoked cannabis (0.9 ± 0.5 g/day)	Humans	Improved mood	Balash et al. (2017)
Effect of cannabinoids on pain			
Cannabis (dose not specified)	Humans	Pain relief	Yust-Katz, Hershkovitz, Gurevich and Djaldetti (2017)
Smoked cannabis (inhaled amount: 0.5 g)	Humans	Pain relief	Lotan, Treves, Roditi and Djaldetti (2014)
Neuroprotective effect of cannabis			
CBD (1 μM dose was most effective)	MPP+ treated PC12 cells from rats	Antioxidant and anti-inflammatory	Santos et al. (2015)
Δ(9)-THCV (2 mg/kg⁻¹ for 14 days)	6-hydroxydopamine-lesioned animals	Anti-oxidant	C. García et al. (2011)
WIN55212 (50 μM)	Glucose and oxygen deprived brain slices from Wistar rats	Prevents cell damage, production of nitric oxide and release of glutamate	Fernández-López et al. (2006)

ing that the presence of CB2 receptors did not affect cell loss (C. García et al., 2011). However, a conflicting study showed that CB2 receptors do play a role in neuroinflammation (Gómez-Gálvez et al., 2016). It was observed in cells of the SNc and striatum that had been treated with lipopolysaccharide (LPS), a neurotoxin used to mimic neuroinflammation and PD in animals, these receptors were up-regulated. The amount of activated microglia was also significantly higher in CB2 receptor deficient cells when compared to the wild type. Furthermore, when compared to the wild type animals, the loss of dopaminergic cells was more rapid, and occurred to a greater extent, in cells lacking CB2 receptors

(Gómez-Gálvez et al., 2016). An overview of the effects of cannabinoids on the motor and non-motor symptoms of PD can be found in Table 1.

7 Conclusion

Medical marijuana has been suggested to have some potential benefits as a treatment for PD. From the literature reviewed here, CBs may be more effective against non-motor symptoms compared to motor symptoms, although meliorates LID. Marijuana has already been legalized in Israel for therapy of pain and tremor in people suffering from PD (Yust-Katz et al., 2017), and it is effective in treating comorbid sleep problems and de-

pression. It may act as a disease-modifying treatment by blocking/delaying disease progression due to oxidative stress and neuroinflammation, an effect that seems to be independent from CB receptors. Among the CBs, the most promising in this regard seems to be CBD.

Although there are contradictory studies which today do not support the therapeutic use of marijuana for different disorders such as PD, medicinal marijuana has been legalized in various areas of the world, including Malta. This step forward has been largely driven by the media, politics and the public, signifying the need for more preclinical/clinical research, including carefully planned clinical trial studies regarding the therapeutic effects of cannabis, in order to enable medical professionals to provide new effective cures for their patients.

References

Alger, B. E. (2002). Retrograde signaling in the regulation of synaptic transmission: focus on endocannabinoids. *Prog. Neurobiol. 68*, 247–286.

Babayeva, M., Assefa, H., Basu, P., Chumki, S. & Loewy, Z. (2016). Marijuana Compounds: A Nonconventional Approach to Parkinson's Disease Therapy. *Parkinson's Disease, 2016*, 19.

Balash, Y., Bar-Lev Schleider, L., Korczyn, A., Shabtai, H., Knaani, J., Rosenberg, A., ... Gurevich, T. (2017). Medical Cannabis in Parkinson Disease. *Clin. Neuropharmacol. 40*, 268–272.

Bambico, F., Hattan, P., Garant, J. & Gobbi, G. (2012). Effect of delta-9- tetrahydrocannabinol on behavioral despair and on pre- and post-synaptic serotonergic transmission. *Prog. Neuro-Pschopharmacology Biol. Psychiatry, 38*, 88–96.

Barrero, F., Ampuero, I., Morales, B., Vives, F., de Dios Luna del Castillo, J., Hoenicka, J. & García Yébenes, J. (2005). Depression in Parkinson's disease is related to a genetic polymorphism of the cannabinoid receptor gene (CNR1). *Pharmacogenomics J. 5*, 135–141.

Basavarajappa, B. (2007). Neuropharmacology of the Endocannabinoid Signaling System-Molecular Mechanisms, Biological Actions and Synaptic Plasticity. *Curr. Neuropharmacol. 5*, 81–97.

Benarroch, E. (2007). Endocannabinoids in basal ganglia circuits: Implications for Parkinson disease. *Neurology, 69*, 306–309.

Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K. & Seitelberger, F. (1973). Brain dopamine and the syndromes of Parkinson and Huntington Clinical, morphological and neurochemical correlations. *J. Neurol. Sci. 20*, 415–455.

Bisogno, T., Berrendero, F., Ambrosino, G., Cebeira, M., Ramos, J., Fernandez-Riuz, J. & Di Marzo, V. (1999). Brain Regional Distribution of Endocannabinoids: Implications for Their Biosynthesis and Biological Function. *Biochem. Biophys. Res. Commun. 256*, 377–380.

Bloomfield, M., Ashok, A., Volkow, N. & Howes, O. (2016). The effects of Δ9- tetrahydrocannabinol on the dopamine system. *Nature, 539*, 369–377.

