PECORA 21 • ISRSE 38

Continuous Monitoring of Our Changing Planet:
From Sensors to Decisions

Call for Abstracts

October 6-11, 2019
Baltimore, MD • Marriott Waterfront

https://www.asprs.org/event/pecora21_isrse38
The Pecora 21/ISRSE 38 Conference will be organized around four session themes spanning the Earth observations and continuous monitoring continuum. Presentations are being sought on diverse science, technology, and applications of remote sensing to understand and sustainably manage the Earth’s environment and natural resources. We encourage contributions along the full value chain of Earth observation, from fundamental research on Earth system processes to operational applications, innovative techniques and future missions, as well as international programmes and coordination.

Submitted abstracts (max. 300 words) will be organized, though not exclusively, along four broad thematic areas described below:

- Understanding the Earth through continuous monitoring
- Societal benefits and empowering decision making
- Technical advances in monitoring using Earth observations
- Envisioning the future of global monitoring

Abstracts addressing the conference themes may be submitted for general consideration, or be considered for inclusion in a proposed special presentation sessions. Additional session proposals will also be accepted during the Call for Abstracts.

**Presentation Types**

- **Standard presentation**: Long-format research talk (15 minutes).
- **Special presentation session**: Inclusion in one of the sessions listed to the right. For more information on the special presentations, visit http://pecora.asprs.org/.
- **Short presentation**: Lightning-style research talk (3 minutes)
- **Short visualization**: Lightning-style talk focused on data visualization, e.g. map products, dashboards, interactive plots, cartographic tools (3 minutes)
- **Poster**: Poster presentation. May be considered for general poster sessions or an illustrated poster sessions where posters will be grouped by topic or theme and presenters will have 1-2 minutes to introduce their poster to attendees.
- **Workshops**: 2-hour and 4-hour pre-conference workshops.

Awards will be offered in each category above, including awards for best young professional/student talk and poster. To be eligible, presenters must request to be considered for judged awards during the abstract submission process.

**Timeline** Abstracts Due—February 25, 2019

**Special Presentation Sessions**

- **SP1**—Open Data Cube: A New Data Technology for Enhancing the use of Satellite Data to Address Sustainable Development Goals
- **SP2**—An Overview of the current Analysis Ready Data products, tools, applications and impacts
- **SP3**—New Technology and Techniques to Increase Scientific and Applications Access to Satellite Earth Observations
- **SP4**—Lidar Vegetation Canopy Metrics—Towards Developing Standards
- **SP5**—High-resolution Land Cover using NAIP
- **SP6**—SAR for Agriculture and Perspective Applications
- **SP7**—Water Colour: The Canadian Perspective
- **SP8**—Applications of NASA Earth Observations for Local Decision Making: 20 Years of the NASA DEVELOP Program
- **SP9**—How No-cost Landsat Data is Reshaping College-level Remote Sensing Courses
- **SP10**—Societal Benefits of Earth Observations in Natural Resource Management Decision Making
- **SP11**—So What, Who Cares: Linking Natural and Social Science to Understand Societal Impact and Improve Decision Making
- **SP12**—Importance of System Calibration and Data Quality on Earth Observation
- **SP13**—UAS, Changing the Future of Remote Sensing
- **SP14**—Satellite Interoperability
- **SP15**—Land Imaging Capabilities and User Needs
- **SP16**—Air Quality Monitoring with Earth Observations for Enhanced Decision Making and Regulatory Support
- **SP17**—Transitional by Nature: Leveraging Remote Sensing Technology for Continuous Monitoring of Dynamic Wetland Ecosystems
- **SP18**—Connecting People and Pixels through Citizen Science to Enhance Global Monitoring
- **SP19**—Open Civil Applications Committee Meetings
- **SP20**—The Next Generation of the Landsat Archive
- **SP21**—Space Agency’s Outlook
- **SP22**—NASA Harvest and Other Recent Advances in Remote Sensing of Agricultural Applications and Food Security
- **SP23**—Sustainable Land Imaging and the Future of Moderate-Resolution Land Observation
- **SP24**—Geospatial Fusion: Observations, Features, Decisions
- **SP25**—The Challenges of Integration for Arctic Monitoring
- **SP26**—Remote Sensing Applications for Water Resources Management, Including Droughts, Floods and Associated Water Cycle Extremes
- **SP27**—Communicating Science Across the Earth Observation Life Cycle
ANNOUNCEMENTS

Dewberry has promoted Phil Thiel to executive vice president. In his new role he will lead Dewberry’s federal market strategy and business development. A consulting industry executive with nearly 35 years of experience, Thiel will direct and broaden Dewberry’s engagement with federal agency clients throughout the enterprise.

Thiel has been with Dewberry for 17 years and will continue to lead the firm’s national Geospatial and Technology Services practice. Under his direction, Dewberry holds numerous federal contracts involving state-of-the-art mapping, remote sensing, GIS, and IT services. Federal clients include the U.S. Army Corps of Engineers, the U.S. Geological Survey, the Federal Emergency Management Agency, the Department of Homeland Security, the National Reconnaissance Office, National Geospatial-Intelligence Agency, and the National Oceanic and Atmospheric Administration. To learn more, visit www.dewberry.com.

AXIS GeoAviation, LLC (AGA) is pleased to announce the newest addition to its fleet; a Vulcan Air P68C Twin Engine aircraft. The additional plane allows AGA to support future aerial surveying and mapping projects particularly in the Southeast US, Caribbean and also nationwide. AGA’s sister company, AXIS GeoSpatial, LLC (AGS), recently acquired Mapping Resource Group, Inc. of Florida with a significant customer base in the Southeast and the Mid-Atlantic regions of the US and the Caribbean. The Vulcan P68C provides the additional capacity to serve AXIS’ clients, as well as versatility in that it is fast enough for aerial imagery collection yet can be flown slower and at lower altitudes to capture higher density lidar data—additional aircraft in AGA’s fleet include a Cessna 206H and a Piper Navajo.

In addition to its growing fleet, AGA has expanded its capability to collect lidar data through the purchase of two RIEGL VQ 1560i dual-channel airborne mapping sensors. AGA invested in the additional lidar sensors to meet the increasing client demand for higher density datasets. As experts in hi-definition lidar acquisition, AGA offers survey data collection, processing and classification for design scale mapping of airports, college campuses, environmental remediation, highways, railroads, pipelines, corridors, and other design and maintenance projects. Currently, AXIS GeoAviation is the only company on the East Coast of the United States to own two RIEGL VQ 1560i sensors.

The new acquisitions bring the total assets owned by AGA to three fixed-wing manned aircraft, two Unmanned Aircraft Systems (UAS), one Vexcel UltraCam Eagle system, one RIEGL LMS-Q1560 lidar sensor, and two RIEGL VQ 1560i lidar sensors. AGA currently employs four pilots and eight crew members operating out of its Easton, Maryland and Ormond Beach, Florida hangars.

GeoCue Group (via its wholly owned AirGon subsidiary) has completed the integration of the new DJI Phantom 4 Pro RTK (P4R) into our widely used AirGon Sensor Processing Suite (ASPSuite). ASPSuite is used as the post-processing solution for our Loki direct geopositioning system for DJI and other manufacturer’s drones.

ASPSuite enables integration of the P4R with third party L1/ L2 GNSS base stations such as systems from Septentrio, Leica, Trimble, Tersus, TopCon, CHC and others in a high accuracy Post-Process Kinematic (PPK) workflow.

Currently, the DJI D-RTK-2 base station (optionally available) for the P4R can only be used in RTK mode and then only if it is being sited on a known location. The D-RTK-2 does not allow access to an observation file, preventing it from being stationed using an online positioning service such as OPUS, AUSPOS, Canadian Geodetic Survey services and so forth. An additional consideration in our integration into ASPSuite was that professional surveyors already have a survey kit that they need incorporated into this workflow.

GeoCue is offering camera calibration services for the P4R for those customers who wish to do minimal or control-free high accuracy mapping projects (the DJI “calibration” is not a rigorous photogrammetric calibration). In a recent test of a GeoCue-calibrated P4R using an OPUS positioned base station and PPK processing with ASPSuite, we achieved about 4 cm horizontal and 5 cm vertical network accuracy (RMSE) with no ground control points. While not quite as accurate as a Loki solution, these results are remarkable for a low-cost drone.

We have partnered with hundreds of engineering, mining and industrial firms as well as government agencies, providing high accuracy drone mapping solutions. For additional information and system quotations, please contact us via email at sales@airgon.com.

CALENDAR

- 3-7 April, AAG 2019 Annual Meeting, Washington, DC. For more information, visit http://annualmeeting.aag.org/.
- 3-5 May, 5th International Conference on Geographical Information Systems Theory, Applications and Management, Heraklion, Crete, Greece. For more information, visit http://www.gistam.org.
- 11-15 August, SPIE— Imaging Spectrometry XXIII, San Diego, California. For more information, visit spie.org/OP423.
See the Cover Description on Page 80
The eastern United States is well known for its widespread show of vibrant foliage each autumn. But western states grow more than just evergreen pines and palms; they also display some spectacular fall color. These images show deciduous trees in northern Utah painting the mountains in shades of red, orange, yellow, and purple.

The Operational Land Imager (OLI) on Landsat-8 acquired the images on 27 September 2018, as daylight hours shortened and temperatures cooled. The cover image, draped over elevation data from the Shuttle Radar Topography Mission (SRTM), shows Ogden Valley in the foreground. The valley is nestled amid part of the Wasatch Range and is accessible by three scenic roads. The Ogden River Scenic Byway, for example, meanders for about 16 kilometers (10 miles) through the Ogden Canyon, connecting the city of North Ogden with Ogden Valley.

The image above is a nadir (straight-down) view showing Ogden city at the base of the colorful mountain slopes. Hikers can reach the 2448-meter (8031-foot) summit of Lewis Peak via a network of trails in the Uinta-Wasatch-Cache National Forest. It is not the tallest mountain in the vicinity, but it offers views of the nearby Willard and Ben Lomond peaks. Along the route, you can find fir trees mingling with wildflowers in July and August and with colorful deciduous trees by September. The U.S. Forest Service noted about the Lewis Peak trail: “There may be no better place to view the autumn colors than along this route.”

Colorful leaves adorn a variety of tree species in Utah’s forests, including canyon maples, quaking aspens, scrub oaks, and Douglas Hawthorn. The timing of the color change in Utah this year was roughly on schedule, according to news reports. Other parts of the country, from Maine to the mid-Atlantic, saw color appear a few weeks late due daily low temperatures that have been warmer than usual. The cooler temperatures and longer nights that arrive with fall tell trees to stop producing chlorophyll. In the absence of this green pigment, the yellow, orange, and red pigments left behind become visible. Weather factors—such as too much or too little water, and the timing of that water—also affect the intensity and timing of color. To see the full images, visit https://landsat.visibleearth.nasa.gov/view.php?id=92880.

Geospatial Certification: Educating Yourself on the Benefits

As the geospatial community grows and opportunities for jobs expand, organizations will search for employees with relevant certifications to fill positions in their respective organizations. For an individual, achieving the full benefit of any geospatial certification will require an understanding of the current geospatial landscape as well as determining personal and professional goals. According to the website Statista.com, market revenue of the global geospatial industry in GIS/Spatial Analytics alone will grow by approximately $10 billion per year between 2018 and 2020. The GeoBuiz 2018 Report on Geospatial Industry Outlook and Readiness Index outlined that the geospatial industry market is witnessing unprecedented growth in all geographies with high double-digit growth in the Asia Pacific, Middle East, Africa and South African regions. During 2013 to 2017, the market grew at an estimated 11.5%, and is forecast to grow around 13.6% between 2017-2020. The geospatial workforce is itself diverse, multidisciplined, and multifaceted, so how does a geospatial professional continue to attain skill sets to demonstrate their value in this booming industry?

Is Certification Right For You?

In 2003, the authors of “Building the Geospatial Workforce” understood that “the growth of this market demands support of the education, training, and development of geospatial professionals and specialists.”

Past debates with respect to certification have ranged from identifying an optimal framework for a path to certification to the criteria that defines an academic versus a professional certification. Those are all worthy topics of research and debate, but what most geospatial practitioners want to know is what is the best certification for me, because is it both a personal and professional goal that requires the dedication of time and money to achieve. There are a few questions to ask when thinking about pursuing a geospatial certification: Is the certification right for your personal and professional goals? Do you need the certification for a particular position, for acquiring a new skill, or for continuing education requirements? Does the certification further your development or does it lock you into a specific discipline? Does the certification expand your opportunities? Is it accepted as a rigorous achievement? Is it a one-time certification or are there continuation requirements and costs? Finally, is it the appropriate certification for your current career level or the level you wish to attain?

The Value of Certification

Over a decade ago, on a training and education program at the National Geospatial-Intelligence Agency (NGA) was developed with the goal of workforce certification, by the Division Chief of the Advanced Geospatial Intelligence Training Program at NGA. The goal was to develop tracks for various tradecrafts creating courses to allow basic, intermediate, and advanced levels of attainment. The central theme was that true certification required education, experience, and training interwoven into a credible program managed by a board of multidisciplinary experts. It was the precursor to the current certification program for the National System for Geospatial Intelligence (NSG), which is managed by NGA. During that time NGA was working towards implementing baseline education and training requirements, and integrating the new requirements with those which had been in the workforce for several years while transferring legacy skills sets to a more digital and software-centric analysis and production environment. Individual as well as team and organizational certification requirements were developed to get to a true and acceptable workforce certification program. Overall, we found that certification must be academically credible, workforce centric, and broadly accepted to hold any real value across the geospatial community.

Which Certification Is Right For You?

Once you determine that “the juice worth the squeeze” in terms of time and cost, the next step is researching which certification is right for you. Is the certification too narrow or too broad for your ultimate goal? Is it focused on a specific discipline? Does the certification expand your opportunities? Is it accepted as a rigorous achievement? Is it a one-time certification or are there continuation requirements and costs? Finally, is it the appropriate certification for your current career level or the level you wish to attain?
ic hardware or software, or is it applicable across multiple geospatial platforms? For example, would an IT certification actually benefit you more as a GIS Manager than a geospatial certification? Would an equipment certification benefit you more as a surveyor than one focused on mostly GIS topics?

Is the certification accredited by a recognized and reputable organization with clout across the geospatial community? Does the certification “have teeth” in that there an enforceable, punitive policy is in place to ensure the long-term validity of the certification? For example, if you have not been personally certified as a GISP by the GIS Certification Institute (GISCI), then you cannot legally use the GISP designation, either as part of your signature or on your resume. Any person found to have used the GISP designation without having been previously granted use of that credential by the GIS Certification Institute will be subject to legal action under federal copyright and trademark code.

Is it organizationally-based, such as the GEOINT Professional Certification (GPC), an NSG program with fundamental and tradecraft certifications? Is the certification only internally recognized or is it internationally recognized? If there is an exam and a recertification process, how often do you have to recertify and how much will it cost? To sort out these types of questions, we discuss several certifications in the sidebar to the right.

To achieve certification under these programs, one must demonstrate a comprehensive understanding of subject matter (GISCI), a solid understanding of various aspects of an occupational specialty (USGIF), and validate a specific technical understanding within an occupational specialty (ASPRS). For internally-based certification (NGA), each certification is a skillset for a tradecraft resident within the organization, and although not specifically designed for use outside of the organization, what is learned may be applicable to skills required in the greater geospatial community.

Is One Umbrella Certification Available?
Developing and agreeing to a basic competency model remains a challenge two decades into the 21st century, but the idea of creating and maintaining a relevant certification is alive and well. In 2009, a study found that “Geographic In-

The American Society for Photogrammetry and Remote Sensing (ASPRS) offers specialty certifications related to technical occupations commonly found in the geospatial community. The ASPRS Certification Program is voluntary and open to all qualified individuals, whether or not they are members of the American Society for Photogrammetry and Remote Sensing. ASPRS has administered a robust certification process for over four decades that continues to revise and add new certifications as innovative technologies are adapted into the geospatial industry. In 2017, Mike Renslow provided a detailed understanding of the ASPRS certification program in a recent SectorInsight.edu article. He wrote that the ASPRS certification is official recognition by one’s colleagues and peers that an individual has demonstrated professional integrity and competence in his or her field. As such, the ASPRS voluntary certification program is considered “specialty certification” and not a substitute for licensure. Prior to 1999, the ASPRS certification process consisted of a peer review of each application and required four confidential references. In 1997, the ASPRS Board of Directors approved a modification to the Certification Program that took effect on January 1, 1998 that required applicants to pass a written examination. In 2013, the ASPRS professional and technologist certification programs each received accreditation from the Council of Engineering and Scientific Specialty Boards (CESB).2
USGIF's Universal GEOINT Certification Program provides a foundation on which GEOINT professionals can certify the knowledge, skills, and abilities necessary for successfully meeting the duties and responsibilities within the multi-faceted GEOINT tradecraft. This program is available internationally to GEOINT practitioners across industry, military, academia, and federal, state, and local governments. The goal of the Certified GEOINT Professional (CGP™) Program is to provide a foundation on which GEOINT professionals can certify the knowledge, skills, and abilities necessary for successfully meeting the duties and responsibilities within the multi-faceted GEOINT tradecraft. GEOINT practitioners who meet the baseline certification requirements and successfully pass the requisite exams earn the right to use the USGIF professional designations. The Certified GEOINT Professional (CGP™) Program is accredited by the National Commission for Certifying Agencies (NCCA), under the Institute for Credentialing Excellence (ICE).

The Geospatial Intelligence (GEOINT) Professional Certification (GPC) program is a National System for GEOINT (NSG) effort to professionally certify the NSG workforce within the GEOINT realm. The GPC, as well as other Defense Intelligence Enterprise programs has come about based on guidance issued by the U.S. Under-Secretary of Defense for Intelligence (USD-I) to certify the Defense Intelligence Enterprise (DIE) workforce. A common perception is the GPC is a National Geospatial-intelligence Agency (NGA) program, but it is actually a full NSG program. It was initially established at NGA Campus East because the Director of NGA also being the NSG Functional Manager; meaning that the NGA Director has the responsibility of establishing GEOINT policy for partner agencies within the NSG. As the GPC is an NSG program, eligibility to participate is limited to U.S. Government civilians, U.S. Government contractors, and military service members who work in the GEOINT field. While most GEOINT civilians and contractors qualify to participate in the GPC-F, in order to participate in any of the PL II certifications there must be an association with that specific tradecraft.

The GPC is a GEOINT tradecraft certification program and is accredited by the National Commission for Certifying Agencies. There are currently two levels of GEOINT certification: Proficiency Level I (PL I) and Proficiency Level II (PL II). PL I, or GPC - Fundamentals (GPC-F), focuses on GEOINT doctrine and policy as well as fundamental tradecraft knowledge. PL II is currently comprised of ten, separate GEOINT tradecraft areas, ranging from Geospatial Analysis to Human Geography, with each based on well-established essential bodies of knowledge. In order to qualify for a GEOINT certification a participant must be in good standing in their tradecraft area and have a minimum of 1-year experience. Recertification requires the participant to earn at least 100 professional development units (PDU’s) over a three-year period.

The PL II tradecraft areas of geospatial analysis, imagery science, and imagery analysis require a completed Professional Qualifications Standards (PQS) workbook, which serve as a demonstration of the minimum skills required for the tradecraft, prior to being permitted to take those exams. While an individual can take a PL II exam, provided that they qualify under that tradecraft area, they will not be conferred as a certified GEOINT professional until the GPC-F exam is passed. As of October 2018, over 15,000 GPC-F exams have been administered to personnel in the GEOINT workforce, with over 10,000 personnel passing to become GEOINT Professionals. Additionally, over 3,700 PL II certifications have been earned.