Borgelt, L., Franson, K., Nussbaum, A. & Wang, G. (2013). The Pharmacologic and Clinical Effects of Medical Cannabis. Pharmacotherapy. *J. Hum. Pharmacol. Drug Ther. 33*, 195–209.

Buczynski, M. & Parsons, L. (2010). Quantification of brain endocannabinoid levels: methods, interpretations and pitfalls. *Br. J. Pharmacol. 160*, 423–442.

Calignano, A., Rana, G., Giuffrida, A. & Piomelli, D. (1998). Control of pain initiation by endogenous cannabinoids. *Nature, 394*, 277–281.

Carroll, C., Bain, P., Teare, L., Liu, X., Joint, C., Wroath, C., ... Zajicek, J. (2004). Cannabis for dyskinesia in Parkinson disease: A randomized double-blind crossover study. *Neurology, 63*, 1245–1250.

Chagas, M., Eckeli, A., Zuardi, A., Pena-Pereira, M., Sobreira-Neto, M., Sobreira, E., ... Crippa, J. (2014). Cannabidiol can improve complex sleep-related behaviours associated with rapid eye movement sleep behaviour disorder in Parkinson's disease patients: a case series. *J. Clin. Pharm. Ther. 39*, 564–566.

Clark, L. & Kadodek, T. (2016). The Immune System and Neuroinflammation as Potential Sources of Blood-Based Biomarkers for Alzheimer's Disease, Parkinson's Disease, and Huntington's Disease. *ACS Chem. Neurosci. 7*, 520–527.

Connolly, B. & Fox, S. (2013). Treatment of Cognitive, Psychiatric, and Affective Disorders Associated with Parkinson's Disease. *Neurotherapeutics, 11*, 78–91.

Consroe, P., Sandyk, R. & Snider, S. (1986). Open label evaluation of cannabidiol in dystonic movement disorders. *Int. J. Neurosci. 30*, 277–282.

Costa, B., Bettoni, I., Petrosino, S., Comelli, F., Giagnoni, G. & Di Marzo, V. (2010). The dual fatty acid amide hydrolase/TRPV1 blocker, N-arachidonoyl-serotonin, relieves carrageenan-induced inflammation and hyperalgesia in mice. *Pharmacol. Res. 61*, 537–546.

Cummings, J. (1992). Depression and Parkinson's disease: a review. *Am. J. Psychiatry, 149*, 443–454.

de Lago, E., de Miguel, R., Lastres-Becker, I., Ramos, J. & Fernández-Ruiz, J. (2004). Involvement of vanilloid-like receptors in the effects of anandamide on motor behavior and nigrostriatal dopaminergic activity: *in vivo* and *in vitro* evidence. *Brain Res. 1007*, 152–159.

Devinsky, O., Cilio, M., Cross, H., Fernández-Ruiz, J., French, J., Hill, C., . . . Friedman, D. (2014). Cannabidiol: Pharmacology and potential therapeutic role in epilepsy and other neuropsychiatric disorders. *Epilepsia*, *55*, 791–802.

Di Marzo, V., Lastres-Becker, I., Bisogno, T., De Petrocellis, L., Milone, A., Davis, J. & Fernández-Ruiz, J. (2001). Hypolocomotor effects in rats of capsaicin and two long chain capsaicin homologues. *Eur. J. Pharmacol. 420*, 123–131.

Dorsey, E., Constantinescu, R., Thompson, J., Biglan, K., Holloway, R., Kieburtz, K., . . . Tanner, C. (2006). Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, *68*, 384–386.

Dworkin, R. H., Backonja, M., Rowbotham, M. C., Allen, R. R., Argoff, C. R., Bennett, G. J., . . . Weinstein, S. M. (2003). Advances in neuropathic pain: diagnosis, mechanisms, and treatment recommendations. *Arch. Neurol. 60*, 1524–1534.

El Manira, A. & Kyriakatos, A. (2010). The Role of Endocannabinoid Signaling in Motor Control. *Physiology*, *25*, 230–238.

Factor, S., McAlarney, T., Sanchez-Ramos, J. & Weiner, W. (1990). Sleep disorders and sleep effect in Parkinson's disease. *Mov. Disord. 5*, 280–285.

Fernández-López, D., Martínez-Orgado, J., Nuñez, E., Romero, J., Lorenzo, P., Moro, M. & Lizasoain, I. (2006). Characterization of the Neuroprotective Effect of the Cannabinoid Agonist WIN-55212 in an *In Vitro* Model of Hypoxic-Ischemic Brain Damage in Newborn Rats. *Pediatr. Res. 60* (2), 169–173.

Fernández-Ruiz, J. (2009). The endocannabinoid system as a target for the treatment of motor dysfunction. *Br. J. Pharmacol. 156*, 1029–1040.

Fernández-Ruiz, J., Lastres-Becker, I., Cabranes, A., González, S. & Ramos, J. (2002). Endocannabinoids and basal ganglia functionality. *Prostaglandins, Leukot. Essent. Fat. Acids (PLEFA)*, *66*, 257–267.