GISCI offers an industry-wide, internationally-recognized, software-agnostic certification available to all geospatial professionals. GISCI's Geospatial Core Technical Knowledge Exam is software-agnostic, based upon a job analysis from a four-year experience level, informed by the GIS&T Body of Knowledge, guided by the Geospatial Technology Competency Model (GTCM), and centered upon the six key knowledge areas. It was developed by GISP's for the GISP Certification, which is open to any geospatial professional with the requisite background to take the exam.
Understanding the diversity of the geospatial workforce and natural resistance is at the heart of why one ubiquitous certification is difficult to create and even more difficult to gain universal acceptance. For example, the competencies of a private sector geospatial professional might not align with competencies required in the private sector. The public sector is not traditionally focused on market growth or stock value, and crucial employee competencies may be inefficient with respect to similar job requirements in the private sector. For example, a photogrammetrist in an intelligence agency might work various forms of digital data and require an understanding of flight patterns of many aerial platforms and several digital formats, but a photogrammetrist in the private sector may work with one specific platform and use both hardcopy and softcopy formats. There are certain skillsets that all photogrammetrists should possess, and those lend themselves to a broad-based certification, but it would be difficult to achieve (and keep up-to-date) one single comprehensive certification. In the future, it might be plausible to develop a series of certifications that can be attained though achieving various levels of proficiency, with a goal of being internationally recognized and cost-effective to develop, maintain, and administer.

The Future of Certification: Opportunities
The various geospatial certifications bring numerous opportunities for the future. Of note, certificates have shown the largest growth among college credentials over the past 30 years, according to a 2013 study by the Georgetown University Center on Education and the Workforce. The future is expected to bring additional collaboration between the public and private sectors, which can improve the long-term certification process. One area where collaboration could be where the introduction of new technologies doesn’t immediately put a certification into legacy status, but actually extends it to a new phase of credentialing. Certification entities develop strategies for students to build skills in existing educational programs to achieve specific certifications while in school. Lowering the cost and barriers to entry while building long-term loyalty to a brand. Colleges and university curriculum developers should explore options to implement certifications at a discounted cost by partnering with private businesses. Companies could also work with community colleges to develop programs that can be taught in house to lower personnel and overhead costs inherently built into the cost of any certification. This allows the members of the geospatial community to determine, but not mandate, pathways to certification, and encourage both the public and private sectors to fill the training, education, and experience required. Any entity can create and maintain a certification, as long as it meets the prescribed requirements outlined by the accrediting body. This allows a ready-made workforce with specific skill sets needed in the geospatial industry. It also allows companies to preserve their precious training dollars, focus on operational requirements, and support relevant certifications that enable the professional growth and long-term viability of the geospatial workforce.

Authors
Jared Ware, GISP, CGP-G, is a Senior GIS Manager for the Texas Railroad Commission. Previously he served as an Assistant Professor with the Geospatial information Science Program in the Department of Geography and Environmental Engineering at the United States Military Academy. He is a retired United States Army Engineer officer and has presented at previous ASPRS Conferences on the topics of GPS and lidar.

LTC Merlin F. Anderson, GISP (GPC-F, GA-II, CA-II) LTC, EN, serves as the Geospatial Plans & Policy Officer on the Army Support Team at the National Geospatial-Intelligence Agency in Springfield, VA. He was commissioned into the Regular Army as a Second Lieutenant in the Engineer Regiment in 2001 through The United States Military Academy, at West Point, New York. He is both a certified GEOINT Professional (GPC-F, GA-II, and CA-II) and GIS-Professional.

4 Carnevale, A. P., Rose, S. J., & Hanson, A. R. (2013). Certificates: Gateway to gainful employment and college degrees. Georgetown University Center on Education and the Workforce.
This is the second edition of a major textbook on lidar, covering the hardware, software and applications of terrestrial, mobile, airborne and spaceborne systems in depth. The editors are well known, highly respected academics from Purdue University (Shan) and The Ohio State University (Toth), who have presided over 32 talented contributors, drawn from industry, government and academia in multiple countries. The publisher enumerated the success of the first edition in support of its decision to go for a second: “With over 1900 hardback copies sold worldwide and thousands of subscriptions’ downloads, the first edition reached students and professionals in [multiple] fields.” Further motivation was the desire to include developments since the first edition, especially UAV-borne and commercial versions of single-photon and Geiger-mode systems, as well as software developments, some of which grew from deeper involvement in lidar of the remote sensing and computer vision communities.

The book begins with an introduction to laser ranging, followed by two hefty chapters covering hardware aspects of terrestrial, mobile, airborne and spaceborne systems. The fastidiousness with which the information has been collected and taxonomies developed to put the systems in perspective for the reader is apparent. Nevertheless, the stream of short sections describing systems produced, to around the end of 2016, could perhaps have been supplemented by graphics showing how some of the main system parameters have evolved. Chapter 4, lidar systems and calibration, explains full systems and the issues involved with registration and calibration. Chapter 5-7 are theoretical treatments of: pulsed laser ranging techniques, including signal strength, waveform digitization, discrete returns and photon counting; the georeferencing component; and full-waveform analysis. Chapter 12, forest inventory, is applications-oriented. Chapter 13 is an important discussion of the integration of lidar and photogrammetric data, focusing on triangulation and orthorectification. Chapter 14, feature extraction in urban areas, begins a long section on city applications, with five further chapters on: global solutions to building segmentation and reconstruction; building and road extraction; progressive modeling of 3D rooftops; automated construction of building models; and quality of extracted building models.

Users of the first edition may be interested specifically in the updates. The first three chapters have been subject to a thorough revision, to include the many new developments and the technological changes on which they depend. Chapters 7, 8, 10 and 17 have undergone significant overhauls and extensions. Chapter 15 a complete replacement of the one in the first edition. There are further updates throughout most of the book, though the preface points out that chapters 5, 6 and 14 are primarily theoretical and have not required revision. As a result, the page count has increased by 48.

*Topographic Laser Ranging and Scanning: Principles and Processing* is well produced, with clear text and well chosen, informative illustrations. The chapters have been written by experts in their fields, so not only are there overlaps, but the authors’ emphases necessarily reflect their research foci. These are hardly demerits, given the underlying didactic purpose. There are copious references at the end of every chapter and a useful, 15-page index. The book is expensive, however, as is typical for textbooks aimed at small markets. Nevertheless, teachers, students and practitioners should certainly have it on their shelves or demand it from their libraries. There are useful complementary works, especially the recently arrived third edition of *Digital Elevation Model Technologies and Applications: The DEM Users Manual* from ASPRS, which also offers the *Airborne Topographic Lidar Manual*. There are less weighty works too. The second edition of Shan and Toth, however, is a *magnum opus*, which will deservedly serve as readers’ primary source for the knowledge essential to be effective in the increasingly important and fast-changing world of lidar.
The 3rd edition of the DEM Users Manual includes 15 chapters and three appendices. References in the eBook version are hyperlinked. Chapter and appendix titles include:

1. Introduction to DEMs
2. Vertical Datums
3. Standards, Guidelines & Specifications
4. The National Elevation Dataset (NED)
5. The 3D Elevation Program (3DEP)
6. Photogrammetry
7. IfSAR
8. Airborne Topographic Lidar
9. Lidar Data Processing
10. Airborne Lidar Bathymetry
11. Sonar
12. Enabling Technologies
13. DEM User Applications
14. DEM User Requirements & Benefits
15. Quality Assessment of Elevations Data
   A. Acronyms
   B. Definitions
   C. Sample Datasets

This book is your guide to 3D elevation technologies, products and applications. It will guide you through the inception and implementation of the U.S. Geological Survey’s (USGS) 3D Elevation Program (3DEP) to provide not just bare earth DEMs, but a full suite of 3D elevation products using Quality Levels (QLs) that are standardized and consistent across the U.S. and territories. The 3DEP is based on the National Enhanced Elevation Assessment (NEEA) which evaluated 602 different mission-critical requirements for and benefits from enhanced elevation data of various QLs for 34 Federal agencies, all 50 states (with local and Tribal input), and 13 non-governmental organizations. The NEEA documented the highest Return on Investment from QL2 lidar for the conterminous states, Hawaii and U.S. territories, and QL5 IfSAR for Alaska.

Chapters 1, 3, 5, 8, 9, 13, 14, and 15 are “must-read” chapters for users and providers of topographic lidar data. Chapter 8 addresses linear mode, single photon and Geiger mode lidar technologies, and Chapter 10 addresses the latest in topobathymetric lidar. The remaining chapters are either relevant to all DEM technologies or address alternative technologies including photogrammetry, IfSAR, and sonar.

As demonstrated by the figures selected for the front cover of this manual, readers will recognize the editors’ vision for the future—a 3D Nation that seamlessly merges topographic and bathymetric data from the tops of the mountains, beneath rivers and lakes, to the depths of the sea.

Co-Editors
David F. Maune, PhD, CP and Amar Nayegandhi, CP, CMS

DEM Users Manual, 3rd Ed.
Stock # 4959

<table>
<thead>
<tr>
<th></th>
<th>Hardcopy</th>
<th>eBook1</th>
</tr>
</thead>
<tbody>
<tr>
<td>List Price</td>
<td>$100</td>
<td>$85</td>
</tr>
<tr>
<td>ASPRS Member</td>
<td>$80</td>
<td>$65</td>
</tr>
<tr>
<td>Student</td>
<td>$50</td>
<td>$50</td>
</tr>
</tbody>
</table>

*Plus shipping/handling

1Coming soon

ORDER ONLINE
HTTPS://WWW.ASPRS.ORG
originally settled in the Bronze Age, the Umm an-Nar’s culture established itself near modern Abu Dhabi in the 3rd century B.C., and its influence extended to the interior of the Arabian Peninsula as well as along the coast to Oman. Some later settlements by the Greeks have been found, and in the Middle Ages most of the region was part of the Kingdom of Hormuz, which controlled trade in the Arabian Gulf. The Portuguese arrived in 1498 and stayed until 1633 until the British took control of the area. By 1820, the British had destroyed or captured all Qawasim pirate ships, imposed a General Treaty of Peace on nine Arab sheikhdoms in the area, and installed a garrison. The area was known as the Trucial Coast until 1971.

The seven emirates are Abu Dhabi (Abu Zaby), ‘Ajman, Dubai (Dubayy), Al Fujayrah, Ra’s al Khaymah, Sharjah (Ash Shariqah), and Umm al Qaywayn. The United Arab Emirates (UAE) cover an area slightly smaller than the state of Maine. Much of the interior of the UAE is desert and runs to the edge of the Empty Quarter of Saudi Arabia, the largest sand desert in the world. The northern and eastern sections are mountainous and green while the coastal areas are marked with salt flats.

The first major geodetic datum of the Arabian Gulf area was established by W.E. Browne of the Iraq Petroleum Company in 1927-1931 at the South End Base at station Nahrwan (East of Baghdad) such that: $\Phi_0 = 33^\circ 19^\prime 10.87^\prime\prime$ North, $\Lambda_0 = +44^\circ 43^\prime 25.54^\prime\prime$ East of Greenwich, and the Clarke 1880 is the ellipsoid of reference. The Sir Bani Yas Island Datum of 1933 was established by the British Royal Navy in 1933 such that: $\Phi_0 = 24^\circ 16^\prime 44.83^\prime\prime$ North, $\Lambda_0 = 52^\circ 37^\prime 17.63^\prime\prime$ East of Greenwich, and the Clarke 1880 is the ellipsoid of reference. The Nahrwan Datum of 1929 is the most prevalent coordinate system of the entire Arabian Gulf area and is still found to this day.

The seven emirates are Abu Dhabi (Abu Zaby), ‘Ajman, Dubai (Dubayy), Al Fujayrah, Ra’s al Khaymah, Sharjah (Ash Shariqah), and Umm al Qaywayn. The United Arab Emirates (UAE) cover an area slightly smaller than the state of Maine. Much of the interior of the UAE is desert and runs to the edge of the Empty Quarter of Saudi Arabia, the largest sand desert in the world. The northern and eastern sections are mountainous and green while the coastal areas are marked with salt flats.

The first major geodetic datum of the Arabian Gulf area was established by W.E. Browne of the Iraq Petroleum Company in 1927-1931 at the South End Base at station Nahrwan (East of Baghdad) such that: $\Phi_0 = 33^\circ 19^\prime 10.87^\prime\prime$ North, $\Lambda_0 = +44^\circ 43^\prime 25.54^\prime\prime$ East of Greenwich, and the Clarke 1880 is the ellipsoid of reference. The Sir Bani Yas Island Datum of 1933 was established by the British Royal Navy in 1933 such that: $\Phi_0 = 24^\circ 16^\prime 44.83^\prime\prime$ North, $\Lambda_0 = 52^\circ 37^\prime 17.63^\prime\prime$ East of Greenwich, and the Clarke 1880 is the ellipsoid of reference. The Nahrwan Datum of 1929 is the most prevalent coordinate system of the entire Arabian Gulf area and is still found to this day.

The first major geodetic datum of the Arabian Gulf area was established by W.E. Browne of the Iraq Petroleum Company in 1927-1931 at the South End Base at station Nahrwan (East of Baghdad) such that: $\Phi_0 = 33^\circ 19^\prime 10.87^\prime\prime$ North, $\Lambda_0 = +44^\circ 43^\prime 25.54^\prime\prime$ East of Greenwich, and the Clarke 1880 is the ellipsoid of reference where: $a = 6,378,300.782$ m, and $1/f = 293.4663077$. The Nahrwan Datum of 1929 is the most prevalent coordinate system of the entire Arabian Gulf area and is still found to this day.

The first major geodetic datum of the Arabian Gulf area was established by W.E. Browne of the Iraq Petroleum Company in 1927-1931 at the South End Base at station Nahrwan (East of Baghdad) such that: $\Phi_0 = 33^\circ 19^\prime 10.87^\prime\prime$ North, $\Lambda_0 = +44^\circ 43^\prime 25.54^\prime\prime$ East of Greenwich, and the Clarke 1880 is the ellipsoid of reference where: $a = 6,378,300.782$ m, and $1/f = 293.4663077$. The Nahrwan Datum of 1929 is the most prevalent coordinate system of the entire Arabian Gulf area and is still found to this day.

The first major geodetic datum of the Arabian Gulf area was established by W.E. Browne of the Iraq Petroleum Company in 1927-1931 at the South End Base at station Nahrwan (East of Baghdad) such that: $\Phi_0 = 33^\circ 19^\prime 10.87^\prime\prime$ North, $\Lambda_0 = +44^\circ 43^\prime 25.54^\prime\prime$ East of Greenwich, and the Clarke 1880 is the ellipsoid of reference where: $a = 6,378,300.782$ m, and $1/f = 293.4663077$. The Nahrwan Datum of 1929 is the most prevalent coordinate system of the entire Arabian Gulf area and is still found to this day.
ic mapping was compiled in the 1960s by the British Directorate of Military Surveys, and a 1:50,000 scale map series (K763) comprising 155 sheets was completed in 1969.

Limited 1:50,000 mapping was produced in the early 1980s with the assistance of Syria, but coverage of Dubai was based on only four third-order points in the southeast corner of the Emirate. A military survey department was set up by the Emirates and new mapping was published between 1989 and 1991 as 138 orthophoto sheets on the Nahrwan Datum of 1929 with the UTM Grid. A new GPS network was initiated for Dubai in 1991 with a new local Grid. The Dubai Local Transverse Mercator (DLTM) Grid is referenced to the WGS 84 ellipsoid, the Central Meridian $\lambda_0 = +55^\circ 20'$ E, and the False Easting = 500 km. The Northings are presumably measured from the Equator. Analysis of the old network indicated a potential positional error of the old classical control of up to 9 meters horizontal. The First Order Geodetic GPS Network of Dubai is composed of 62 monumented points with distances between points ranging from 5 to 10 km. Of particular interest is that Dubai has completely abandoned the previous classical geodetic work extant in the Emirate. Zero effort was (apparently) made to relate the old to the new! I personally do not agree with this philosophy because I prefer to relate historical records to current and future work. However, I suspect that this unfortunate tack may be chosen from time-to-time for the sake of expediency.

Satellite positioning studies (by others) in the United Arab Emirates derived a set of Datum shift parameters from WGS72 Datum to Nahrwan Datum of 1929 where: $\Delta X = +225.4$ m, $\Delta Y = +158.7$ m, $\Delta Z = +378.9$ m, based on observations of 8 stations. I personally would consider the tenths of a meter used in these parameters as very optimistic. Interestingly, NIMA lists the transformation from Nahrwan 1929 to WGS 84 as $\Delta X = -249$ m, $\Delta Y = -156$ m, $\Delta Z = -381$ m, $\pm 25$ m, based on two stations observed in 1987.

**Update**

Significant developments have been implemented in the coordinate reference systems for two primary areas of the United Arab Emirates: in Al Ain Region, Abu Dhabi and in Dubai. In Al Ain Municipality, Professor Kamal A. Abdalla of the University of Khartoum, Sudan spent a number of years as a consultant to the region, and in 2005 published a paper that stated, "The local geodetic network adopt(ed) Ras Ghantut datum and (is) based on the modified Clarke 1880 ellipsoid. The local control stations are non-homogeneous, unadjusted and have many limitations in terms of spatial data applications. While the global geodetic network is tied to the ITRF system, containing 33 well-distributed geodetic control stations. The transformation parameters between the global datum and the local datum were computed." 

In the Dubai Municipality Survey Department, a Leica SmartNet™ has been established in order to provide a virtual reference system (VRS) to the emirate. Furthermore, the paper, An Absolute/Relative Gravity Base Net in the Emirate of Dubai, details a new fundamental gravity network of nine relative and absolute gravity stations and that "These measurements were the first absolute gravity determinations in the whole South-West Asia."

---

1 Al Ain Local and Global Spatial Reference Systems, Map Middle East 2005.
2 http://www.m1.ae/DVRS.html.

The contents of this column reflect the views of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the American Society for Photogrammetry and Remote Sensing and/or the Louisiana State University Center for GeoInformatics (C4G).

This column was previously published in *PE&RS*. 

---
GREETINGS FROM THE STUDENT ADVISORY COMMITTEE (SAC)!

Geo Week 2019 anchored by the APSRS annual conference was a resounding success. This year’s conference took place in Denver, Colorado. The attraction of industry, government, and academia made this year’s conference a rewarding experience for students and young professionals. We will use the remainder of the page to update everyone on the SAC’s sessions and remind students to apply to join the SAC!

ASPRS ANNUAL CONFERENCE
The SAC co-hosted a very successful session which highlighted the student research. Many student members also participated in the poster sessions. These posters covered many aspects of geospatial technologies and remote sensing. Another highlight was the GeoLeague, which included members of the SAC and wider ASPRS student community, competing in a test of geospatial knowledge.

JOIN THE SAC!
The SAC is a great place to learn, grow, and develop as a leader and professional society member. The SAC facilitates student participation in conference planning and other aspects of the societies functioning. All graduate and undergraduate students are encouraged to apply at https://www.asprs.org/councils/student-advisory-council-positions.

If you did not attend the Geo Week this year, we hope to see you in the fall at Pecora 21 in Baltimore, Maryland!