Fernández-Ruiz, J., Sagredo, O., Pazos, M., García, C., Pertwee, R. G., Mechoulam, R. & Martínez-Orgado, J. (2013). Cannabidiol for neurodegenerative disorders: important new clinical applications for this phytocannabinoid? *Br. J. Clin. Pharmacol. 75*, 323–333.

Ferré, S., Goldberg, S., Lluis, C. & Franco, R. (2009). Looking for the role of cannabinoid receptor heteromers in striatal function. *Neuropharmacology*, *56*, 226–234.

Fishbein-Kaminietsky, M., Gafni, M. & Sarne, Y. (2014). Ultralow doses of cannabinoid drugs protect the mouse brain from inflammation-induced cognitive damage. *J. Neurosci. Res. 92*, 1669–1677.

Fitzcharles, M. & Eisenberg, E. (2018). Medical cannabis: A forward vision for the clinician. *Eur. J. Pain*, *22*, 485–491.

Fox, S., Henry, B., Hill, M., Crossman, A. & Brotchie, J. (2002). Stimulation of cannabinoid receptors reduces levodopa-induced dyskinesia in the MPTP-lesioned nonhuman primate model of Parkinson's disease. *Mov. Disord. 17*, 1180–1187.

Fusco, F., Martorana, A., Giampà, C., De March, Z., Farini, D., D'Angelo, V., . . . Bernardi, G. (2004). Immunolocalization of CB1receptor in rat striatal neurons: A confocal microscopy study. *Synapse*, *53*, 159–167.

Galvan, A. & Wichmann, T. (2008). Pathophysiology of Parkinsonism. *Clin. Neuropharmacol. 119*, 1459–1474.

García, C., Palomo-Garo, C., García-Arencibia, M., Ramos, J., Pertwee, R. G. & Fernández- Ruiz, J. (2011). Symptom-relieving and neuroprotective effects of the phytocannabinoid Δ9-THCV in animal models of Parkinson's disease. *Br. J. Pharmacol. 163*, 1495–1506.

García, M., Cinquina, V., Palomo-Garo, C., Rábano, A. & Fernández-Ruiz, J. (2015). Identification of CB2 receptors in human nigral neurons that degenerate in Parkinson's disease. *Neurosci. Lett. 587*, 1–4.

García-Arencibia, M., Ferraro, L., Tanganelli, S. & Fernández-Ruiz, J. (2008). Enhanced striatal glutamate release after the administration of rimonabant to 6-hydroxydopamine-lesioned rats. *Neurosci. Lett. 438*, 10–13.

García-Arencibia, M., González, S., de Lago, E., Ramos, J., Mechoulam, R. & Fernández-Ruiz, J. (2007). Evaluation of the neuroprotective effect of cannabinoids in a rat model of Parkinson's disease: Importance of antioxidant and cannabinoid receptor-independent properties. *Brain Res. 1134*, 162–170.

Gilgun-Sherki, Y., Melamed, E., Mechoulam, R. & Offen, D. (2003). The CB1Cannabinoid Receptor Agonist, HU-210, Reduces Levodopa-Induced Rotations in 6- Hydroxydopamine-Lesioned Rats. *Pharmacol. Toxicol. 93*, 66–70.

Giuffrida, A., Leweke, F., Gerth, C., Schreiber, D., Koethe, D., Faulhaber, J., . . . Piomelli, D. (2004). Cerebrospinal Anandamide Levels are Elevated in Acute Schizophrenia and are Inversely Correlated with Psychotic Symptoms. *Neuropsychopharmacology*, *29*, 2108–2114.

Giuffrida, A., Parsons, L., Kerr, T., de Fonseca, F., Navarro, M. & Piomelli, D. (1999). Dopamine activation of endogenous cannabinoid signaling in dorsal striatum. *Nat. Neurosci. 2*, 358–363.

Gobbi, G., Bambico, F., Mangieri, R., Bortolato, M., Campolongo, P., Solinas, M., . . . Piomelli, D.

(2005). Antidepressant-like activity and modulation of brain monoaminergic transmission by blockade of anandamide hydrolysis. *Proc. Natl. Acad. Sci. 102*, 18620–18625.

Goetz, C. & Pal, G. (2014). Initial management of Parkinson's disease. *BMJ, 349*, g6258.

Gómez-Gálvez, Y., Palomo-Garo, C., Fernández-Ruiz, J. & García, C. (2016). Potential of the cannabinoid CB2 receptor as a pharmacological target against inflammation in Parkinson's disease. *Prog. Neuro-Pschopharmacology Biol. Psychiatry, 64*, 200–208.

González, S., Mena, M., Lastres-Becker, I., Serrano, A., de Yébenes, J., Ramos, J. & Fernández-Ruiz, J. (2005). Cannabinoid CB1 receptors in the basal ganglia and motor response to activation or blockade of these receptors in parkin-null mice. *Brain Res. 1046*, 195–206.

González, S., Scorticati, C., García-Arencibia, M., de Miguel, R., Ramos, J. & Fernández-Ruiz, J. (2006). Effects of rimonabant, a selective cannabinoid CB1 receptor antagonist, in a rat model of Parkinson's disease. *Brain Res. 1073-1074*, 209–219.