As always send any questions or inquiries to SAC@asprs.org
ASPRS Aerial Data Catalog

“The Source for Finding Aerial Collections”

HTTP://DPAC.ASPRS.ORG

The ASPRS Aerial Data Catalog is a tool allowing owners of aerial photography to list details and contact information about individual collections.

By providing this free and open metadata catalog with no commercial interests, the Data Preservation and Archiving Committee (DPAC) aims to provide a definitive metadata resource for all users in the geospatial community to locate previously unknown imagery.

DPAC hopes this Catalog will contribute to the protection and preservation of aerial photography around the world!

ASPRS Members: We Need Your Help!
There are three ways to get involved:

1. Use
   Use the catalog to browse over 5,000 entries from all 50 states and many countries. Millions of frames from as early as 1924!

2. Supply
   Caretakers of collections, with or without metadata, should contact DPAC to add their datasets to the catalog free of charge!

3. Tell
   Spread the word about the catalog! New users and data collections are key to making this a useful tool for the community!

For More Details or To Get Involved Contact:

DAVID RUZ • druiz@quantumspatial.com • 510-834-2001 OR DAVID DAY • dday@kasurveys.com • 215-677-3119
ASPRS MEMBERSHIP

ASPRS would like to welcome the following new members!

Alaska
Philip Tworek

At Large
C. Kelly
Justin Lindeman
Yitzhak Wolf Marmelstein
Lee Ann Nolan
Ingrit Lolita Sari
Rennie Smith

Columbia River
Brittany Biro
Stephen Carter
Lindsay Cumella
Robin Flowers
Thomas Hart
Brooke Hunter

Stephanie Jameson
Josh Kowalski
Dean Olson
Ethan Poole
Lauren Sharwood
Devin Shiang
Kyler Speich
Amanda Temple
Victoria Zohner

Eastern Great Lakes
Kay Yamakawa

Florida
Jackson Alan Beebe
David F. McKay, CP
Penny Simone

Inter-Mountain
Ryan Howell

Mid-South
Nima Ekhtari
Jonah Gabriel Hall
John Bosco Osimbo Namwamba

New England
Scott Stetson

Pacific Southwest
Michael Carlson
Joshua Wagner

Poconos
Alan Mikuni

Puget Sound
Bradford James Hamson

Western Great Lakes
Andrew Williams

ASPRS MEMBERSHIP

ASPRS would like to welcome the following new members!

Alaska
Philip Tworek

At Large
C. Kelly
Justin Lindeman
Yitzhak Wolf Marmelstein
Lee Ann Nolan
Ingrit Lolita Sari
Rennie Smith

Columbia River
Brittany Biro
Stephen Carter
Lindsay Cumella
Robin Flowers
Thomas Hart
Brooke Hunter

Stephanie Jameson
Josh Kowalski
Dean Olson
Ethan Poole
Lauren Sharwood
Devin Shiang
Kyler Speich
Amanda Temple
Victoria Zohner

Eastern Great Lakes
Kay Yamakawa

Florida
Jackson Alan Beebe
David F. McKay, CP
Penny Simone

Inter-Mountain
Ryan Howell

Mid-South
Nima Ekhtari
Jonah Gabriel Hall
John Bosco Osimbo Namwamba

New England
Scott Stetson

Pacific Southwest
Michael Carlson
Joshua Wagner

Poconos
Alan Mikuni

Puget Sound
Bradford James Hamson

Western Great Lakes
Andrew Williams

For more information on ASPRS Membership, visit
HTTP://WWW.ASPRS.ORG/JOIN-NOW

Your path to success in the geospatial community

STAND OUT FROM THE REST

EARN ASPRS CERTIFICATION

ASPRS congratulates these recently Certified and Re-certified individuals:

CERTIFIED MAPPING SCIENTIST UAS
Mark Treon, Certification #UAS026
Effective December 21, 2018, expires December 21, 2023

CERTIFIED PHOTOGRAMMETRIST
Sean Kampe, Certification #1644
Effective December 21, 2018, expires December 21, 2023

RECERTIFIED PHOTOGRAMMETRISTS
James R. Cannistra, Certification #R850
Effective September 18, 2018, expires September 18, 2023

Clayton R. Hopkins, Certification #R1379
Effective October 29, 2018, expires October 29, 2023

R. Steven Denney, Certification #R684
Effective October 8, 2018, expires October 8, 2023

Jason B. Krueger, Certification #R1558
Effective October 27, 2018, expires October 27, 2023

Clay D. Smith, Certification #R939
Effective January 11, 2019, expires January 11, 2024

Alan Mikuni, Certification #R1127
Effective October 19, 2018, expires October 19, 2023

Patrick G. Moroney, Certification #R1377
Effective October 19, 2018, expires October 19, 2023

Debra A. Taylor, Certification #R1539
Effective December 19, 2017, expires December 19, 2022

Kevin Reid, Certification #R1557
Effective October 27, 2018, expires October 27, 2023

Evan H. Brown, Certification #R1342
Effective December 4, 2017, expires December 4, 2022

RECERTIFIED PHOTOGRAMMETRIC TECHNOLOGIST
John Ong, Certification #R1578PT
Effective January 1, 2018, expires January 1, 2021

ASPRS Certification validates your professional practice and experience. It differentiates you from others in the profession.

For more information on the ASPRS Certification program: contact certification@asprs.org, visit
https://www.asprs.org/general/asprs-certification-program.html
ASPRS Begins 2019 with a New Association Management Firm


P & N Association Management will provide association management services to ASPRS that will complement and supplement the efforts of ASPRS staff. P & N Association Management is a subsidiary of Postlethwaite & Netterville, a professional services firm of over 400 professionals providing Accounting and Assurance, Consulting, Tax and Technology services to clients. P & N Association Management has served the needs of its association clients since 1991 and currently serves the needs of 18 clients.

After several months of study and deliberation, the ASPRS Board of Directors decided to hire a company to provide association management services. This collaboration with P & N will provide increased value to ASPRS membership and business partners. ASPRS is pleased to partner with P & N and looks forward to a long and mutually beneficial relationship.

Established in 1934, the American Society for Photogrammetry and Remote Sensing is a scientific association serving over 2000 professional members around the world, providing its members professional development through education and networking experiences, professional certification, publications, scholarships, and other services. The American Society for Photogrammetry and Remote Sensing advances the knowledge and improves understanding of mapping sciences to promote the responsible applications of photogrammetry, remote sensing, geographic information systems and supporting technologies.

The ASPRS Board of Directors consists of: President, Anne Hillyer; President-Elect, Tommy Jordan; Vice-President, Jeff Lovin; Immediate Past President, Rebecca A. Morton;

Treasurer, A. Stewart Walker; Secretary, Roberta E. Lenczowski; Acting Executive Director, James T. “Jesse” Winch; and Council Chairs: Region Officers Council, Lorraine Amendia; Technical Division Directors Council, John McCombs; Student Advisory Council, Ge “Jeff” Pu; Early-Career Professionals Council, Amanda Aragon; Corporate Members Council, Joe Cantz.
Repeated Structure Detection for 3D Reconstruction of Building Façade from Mobile Lidar Data

Yanming Chen, Xiaoliang Liu, Mengru Yao, Shulin Deng, Feixue Li, Liang Cheng, and Manchun Li

Abstract

This study proposes a new method for repeated structure detection and the three-dimensional (3D) reconstruction of building façades from mobile lidar data. Firstly, the building façade is divided and unrolled to simplify the complex façade structures, improving the automation of structure detection and building reconstruction. Subsequently, the unrolled façade is decomposed into tiles by analyzing the repeated structures. Tiles with strong similarities are matched and merged to restore the imperfect façade points. Based on the restored points and repeated structures, a 3D building façade can be reconstructed with a complete structure and fine detail. An analysis is conducted to compare the constructed 3D model with the lidar points of actual façade. The results of this analysis demonstrate that the proposed method can effectively deal with missing areas caused by occlusion, viewpoint limitation, and uneven point density, as well as realizing the highly complete 3D reconstruction of a building façade.

Introduction

The three-dimensional (3D) reconstruction of a building model is a hot topic of research in many fields such as architecture, engineering, construction, change detection and urban planning (Chen et al., 2014; Huang et al., 2017; Qin et al., 2015; Xu et al., 2015). It is an important task to detect and reconstruct buildings from optical images or lidar data, and therefore has attracted considerable attention over the past few decades (Huang and Zhang, 2012; Huang et al., 2011; Wang et al., 2016). In recent years, great progress has been made to reconstruct 3D models of buildings and façades quickly and accurately (Li et al., 2017; Wang et al., 2016). The modeling of different features on a building façade has attracted increased attention because it features in many applications (Cheng et al., 2018; Wang et al., 2018).

Mobile lidar has become an efficient means of data acquisition for the modeling of building façades (Leberl et al., 2010). It is an emerging mobile mapping system that can rapidly capture road surface features from high-speed vehicles with an overlooking or upward-looking viewpoint (Cheng et al., 2014). Despite the complications of urban landscapes, this system offers the advantages of highly precise positioning, quick acquisition, and abundant façade detail (Yu et al., 2015). Thus, the mobile lidar is extremely advantageous to the 3D reconstruction of large-scale urban landscapes (Musialski et al., 2013). However, due to the existence of extremely complex structures, the limitations of data collection, and the occlusion and disturbance of objects on the street (e.g., trees and vehicles), the reconstruction of a high-quality building façade model with a high level of automation presents a challenging but very worthwhile research topic (Wan and Shafar, 2012).

Consequently, exploratory research into the reconstruction of building façade models has been undertaken by several researchers and relevant reviews (Haala and Kada, 2010; Musialski et al., 2013; Tang et al., 2010). Specifically, Becker (2009) constructed the details of windows, doors, and protrusions by utilizing the transmission characteristics of lidar points, and a formal grammar was generated for the reconstruction of building façades. Similarly, Pu and Vosselman (2009) presented a knowledge-based approach to the automatic reconstruction of building façade models. However, these methods can be easily affected by the data quality, especially by occlusions and noise existing in the points. In addition, other details like balconies also make reconstruction more difficult.

Urban buildings usually have characteristics stemming from the time at which they were built, as well as the culture and tastes of the region. Therefore, similar structures are likely to be seen in many building façades (Mitra et al., 2013). The detection of similar structures is usually based on the hypothesis that structure cells present a repetitive lattice distribution (Minwoo et al., 2009). Many researchers have extracted building façade structures from images (Ceylan et al., 2012; Xiao et al., 2008). Müller et al. (2007) proposed shape grammar rules that had been derived from façade images to subdivide the texture of a façade into repeated elements such as windows, doors, and tiles. Li et al. (2011) presented a state-of-the-art method by fusing 2D photographs and 3D lidar points to decompose depth-layer of façades and produce 3D consolidated models. Teeravech et al. (2014) discovered repetitive patterns by using a RANSAC-styled sine wave fitting algorithm, and then decomposed the façade images of buildings into floors and tiles automatically.

For building façade point clouds, researchers have also taken advantage of structural regularities to solve the problems caused by occlusion and noise. Pauly et al. (2008) introduced a framework for identifying repeated elements as regular lattice structures in point- or mesh-based models. However, this framework cannot handle “warped” sequential repetitive structures. Bokeloh et al. (2009) identified similar and repetitive structures through the line feature matching algorithm. This algorithm works only when all the structures are identical, and the reconstruction result is highly dependent.
on the detection of line features. Zheng et al. (2010) proposed a new method for modifying building façade points by fusing non-local point cloud parts with repetitive and self-similarity structures. This approach requires manual intervention for the detection of repetitive structures. Nan et al. (2010) presented an interactive interface, which uses simple building blocks in the form of axis-aligned rectangular cuboids called Smart-Boxes, to improve the reconstruction of sparse point clouds. This interactive tool assists users in fitting highly repetitive structures with Smart-Boxes from 3D point clouds. Shen et al. (2011) proposed an adaptive partition method of urban façade based on repetitions and symmetry detection in terrestrial lidar points. This method can extend the technique of Pauly et al. (2008) to be applicable to building facade that is not globally rectilinear. However, it focuses on how to split façade structures rather than on detecting repetitive patterns. Further, it is not suitable for complicated facades in different planes, and the partition results are dependent on the first splitting step. Based on these findings, Wan and Sharf (2012) introduced a grammar-based segmentation method to deal with the depth layer of the façade points. This method is restricted to a small number of possible input buildings, and the geometry details might be affected due to the lack of expression in the shape grammar. To attain better decomposition results, some researchers have attempted to use waveform-fitting methods to solve the problems caused by occlusion and gaps. Friedman and Stamos (2013), for instance, introduced an online method for the detection of repeated features through Fourier analysis of column functions, and applied a square wave to fit repetitive façade features. However, although this method can detect repetitive planar façade structures, it cannot detect regularity and is not suitable for low-rise residential structures. Li et al. (2017) proposed a hierarchical modeling method by using the semantic segmentation and underlying façade structures of building points. However, this method just used repetitive and symmetric patterns of façade to parse the façade elements, which were not suitable for the consolidation of missing data. In general, due to the large number of points, uneven point density, significant amounts of occlusion, and complicated landscape, extracting building points and reconstructing building models remains a challenging task. Further, façade structure detection is still in its exploratory stage. Most of these methods are applicable only to planar façades, while complex buildings with different planar façades have to be divided separately before further processing. Some researches rely on user assistance to detect repetitive structures or reconstruct façade models. It is necessary to improve the level of automation. Furthermore, only a few studies have investigated the optimal use of strong repeated structures for the restoration of imperfect building points and reconstruction of integrity façade models. Thus, these methods are not capable of large-scale reconstruction.

In this study, we focused on repeated structure detection in building façades for the reconstruction of high integrity 3D building façade models based on lidar data collected by mobile mapping systems. The primary contributions of this study are: (1) To reduce the difficulties associated with building structure detection and reconstruction of multiple façades, an unfolding method that flattens uneven wall surfaces, and front and side building façades onto a simple plane with depth values is introduced; (2) To identify splitting boundaries between similar structures and non-repetitive components, a two-direction detection strategy that divides the unfolded façade into small periodic tiles, with façade structure detection performed on cumulative histograms of façade points distributed along the vertical and horizontal directions successively, is proposed; (3) To consolidate noisy and incomplete façade points, a restoration method that restores imperfect data with repeated structures, according to the procedure of tile matching and similarity measurement, is developed.; and (4) To reconstruct 3D building façade models more easily, repeated façade structures with smaller size and simpler geometry are adopted, and the restored façade points improve the accuracy of façade reconstruction. These procedures will guarantee the restoration of imperfect lidar points and reconstruction of high integrity 3D building models in a large area.

### Methodology

The method includes three main stages: the preprocessing of façade points, the detection of repeated façade structures, and the 3D reconstruction of building façade, as shown in Figure 1. First, preprocessing procedure is adopted to divide building façades into simpler parts, and an unrolling method is proposed to transform the building façade points to a planar surface. Second, a structure detection method is presented to discover the tile structures, including the extraction of 3D façade edge points, the detection of vertical and horizontal structures. Third, building façade points are restored according to the detection of repeated façade structures, and 3D building façade models are reconstructed by using the restored façade points and the repeated façade structures.

### Preprocessing

The purpose of this section is to divide the complex structures of each individual building into simple parts, and then flatten the façade points of each building parts from 3D space to a 2D plane. The subsequent procedures are performed on the unrolled building façades.

### Dividing of Building Façade Parts

The building façade points were scanned by mobile lidar (such as Figure 2a) often have complex structures, as the lower and upper parts of the building do not coincide, which are difficult to unroll these façade points into a 2D plane. It is needed to divide such complex structure of building façades into simple parts to extract the building contours individually. As Figure 2b shows, for structural stability, the lower part of these complex buildings usually has a broader structure than the upper part. Building façade points are acquired using

---

**Figure 1. Flow of building façade reconstruction.**
a side-looking vehicle-based mobile laser scanner. Thus, the initial section of the upper part (as indicated by the red region in Figure 2a and 2b) is occluded by the lower part, which leads to missing façade point data for the occlusion area. Therefore, histogram value along the elevation direction is the minimum at the height of occlusion area.

By taking advantage of this feature, a histogram (as black line shown in Figure 3) is constructed by counting the cumulative point number along the elevation direction of the building façade points. There is a local minimal value (as the elevation of 18.5 m in Figure 3) in the histogram along the vertical direction. And, it is clearer as the smooth result of histogram (as the red line shown in Figure 3) based on the FFT filter method. According to the height of 18.5 m (the local extremum low value), the building is divided into the lower part and the upper part. Within the upper part, two disconnected components are detected as island groups which are distinct from other groups by evaluating the proximity of points. Finally, three simple components (as part A, B, and C shown in Figure 4) have been detected. The subsequent steps are processed for each decomposed parts individually.

**Unrolling of Building Façade Points**

A building façade is usually composed of many developable wall surfaces, such as planes, single-curved surfaces, and cylinders. These surfaces can be unrolled (unfolded or flattened) onto a plane as rigidly as possible. In other words, the surface details, such as the depth information of windows, balconies, and other objects on the walls, tend to be preserved without distortion. As shown in Figure 5, the building façade is composed of surfaces A, B, C, D, and E. Unrolling of these surfaces essentially involves mapping of these 3D rectangles onto a planar plane, as surfaces A’, B’, C’, Dv, and E’ in Figure 5.

Accordingly, in order to reduce the difficulty associated with data processing for multiple façade surfaces, the lidar points of each building façade will be unrolled separately onto a plane. Furthermore, the unrolling of building façade surfaces can be simplified as the unrolling of 2D façade footprints (indicated by the red curve of the building façade in Figure 5) to a straight line (indicated by the straight red line of the unrolled building façade in Figure 5), by following the five-step process described below.

1. **Extraction of the building façade footprints.** During the unrolling, the footprints of the building façade (the red
segment ABCD in Figure 6) are extracted first. A coarse-to-fine approach proposed by Yang et al. (2013) is used to accurately extract the building façade footprints from the mobile lidar point clouds in the urban environment. This method uses georeferenced feature images, the RANSAC and PCA algorithms to extract coarse façade footprints. Thereafter, the extracted footprints for each building façade are grouped, and the redundant footprints are eliminated by calculating the length and slope of each footprint. Finally, the extracted façade footprints that belong to a single building are connected and harmonized as the result.

2. Detecting control points of the building façade footprints. These extracted building façade footprints are set as the directrix curves of the building façade surfaces. The curvature \( k \) of the directrix curve is computed to detect the points of the curve endings and corners, where the value of the curvature \( k \) is the local maximum. These points are selected as the control points \( P_i = \{x_i, y_i, z_i\} \), such as the points A, B, C, and D on the curve shown in Figure 6.

3. Dividing the building façade points into rigid objects. The normal vector \( n \) of the control points is calculated. Subsequently, a plane along the normal direction through the control point is constructed, such as the planes \( ab \) and \( cd \) through the control points B and C, respectively, in Figure 6a. These normal direction planes divide the lidar points of the original building façade into several small parts. Each part is regarded as an individual rigid object. This means that the object itself will not be deformed; however, its position will be transformed.

4. Establishing a mapping relationship between the rigid objects and control points. A corresponding relationship between the divided rigid objects and the control points is then established. This step converts the unrolling of the complicated building lidar points into the rigid transformation of straightening the building façade footprints, which can effectively avoid the unrolling of the original points, and improve the processing efficiency.