Gowran, A., Noonan, J. & Campbell, V. (2010). The Multiplicity of Action of Cannabinoids: Implications for Treating Neurodegeneration. *CNS Neurosci. Ther. 17*, 637–644.

Greenbaum, L., Tegeder, I., Barhum, Y., Melamed, E., Roditi, Y. & Djaldetti, R. (2012). Contribution of genetic variants to pain susceptibility in Parkinson disease. *Eur. J. Pain, 16*, 1243–1250.

Heifets, B. & Castillo, P. (2009). Endocannabinoid Signaling and Long-Term Synaptic Plasticity. *Annu. Rev. Physiol. 71*, 283–306.

Heinbockel, T. (2005). Endocannabinoid Signaling Dynamics Probed with Optical Tools. *J. Neurosci. 25*, 9449–9459.

Herkenham, M., Lynn, A., Johnson, M. & Melvin, L. (1991). Characterization and localization of cannabinoid receptors in rat brain: a quantitative *in vitro* autoradiographic study. *J. Neurosci. 11*(2), 563–583.

Hernandes, M., Café-Mendes, C. & Britto, L. (2013). NADPH Oxidase and the Degeneration of Dopaminergic Neurons in Parkinsonian Mice. *Oxid. Med. Cell. Longev. 2013*, 1–13.

Iuvone, T., Esposito, G., De Filippis, D., Scuderi, C. & Steardo, L. (2009). Cannabidiol: A Promising Drug for Neurodegenerative Disorders? *CNS Neurosci. Ther. 15*, 65–75.

Johnston, T., Huot, P., Fox, S., Wakefield, J., Skyes, K., Barolini, W., ... Brotchie, J. (2010). Fatty Acid Amide Hydrolase (FAAH) Inhibition Reduces L-3,4-Dihydroxyphenylalanine-Induced Hyperactivity in the 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine-Lesioned Non-Human Primate Model of Parkinson's Disease. *J. Pharmacol. Exp. Ther. 336*, 423–430.

Karniol, I. & Carlini, E. (1973). Pharmacological interaction between cannabidiol and Δ9- tetrahydrocannabinol. *Psychopharmacologia, 33*, 53–70.

Koppel, B., Brust, J., Fife, T., Bronstein, J., Youssof, S., Gronseth, G. & Gloss, D. (2014). Systematic review: Efficacy and safety of medical marijuana in selected neurologic disorders: Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology, 82*, 1556–1563.

Kreitzer, A. & Regehr, W. (2001). Retrograde Inhibition of Presynaptic Calcium Influx by Endogenous Cannabinoids at Excitatory Synapses onto Purkinje Cells. *Neuron, 29*, 717–727.

Lambert, D. M. & Fowler, C. J. (2005). The endocannabinoid system: drug targets, lead compounds, and potential therapeutic applications. *J Med Chem, 48*(16), 5059–5087.

Lastres-Becker, I., Molina-Holgado, F., Ramos, J., Mechoulam, R. & Fernández-Ruiz, J. (2005). Cannabinoids provide neuroprotection against 6-hydroxydopamine toxicity *in vivo* and *in vitro*: Relevance to Parkinson's disease. *Neurobiol. Dis. 19*, 96–107.

Le Boisselier, R., Alexandre, J., Lelong-Boulouard, V. & Debruyne, D. (2017). Focus on cannabinoids and synthetic cannabinoids. *Clin Pharmacol Ther, 101*(2), 220–229.

Lees, A. (2005). Alternatives to Levodopa in the Initial Treatment of Early Parkinson's Disease. *Drugs Aging, 22*, 731–740.

Lévénès, C., Daniel, H., Soubrié, P. & Crépel, F. (1998). Cannabinoids decrease excitatory synaptic transmission and impair long-term depression in rat cerebellar Purkinje cells. *J. Physiol. 510*, 867–879.

Leweke, F., Piomelli, D., Pahlisch, F., Muhl, D., Gerth, C., Hoyer, C., ... Koethe, D. (2012). Cannabidiol enhances anandamide signaling and alleviates psychotic symptoms of schizophrenia. *Transl. Psychiatry, 2*, e94.

Lin, Y. & Chen, S. (2018). RBD: a red flag for cognitive impairment in Parkinson's disease? *Sleep Med. 44*, 38–44.

Lotan, I., Treves, T., Roditi, Y. & Djaldetti, R. (2014). Cannabis (Medical Marijuana) Treatment for Motor and Non-Motor Symptoms of Parkinson Disease. *Clin. Neuropharmacol. 37*, 41–44.

Lovinger, D. (2008). Presynaptic Modulation by Endocannabinoids. *Handb. Exp. Pharmacol.* (184), 435–477.

Marsden, C. & Parkes, J. (1977). Success and Problems of Long-term Levodopa Therapy in Parkinson's Disease. *Lancet*, *309*, 345–349.

Marsh, L. (2013). Depression and Parkinson's Disease: Current Knowledge. *Curr. Neurol. Neurosci. Rep.* *13*(12), 409.

Matsuda, L., Bonner, T. & Lolait, S. (1993). Localization of cannabinoid receptor mRNA in rat brain. *J. Comp. Neurol. 327*, 535–550.