5. Unrolling of the building façade points. The surface deformation method (Botsch and Sorkine, 2008) is used to straighten the building footprints and the converted control points \( P_i \) are recorded as \( P'_i = \{u_i, v_i\} \). Then, lidar points of each rigid object are translated and rotated according to the control points. The coordinates of \( P_i \) and \( P'_i \) can be operated as the column vectors, and the conversion of \( P_i \) is completed by applying Formula 1.
Finally, the unrolling of the building façade points is realized as shown in Figure 6b and Figure 7. The depth information of the windows, balconies, and other objects on the walls is preserved, as shown in Figure 7. Therefore, the details of the difference in height between the windows on the side and front walls can be preserved with little influence on the detection of repetitive structures.

**Detection of Façade Structures**

The walls, windows, balconies, and other objects in a building façade usually present similar structures that are distributed regularly and repetitively. In this study, the similarities between these objects in the façade are utilized to enable the detection of repeated structures and to solve the problem of missing data, so as to improve the completeness of the building façade.

Based on the unrolled lidar point, the neighborhood statistics algorithm (Rusu et al., 2008) is applied to filter the noisy points. The alpha shapes algorithm (Edelsbrunner et al., 1994) is then adopted to extract the edge points (the red points in Figure 8), which actually correspond to the points at the edges of the walls, windows, balconies, and so on. In this way, the initial features of façade structures become more distinct, and the quantity of façade points is reduced significantly. The subsequent processes include two steps: the detection of vertical structures and then the detection of tile structures.

**Detection of Vertical Structures**

A histogram is constructed by counting the cumulative point number along the elevation direction of the unrolled edge points, as shown in Figure 9. The vertical structures (building floors) can be detected by extracting the periodic peaks in the histogram.

1. **Construct and normalize histogram.** First, points within different elevation intervals are added to obtain the histogram (Figure 9 and the red waveform in Figure 10). The interval value was selected as 0.4 m experimentally to obtain a good shape of the statistical data in a histogram. Then, the base waveform (blue waveform in Figure 10) is calculated through the application of Formula 4 with a sliding window (\(i = 5\), which is the optimal experimental result obtained with major trends of the lower envelope and details of the waveform) for the lower envelope value. The normalized histogram (yellow waveform in Figure 10) is obtained by applying a subtraction operation between the accumulation waveform and base waveform, to solve the problem of uneven point density.

\[
y_k = \min(x_{k-i}, x_{k-i+1}, \ldots, x_{k-i+t}, x_{k+i})
\]

2. **Detect building floor structures.** According to the periodic characteristics of the building floors, a sine wave is used to fit the peaks in the normalized histogram. Formula 5 is used to fit the candidate sine waves from the histogram, as shown in Figure 11. In the formula, the amplitude \(A\) is the maximum height of the sine wave, \(y_0\) is the offset position, and \(y_k\) is the result obtained with major trends of the lower envelope and details of the waveform.

\[
A = \frac{\text{Amplitude of Sine Wave}}{2}
\]

\[
y_k = A \sin(\frac{2\pi}{\text{Period of Sine Wave}} k) + y_0
\]
of the central amplitude, \( x_c \) is phase shift position, \( \lambda \) is the wave-length between two successive positive peaks, \( x_0 \) is the first positive peak corresponding to the elevation of the lowest floor, and \( x_i \) corresponds to the positions of the other floors. During the fitting procedure, a low priority is set for the lowest and highest floors, so as to avoid the offset might be caused by their irregular structures. In addition, the Levenberg-Marquardt algorithm (Pujol, 2007) is adopted to minimize the residual sum of squares for the optimal fitting result.

\[
y = y_0 + A \sin \left( 2\pi \frac{x - x_c}{\lambda} \right) \\
x_c = x_0 + \frac{\lambda}{4} \\
x_i = x_0 + k\lambda, \quad k = 0, 1, \ldots, n
\]  

(5)

**Detection of Tile Structures**

Compared to the repetitive characteristics of a building’s floor structures, the façade structures might be particularly random in the horizontal direction. To overcome this, waveform smoothing and peak extraction are performed to detect the horizontal structures. Finally, tile structures are determined by the cross of vertical and horizontal structures.

1. **Smoothing the waveform to detect preliminary horizontal structures.** The histogram in the horizontal direction (green waveform in Figure 12) is generated using the method described above. By using Formula 4 with a smaller sliding window size \( i = 2 \) to obtain more waveform details of the histogram, a base waveform (red waveform in Figure 12) approximating the original shape of the histogram is constructed. Then, based on the local minimum method, the valley positions of the base waveform are extracted as the decomposition positions for the preliminary horizontal structures (red dotted lines \( a, b, \ldots, m \) in Figure 12). In addition, the higher value of the base waveform means the higher point density, and the structure contains the more detailed information. Thus, according to the base
waveform values at the position of the dotted line \(a\) to \(m\), the ranks of the decomposed structures can be obtained with a positive correlation.

The structure ranks are used to decide the priority of tile matching in the following process. In Figure 12, the white dotted lines \(A, B,\) and \(C\) were drawn according to the 4-quantiles for the \(Z\) values of the red base waveform, dividing the distribution into four groups of equal size. In other words, the height of the white dotted line \(A\) is determined by the first quantile (25%) of the ascending ranked \(Z\) value, the white dotted line \(B\) is based on the median or second quantile (50%) of the \(Z\) value, and the white dotted line \(C\) depends on the third quantile (75%) of the \(Z\) value. As indicated by the white dotted line \(A\), the façade is decomposed into regions \(ab, bd, dj, jk, kl, lm\), which are assigned the low rank. The white dotted line \(B\) further decomposes \(dj\) into \(de, el\), and \(lj\), assigned as medium rank. Dotted line \(C\) decomposes \(bd\) into \(bc\) and \(cd\) and decomposes \(el\) into \(ef, fg, gh,\) and \(hl\), all of which are assigned as high rank. Thus, more elaborate details can be obtained for the tile grouping in the following section.

2. **Peak extraction for refining position of horizontal structures.** The preliminary horizontal structures are detected from the base waveform, in which it is difficult to precisely locate the boundaries of repeated building structures, such as the edges of walls, windows, and balconies. Thus, the peak position of the original histogram is detected for the detailed horizontal structures, and then used to refine the preliminary horizontal structures. The peaks in the histogram are extracted by adopting the local maximum method, and those peaks with extremely low values are eliminated. The extracted peaks are shown as the red dotted lines in Figure 13, and the blue lines in the figure are the edges of the building walls. Then, by searching the neighborhood of each decomposition position (red dotted lines in Figure 12), the location of the nearest high-value peak (red dotted lines in Figure 13) is used to refine the preliminary horizontal structure. Finally, by superimposing the horizontal and vertical structures together, the tile structures of the building façade are detected as shown in Figure 14.

### Reconstruction of Building Façade

According to the results of detecting the tile structures, the building façade can be split into pieces with smaller size and simpler geometry, which make the reconstruction much easier. In this section, repeated structures are detected by similarity measurement, and the imperfect façade points are restored according to the repeated structures. After that, tile models are reconstructed, and a complete model of the building façade can be generated by combining these repeated tile models.

### Restoration of Repeated Façade Structures

Detected tiles with high levels of similarity are merged and any noise points are eliminated to restore the imperfect parts, such that the building façade points can be obtained with a high degree of completeness. To attain this, rigid registration is used to match the repetitive tiles, and the root mean square (RMS) distance is then calculated to determine the similarity of each tile pair.

Before the processing, tiles are divided into groups with different matching priorities. The tiles in any one column are grouped and assigned the highest priority; the tiles having a similar bounding-box size are grouped and assigned a higher priority; and a high priority is assigned to that group of tiles with the same rank, as obtained in the previous section in which preliminary horizontal structures were detected. In each group, the tiles with the high point density are taken as the “reference” data \(Y\), while the other tiles with sparse points are regarded as being “matched” data \(X\), and the reference tile and a matched tile constitute a tile pair. Take building façade tiles in the same column for instance (the yellow marked region in Figure 15), there are 25 floors in the vertical structure, dividing the column into 25 tiles. Choose the tile on the 8th floor, which has the highest point density, as the “reference” tile, and the other 24 tiles as the “matched” tile.

1. **Tile matching.** Firstly, a tile pair is roughly matched by using the bounding-box centers, which transform the two centers of the tiles to coincide. Then, the iterative closest point (ICP) algorithm (Besl and McKay, 1992) is used to match a tile pair precisely, and the rotation parameter \(Rot\) and translation parameter \(Trans\) in Formula 6 are calculated to obtain rigid transformation result \(S\).

\[
S = Rot(S) + Trans \tag{6}
\]
Figure 15. Façade points restoration of the repeated façade structures: (a) Similarity measurement of the tile pairs in each floor. The tile on the 8th floor is taken as the “reference” data to calculate the similarity $d_{RMS}$ from floors 01 to 25, and the tile in the 22nd floor is taken as the “reference” data to determine the similarity $d_{RMS}'$ from floors 22 to 25; and (b) Restoration of the building façade points by merging repetitive tiles and removing outlier points, such as the tiles from floors 01 to 21 and the tiles from floors 22 to 24.