McHugh, D., Hu, S., Rimmerman, N., Juknat, A., Vogel, Z., Walker, J. & Bradshaw, H. (2010). N-arachidonoyl glycine, an abundant endogenous lipid, potently drives directed cellular migration through GPR18, the putative abnormal cannabidiol receptor. *BMC Neurosci. 11*, 44.

Melis, M. (2004). Endocannabinoids Mediate Presynaptic Inhibition of Glutamatergic Transmission in Rat Ventral Tegmental Area Dopamine Neurons through Activation of CB1 Receptors. *J. Neurosci. 24*, 53–62.

Melis, M. & Pistis, M. (2007). Endocannabinoid Signaling in Midbrain Dopamine Neurons: More than Physiology? *Curr. Neuropharmacol. 5*, 268–277.

Meschler, J., Howlett, A. & Madras, B. (2001). Cannabinoid receptor agonist and antagonist effects on motor function in normal and 1-methyl-4-phenyl-1,2,5,6- tetrahydropyridine(MPTP)-treatednonhuman primates. *Psychopharmacology (Berl). 156*, 79–85.

Mesnage, V., Houeto, J., Bonnet, A., Clavier, I., Arnulf, I., Cattelin, F., . . . Agid, Y. (2004). Neurokinin B, Neurotensin, and Cannabinoid Receptor Antagonists and Parkinson Disease. *Clin. Neuropharmacol. 27*(3), 108–110.

Mezey, E., Toth, Z., Cortright, D., Arzubi, M., Krause, J., Elde, R., . . . Szallasi, A. (2000). Distribution of mRNA for vanilloid receptor subtype 1 (VR1), and VR1-like immunoreactivity, in the central nervous system of the rat and human. *Proc. Natl. Acad. Sci. 97*, 3655–3660.

More, S. & Choi, D. (2015). Promising cannabinoid-based therapies for Parkinson's disease: motor symptoms to neuroprotection. *Mol. Neurodegener. 10*, 17.

Murillo-Rodriguez, E., Pastrana-Trejo, J., Salas-Crisóstomo, M. & De-la-Cruz, M. (2017). The Endocannabinoid System Modulating Levels of Consciousness, Emotions and Likely Dream Contents. *CNS Neurol. Disord. - Drug Targets*, *16*(4), 370–379.

Musty, R. & Kaback, L. (1995). Relationships between motivation and depression in chronic marijuana users. *Life Sci. 56*, 2151–2158.

Navarro, M., Hernández, E., Muñoz, R., del Arco, I., Villanúa, M., Carrera, M. & Rodríguez de Fonseca, F. (1997). Acute administration of the CB1 cannabinoid receptor antagonist SR 141716A induces anxiety-like responses in the rat. *Neuroreport*, *8*, 491–496.

Núñez, E., Benito, C., Tolón, R., Hillard, C., Griffin, W. & Romero, J. (2008). Glial expression of cannabinoid CB2 receptors and fatty acid amide hydrolase are beta amyloid-linked events in Down's syndrome. *Neuroscience*, *151*, 104–110.

O'Sullivan, S. (2009). Cannabinoids go nuclear: evidence for activation of peroxisome proliferator-activated receptors. *Br. J. Pharmacol. 152*, 576–582.

Pacher, P. (2006). The Endocannabinoid System as an Emerging Target of Pharmacotherapy. *Pharmacol. Rev. 58*, 389–462.

Pahwa, R., Factor, S., Lyons, K., Ondo, W., Gronseth, G., Bronte-Stewart, H., . . . Weiner, W. (2006). Practice Parameter: Treatment of Parkinson disease with motor fluctuations and dyskinesia (an evidence-based review): [RETIRED]. *Neurology*, *66*, 983–995.

Pan, H., Mukhopadhyay, P., Rajesh, M., Patel, V., Mukhopadhyay, B., Gao, B., . . . Pacher, P. (2008). Cannabidiol Attenuates Cisplatin-Induced Nephrotoxicity by Decreasing Oxidative/Nitrosative Stress, Inflammation, and Cell Death. *J. Pharmacol. Exp. Ther. 328*, 708–714.

Pantcheva, P., Reyes, S., Hoover, J., Kaelber, S. & Borlongan, C. (2015). Treating non-motor symptoms of Parkinson's disease with transplantation of stem cells. *Expert Rev. Neurother. 15*, 1231–1240.

Pertwee, R. G. (2006). Cannabinoid pharmacology: the first 66 years. *Br. J. Pharmacol. 147*(Suppl 1), S163–S171.

Pertwee, R. G. (2008). The diverse CB1and CB2receptor pharmacology of three plant cannabinoids: ∆9-tetrahydrocannabinol, cannabidiol and ∆9- tetrahydrocannabivarin. *Br. J. Pharmacol. 153*, 199–215.

Petrocellis, L., Cascio, M. & Marzo, V. (2004). The endocannabinoid system: a general view and latest additions. *Br. J. Pharmacol. 141*(5), 765–774.

Pisani, A., Fezza, F., Galati, S., Battista, N., Napolitano, S., Finazzi-Agrò, A., . . . Maccarrone, M. (2005). High endogenous cannabinoid levels in the cerebrospinal fluid of untreated Parkinson's disease patients. *Ann. Neurol. 57*, 777–779.

Pisani, V., Madeo, G., Tassone, A., Sciamanna, G., Maccarrone, M., Stanzione, P. & Pisani, A. (2010). Homeostatic changes of the endocannabinoid system in Parkinson's disease. *Mov. Disord. 26*, 216–222.

Reggio, P. (2010). Endocannabinoid Binding to the Cannabinoid Receptors: What Is Known and What Remains Unknown. *Curr. Med. Chem. 17*, 1468–1486.

Riegel, A. (2004). Independent Presynaptic and Postsynaptic Mechanisms Regulate Endocannabinoid Signaling at Multiple Synapses in the Ventral Tegmental Area. *J. Neurosci. 24*, 11070–11078.

Rodríguez de Fonseca, F., Del Arco, I., Bermudez-Silva, F. J., Bilbao, A., Cippitelli, A. & Navarro, M. (2005). The endocannabinoid system: physiology and pharmacology. *Alcohol Alcohol. 40*(1), 2–14.

Rodriguez-Oroz, M., Jahanshahi, M., Krack, P., Litvan, I., Macias, R., Bezard, E. & Obeso, J. (2009). Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. *Lancet Neurol. 8*, 1128–1139.

Romero, J., Lastres-Becker, I., de Miguel, R., Berrendero, F., Ramos, J. & Fernández-Ruiz, J. (2002). The endogenous cannabinoid system and the basal ganglia. *Pharmacol. Ther. 95*, 137–152.

Ross, R. (2009). The enigmatic pharmacology of GPR55. *Trends Pharmacol. Sci. 30*, 156–163.

Russo, E. (2011). Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *Br. J. Pharmacol. 163*, 1344–1364.

Sagredo, O., García-Arencibia, M., de Lago, E., Finetti, S., Decio, A. & Fernández-Ruiz, J. (2007). Cannabinoids and Neuroprotection in Basal Ganglia Disorders. *Mol. Neurobiol. 36*, 82–91.

Sagredo, O., Pazos, M., Satta, V., Ramos, J., Pertwee, R. G. & Fernández-Ruiz, J. (2011). Neuroprotective effects of phytocannabinoid-based medicines in experimental models of Huntington's disease. *J. Neurosci. Res. 89*, 1509–1518.

Sam, A. H., Salem, V. & Ghatei, M. A. (2011). Rimonabant: From RIO to Ban. *J. Obes. 2011*, 432607.

Santos, N., Martins, N., Sisti, F., Fernandes, L., Ferreira, R., Queiroz, R. & Santos, A. (2015). The neuroprotection of cannabidiol against MPP + -induced toxicity in PC12 cells involves trkA receptors, upregulation of axonal and synaptic proteins, neuritogenesis, and might be relevant to Parkinson's disease. *Toxicol. Vitr. 30*, 231–240.

Sayd, A., Anton, M., Alen, F., Caso, J., Pavon, J., Leza, J., . . . Orio, L. (2014). Systemic Administration of Oleoylethanolamide Protects from Neuroinflammation and Anhedonia Induced by LPS in Rats. *Int. J. Neuropsychopharmacol. 18*(6), pii: pyu111.

Segovia, G., Mora, F., Crossman, A. & Brotchie, J. (2003). Effects of CB1 cannabinoid receptor modulating compounds on the hyperkinesia induced by high-dose levodopa in the reserpine-treated rat model of Parkinson's disease. *Mov. Disord. 18*, 138–149.

Sieradzan, K., Fox, S., Hill, M., Dick, J., Crossman, A. & Brotchie, J. (2001). Cannabinoids reduce levodopa-induced dyskinesia in Parkinson's disease: A pilot study. *Neurology, 57*, 2108–2111.

Solinas, M., Justinova, Z., Golberg, S. & Tanda, G. (2006). Anandamide administration alone and after inhibition of fatty acid amide hydrolase (FAAH) increases dopamine levels in the nucleus accumbens shell in rats. *J. Neurochem. 98*, 408–419.

Stampanoni Bassi, M., Sancesario, A., Morace, R., Centonze, D. & Iezzi, E. (2017). Cannabinoids in Parkinson's Disease. *Cannabis Cannabinoid Res. 2*, 21–29.

Stocchi, F., Abbruzzese, G., Ceravolo, R., Cortelli, P., D'Amelio, M., De Pandis, M., . . . Zappia, M. (2014). Prevalence of fatigue in Parkinson disease and its clinical correlates. *Neurology, 83*, 215–220.

Tanda, G. (1997). Cannabinoid and Heroin Activation of Mesolimbic Dopamine Transmission by a Common μ1 Opioid Receptor Mechanism. *Science (80-.). 276*, 2048–2050.

Tandberg, E. (1996). The Occurrence of Depression in Parkinson's Disease. *Arch. Neurol. 53*(2), 175–179.

Thomas, B. & Beal, M. (2007). Parkinson's disease. *Hum. Mol. Genet. 16*, R183–R194.

Tóth, A., Blumberg, P. & Boczán, J. (2009). Anandamide and the Vanilloid Receptor (TRPV1). *Vitam. Horm. 81*, 389–419.

van der Stelt, M. & Di Marzo, V. (2003). The endocannabinoid system in the basal ganglia and in the mesolimbic reward system: implications for neurological and psychiatric disorders. *Eur. J. Pharmacol. 480*, 133–150.

van der Stelt, M., Veldhuis, W., Maccarrone, M., Bär, P., Nicolay, K., Veldink, G., . . . Vliegenthart, J. (2002). Acute Neuronal Injury, Excitotoxicity, and the Endocannabinoid System. *Mol. Neurobiol. 26*, 317–346.

Van Sickle, M. (2005). Identification and Functional Characterization of Brainstem Cannabinoid CB2 Receptors. *Science (80-.). 310*, 329–332.

Velseboer, D., de Haan, R., Wieling, W., Goldstein, D. & de Bie, R. (2011). Prevalence of orthostatic hypotension in Parkinson's disease: A systematic review and meta-analysis. *Parkinsonism Relat. Disord. 17*, 724–729.

Venderová, K., Ružička, E., Vořísek, V. & Višňovský, P. (2004). Survey on cannabis use in Parkinson's disease: Subjective improvement of motor symptoms. *Mov. Disord. 19*, 1102–1106.

Xu, J., Kochanek, K., Murphy, S. & Tejada-Vera, B. (2010). Deaths: final data for 2007. *Natl. Vital Stat. Reports*, *58*(19), 1–19.

Yust-Katz, S., Hershkovitz, R., Gurevich, T. & Djaldetti, R. (2017). Pain in Extrapyramidal Neurodegenerative Diseases. *Clin. J. Pain*, *33*, 635–639.

Zeissler, M., Hanemann, C., Zajicek, J. & Carroll, C. (2012). FAAH Inhibition is Protective in a Cell Culture Model of Parkinson's Disease. *J. Neurol. Neurosurg. Psychiatry*, *83*, A15.

News Article

Malta's Science and Arts Festival Focuses on the Science of YOU

Edward Duca[*1]

[1]*Centre for Entrepreneurship and Business Incubation, University of Malta, Msida, Malta*

Malta organises one of the largest celebrations of researchers around Europe. As part of the EU-wide European Researchers Night (ERN; Marie Skłodowska-Curie Actions of the Horizon 2020 Program (H2020, 2014–2020)), Science in the City is Malta's platform to engage citizens with the latest research and scientific projects happening around the Maltese Islands.

From 2018–2019, 55 ERN projects ran in 371 cities across Europe, Turkey, Israel and other countries. During the 2018 ERN, over 1.5 million visitors attended. The Maltese ERN ranked 7th from 128 applications and began in 2012, attracting just over 10,000 visitors. It has now grown into a science and arts festival that attracts 30,000 people, around 6% of the Maltese Islands' population.

The festival uses an approach that combines science with the Arts, including research from every academic discipline, from history and accountancy, to medicine and ICT. This reflects the worldwide shift from using a STEM (science, technology, engineering and maths) approach in education to further include the Arts, now recognised as a STEAM approach. In line with several aims of Science in the City, the combination of humanities and STEM disciplines has been correlated with an increased interest and passion in scientific subjects, leadind to STEM careers (Land, 2013; Reiter, 2017; Segarra, Natalizio, Falkenberg, Pulford & Holmes, 2018).

The festival uses an open and inclusive approach to involve and provide a platform for as many researchers as possible. Open calls (followed by a rigorous evaluation) are used to attract the most creative ideas from both artists and researchers in order to engage citizens. The festival also approaches many research groups, departments, faculties, companies, NGOs, and government institutions, resulting in 94 partners being involved in the festival in 2018. This resulted in over 600 personnel (including 400 University of Malta, Junior College and MCAST students) participating in the festival, around 100 of which were active researchers.

The festival encourages everyone to participate, provided that they use a creative inquiry-based science education approach which is hands on and engaging (Murphy & Beggs, 2003; Rocard et al., 2007; Chappell, Hetherington, Ruck Keene, Slade & Cukorova, 2016). This approach fosters dialogue between experts and participants, while building participants' confidence in the subject and inspiring further interest. It can be challenging to implement but maintains the public appeal and festival quality.

The festival is rigorously evaluated every year to inform organisers in order to enable them to continuously improve the festival. A survey methodology is used that has shown numerous important findings. The festival is seen as a largely family friendly event; in 2016 more than 50% of people surveyed attended with their family. However, a large percentage of the festival is attended by 16–25 year olds (teenagers and students), thus the festival is now designing tailored events to cater for them. In the same year, 42% of participants attended for the first time, showing that Science in the City attracts new audiences. Despite these findings, only 54.6% of festival volunteers initiated conversations with participants and 14.6% used difficult jargon, hence the festival has now focused volunteer training in order to address these issues. Several other findings are in preparation for publication in order to aid other festivals to use similar evaluation processes.

In 2019, the festival theme, 'The Science of YOU', focussed on the science of human development. This concept aimed to demonstrate how scientific research is geared towards human advancement. Science creates knowledge and solves problems, giving us the power to make decisions and shaping our actions. It defines who we are by giving us options, affecting the way we think about ourselves, others and the environment.

Correspondence to: Edward Duca (edward.duca@um.edu.mt)

Table of Contents

The festival covered topics from personalised genomic medicine and bio banking, to Artificial Intelligence and the Environment. One of the main artworks was *Xewk*, a project led by Prof. Ruben Paul Borg with a team of architects (ICE & MGPEI (Malta Group of Professional Engineering Institutions)), and Student Chapter University of Malta from the Faculty for the Built Environment, who used computer programmes to create an emergent design based on parametric modelling techniques. The team created a light strong design inspired by Maltese trees and spines, which was around 2 m in length and width, 1 m high and only weighed 25 kg. The artwork was made out of high-performance fibre-reinforced cement, which is a sustainable solution that can be implemented in current building designs (Borg, Bondin & Mangion, 2019).

The other main artwork *Spherical* was created by Louis Briffa. Five illuminated glass spheres reflected the topics of Artificial Intelligence, Biotechnology, Cloud Computing, Blockchain and Nanotechnology. These subjects are intensely researched at the University of Malta and celebrate local achievements.

The festival also featured *Creative Jam*, a collaboration between Diccon Cooper's band Stretta Swing and The Edward De Bono Institute (University of Malta), who used creativity tests to analyse if the techniques behind jazz improvisation affected a listener's creativity. Furthermore, to raise awareness about neurodiversity, Moveo Dance Company performed a dance called *Diversely Typical*. Neurodiversity is the concept that neurological disorders such as ADHD and autism are just another way of behaving, and that people living with these conditions need to be accepted and included in an environment that facilitates them to live fulfilling lives within society. The festival also worked with NGO Green House to create *Hanging On*, an interactive performance that allows participants to choose whether or not to save bats in Malta through research. Alongside these unique activities, the festival includes many science shows, hands-on experiments and interactive technology (digital games, VR and others) from every major education institution on the Maltese Islands.

The festival now attracts science communicators, researchers, performers and artists from around 10 different countries.

All of these events are attempting to attract diverse groups of people within society. It is evident that using a one size fits all approach to marketing and communications does not work, therefore, a targeted approach is used to attract large audience numbers effectively, in order to achieve the festival's aims. These include, increasing the viability of researchers in Malta and showing how their work impacts everyday life. In addition, highlighting that research is a creative profession for all genders and sexualities, that it is fun and inspiring to encourage STEM career uptake, and that scientists and artists can work together to enhance both professions

for the social and economic progress of a country. Furthermore, achievement of these ambitious goals will require multiple initiatives, time, partners, and significant resources.

Science in the City is part of the EU-wide celebration-European Researchers' Night. It is funded by the Marie Skłodowska-Curie Actions of the Horizon 2020 Program of the EU, as well as a number of corporate sponsors, and is recognized as a Festival by 'Europe for Festivals and Festivals for Europe' (EFFE).

The consortium is led by the University of Malta, Malta Chamber of Scientists and the University's Research Trust (RIDT), in partnership with Parliamentary Secretary for Financial Services, Digital Economy and Innovation, Ministry of Education and Employment, MCAST, Esplora, JUGS Ltd, Studio 7, BPC International, GSD Marketing Ltd, Aquabiotech Ltd, MEUSAC, PBS, Spazju Kreattiv, Valletta Local Council, Malta Council for the Voluntary Sector, Small Initiatives Support Scheme, More or Less Theatre, Transport Malta, Kreattiv and Creative Community funds, and Arts Council Malta.

For more info about the festival see www.scienceinthecity.org.mt and www.facebook.com/ScienceInTheCityMalta.

To be part of the festival email: info@scienceinthecity.org.mt or edward.duca@um.edu.mt.

References

Borg, R. P., Bondin, W. & Mangion, F. (2019). *Emergent Design for Complex structures based on High-Performance Fibre-Reinforced Cement Based Materials*. Manuscript Submitted for Publication.

Chappell, K., Hetherington, L., Ruck Keene, H., Slade, C. & Cukorova, M. (2016). *D2.1 The features of inquiry learning: theory, research and practice*. Retreived from http://creations-project.eu/wp-content/uploads/2016/09/Download-3.pdf. CREATIONS EU project.

Land, M. H. (2013). Full STEAM Ahead: The Benefits of Integrating the Arts Into STEM. *Procedia Comput. Sci. 20*, 547–552.

Murphy, C. & Beggs, J. (2003). Children's perceptions of school science. *Sch. Sci. Rev. 84*(308), 109–116.

Reiter, C. M. (2017). 21st Century Education: The Importance of the Humanities in Primary Education in the Age of STEM. *Senior Theses and Capstone Projects*, 65.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H. & Hemmo, V. (2007). *Science education now: A renewed pedagogy for the future of Europe*. Brussels: European Commission.

Segarra, V. A., Natalizio, B., Falkenberg, C. V., Pulford, S. & Holmes, R. M. (2018). Using the Arts to Train Well-Rounded and Creative Scientists. *J. Microbiol. Biol. Educ. 19*(1), 19.1.53.