Figure 16. Restoration result for building façade (Red dotted lines are the detailed building structures).
The Rot can be derived from the cross-covariance matrix $C$, which is formed for the $n$ correspondences $(s_i, y_i)$ of the matched tile $S$ and referenced tile $Y$.

$$C = \frac{1}{n} \sum_{i=1}^{n} (s_i - \bar{s})(y_i - \bar{y})^T$$  \hspace{1cm} (7)

Performing the SVD of $C$:

$$USV^T = C$$ \hspace{1cm} (8)

where $U$ and $V$ are two orthogonal matrices and $S$ is a diagonal matrix of singular values. The rotation matrix $Rot$ can be calculated from the orthogonal matrices as:

$$Rot = VU^T$$ \hspace{1cm} (9)

The translation vector $Trans$ can be estimated as:

$$Trans = Y - VU^T S$$ \hspace{1cm} (10)

2. **Similarity measurement.** The distance error for the tile pair is computed by calculating the distance $d_i$ from each point $s_i'$ of the matched tile $S'$ to its nearest point $y_i$ on the reference tile $Y$, as Formula 11 shows. After that, the RMS distance $d_{RMS}$ of the tile pair between “reference” tile and each “matched” tile is calculated to measure the similarity according to Formula (12). Due to discreteness of point cloud, the points with distance $d_i > 0.5$ m will be removed to avoid the cumulative errors between non-correspondence point pairs, then the number of reserved points is counted as $n$. The similarity results are classified into high and low values by using the K-Means clustering algorithm. In addition, the smaller value of $d_{RMS}$ indicates the greater degree of similarity.

As in Figure 15a, compared with the “reference” tile on the 8th floor, the tiles from floor 01~21 have a higher level of similarity, and thus are taken as repetitive tiles; and the tiles from floor 22 to 25 have a lower level of similarity, and therefore taken as dissimilar tiles. After that, these dissimilar tiles from floor 22 to 25 are set as a new group to recalculate RMS distance $d_{RMS}$. Among them, the tile on the 22nd floor, which has the highest point density, is taken as the new “reference” tile for similarity measurement form floor 22 to 25. At this time, compared with the new “reference” tile on the 22nd floor, the tiles from 23 to 24 with smaller $d_{RMS}$ values are classified as repetitive tiles, and the tile on the 25th floor is different from the other tiles.

$$d_i = \min_{y_i \neq s_i'} \|y_i - s_i'\|$$ \hspace{1cm} (11)

3. **Restoration of building façade points.** Those tiles with a high level similarity are merged as repetitive structure, and then the noisy points are eliminated for the high-integrity tiles by adopting the sparse outlier removal algorithm (Rusu et al., 2008), as shown in Figure 15b. Subsequently, the restored tile is used to fill the repetitive parts with the application of inverse transformation. As a result, a detailed building façade is obtained as shown in Figure 16. The red dotted lines in the figure are the detailed building structures obtained in the previous section.

**3D Reconstruction of Building Façade**
To reconstruct the tile model, line features are first fitted from the restored edge points. Most façade features, such as the walls, windows, and balconies, have rectangular shapes, for which the boundaries have perpendicular structures. Therefore, by using the detailed building structures (red dotted lines in Figure 16 and Figure 17a) as a guide, the 3D line segments are efficiently fitted by the application of the least squares algorithm along the restored edge points, and then the enclosed regions surrounded by the outlines are constructed as the façade surfaces, such as the gray areas in Figure 17b. Besides, the width of window edge is set to 5 cm, and the window surfaces are positioned in the wall with an offset parameter about 10 cm, such as the dark green areas in Figure 17c.

Topology relationships of the surfaces are generated using an improved split-merge-shape (SMS) method that we proposed in Cheng et al. (2011). The intersection surfaces are extended, trimmed and merged as shown in Figure 18a. For parallel surfaces, the distance $d$ between two surfaces are calculated (Figure 18b). If the $d$ is smaller than 0.1 m, the two parallel surfaces are regarded as a continuous surface, and construct a new surface in the middle position of two parallel surfaces. If the $d$ is larger than 0.1 m, the two parallel surfaces are regarded as a three-folded surface, and connect the cross-section of two parallel surfaces. This provides an effective solution for automatically recovering the topology relationships between these 3D surfaces, including the automatic recovery of lost boundary segments for which laser points are either unavailable or only partially available.

Subsequently, repeated tile models are duplicated, transformed, and merged to compose the entire building façade model. By means of an inverse unrolling transformation, the building polyhedral model is obtained. After that, visual inspections are performed, and additional manual modification is applied to deal with the connection problems and topological errors, while combining small tile models into a complete building façade model. Except for such modification that requires user interaction, the other steps in the pipeline are

---

**Figure 17.** Local model reconstruction of building façade tiles: (a) Fitting linear features of walls, doors, windows and balconies; (b) Constructing façade surfaces by enclosing boundaries, and (c) Constructing the window surfaces and window edges.
Finally, a building façade model with high geometric quality is reconstructed as shown in Figure 19.

**Experiment and Analysis**

**Experimental Data**

The mobile laser scanning system selected for this study was the Lynx Mobile Mapper, which we used to acquire the data for our experiments on 12 October 2011. This system has a 360° field of view with a maximum range of 200 m, a positioning accuracy of about 0.02 m in the X and Y directions, and 0.05 m in the Z direction.

The part of Shuiximen Street in Nanjing, China, which covers an area of about 500 m × 300 m, is selected as the experimental area. The building façades in the experimental area have diversified structures, which are representative for 3D reconstruction research of building façade based on mobile lidar data. As Figure 20 shows, the point cloud of building façades contains 2.72×10^6 lidar points: its bounding box is about 378 m × 222 m × 85 m, and the point cloud is overlaid with a sliding scale coloration ranging from blue at minimum Z to red at maximum Z.

Additionally, another two datasets were used for method comparison. The Stdh façade (Figure 21a) contains 1.09×10^5 points and its bounding box is approximately 37 m × 14 m × 7 m. The Apartment façade (Figure 21b) contains 4.63 × 10^5
points and its bounding box is approximately $53 \times 27 \times 4$ m. These two building façade point clouds (Figure 21) were also tested by the state-of-the-art method of Li et al. (2011) grammar-based method of Wan and Sharf (2012), and hierarchical-operational method of Li et al. (2017).

Visual Check
Figure 22 and Figure 23 demonstrate the reconstructed 3D building façade models of experimental data. To visually check the quality of the 3D building roof models, we compared the 3D building models with the lidar data, as Figure 23 shows.

The results of observation demonstrate that the reconstructed 3D façade models provide a reliable appearance, and most of them have a high coincidence with the lidar data. Furthermore, the proposed method can successfully lead to 3D building façade models, even with complex structures.

Furthermore, the building façade model (as building A in Figure 23 shows) was selected as a representative sample for the check, as shown in Figure 24.

The reconstructed 3D model (Figure 24a) was compared with the image (Figure 24b) as captured from the same viewpoint. From a visual standpoint, the constructed model...
Figure 24. Visual comparison of 3D building façade model: (a) Reconstructed 3D building model, (b) Actual building image, (c) and (d) Comparison of model details.

Figure 25. Offset distances between the lidar points and 3D model.
is highly realistic. However, there are several defects in the building model. In Figure 24c, the convex structure in the model (outlined with a red ellipse) can be seen to have been only partly reconstructed. The reason for this error is that during the similarity measurement, such tiny differences between similar tiles are not detected, and the convex structure of the upper floors is repeated to the lower floors. In addition, the tiny structure outlined in Figure 24d is lost during the reconstruction processes, as a result of the edge point extraction method being unable to deal with such tiny structures.

**Accuracy Evaluation**

The accuracy of the reconstructed building model is evaluated based on the offset distance. By taking the original lidar points as the true values, the minimum offset distance is calculated between the lidar points and the meshes of the 3D model. Specifically, the CloudCompare open-source software was used to make this comparison. To improve the efficiency and reduce the statistical error caused by unevenly distributed points, the original points are subsampled with the interval of 0.2 m. The offset distance of the building model is calculated as \( d_{Mi} \). Firstly, a random point \( R_i \) is selected from the true value \( R \). Then, the nearest triangular patch in the building model is searched for and labeled \( M_i \). The offset distance between point \( R_i \) and triangular patch \( M_i \) is determined as \( d_{Mi} \). Finally, \( d = \{d_{M1}, d_{M2}, ..., d_{Mn}\} \) is obtained, and the statistical maximum distance, average distance, and standard deviation are calculated.

The experiment takes 11 building models (as labeled A–K in Figure 23) to evaluate the accuracy, the result of evaluation is shown in Table 1. According to the result, 3D building models have small distance error compared with the evaluation lidar points. The total average offset distance is 0.101 m, and about 99% evaluation points’ distance error is less than 0.454 m.

<table>
<thead>
<tr>
<th>Building Models</th>
<th>Points Number</th>
<th>Average Distance</th>
<th>Standard Deviation</th>
<th>90th Percentile</th>
<th>Maximum Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90960</td>
<td>0.109</td>
<td>0.097</td>
<td>0.434</td>
<td>1.119</td>
</tr>
<tr>
<td>B</td>
<td>41654</td>
<td>0.070</td>
<td>0.090</td>
<td>0.467</td>
<td>0.673</td>
</tr>
<tr>
<td>C</td>
<td>19105</td>
<td>0.094</td>
<td>0.093</td>
<td>0.453</td>
<td>0.840</td>
</tr>
<tr>
<td>D</td>
<td>26063</td>
<td>0.118</td>
<td>0.095</td>
<td>0.451</td>
<td>0.634</td>
</tr>
<tr>
<td>E</td>
<td>11443</td>
<td>0.148</td>
<td>0.112</td>
<td>0.466</td>
<td>1.202</td>
</tr>
<tr>
<td>F</td>
<td>7891</td>
<td>0.109</td>
<td>0.107</td>
<td>0.460</td>
<td>0.937</td>
</tr>
<tr>
<td>G</td>
<td>49618</td>
<td>0.132</td>
<td>0.118</td>
<td>0.466</td>
<td>1.061</td>
</tr>
<tr>
<td>H</td>
<td>19954</td>
<td>0.092</td>
<td>0.098</td>
<td>0.446</td>
<td>0.865</td>
</tr>
<tr>
<td>I</td>
<td>31538</td>
<td>0.087</td>
<td>0.099</td>
<td>0.442</td>
<td>1.016</td>
</tr>
<tr>
<td>J</td>
<td>31518</td>
<td>0.082</td>
<td>0.101</td>
<td>0.443</td>
<td>0.885</td>
</tr>
<tr>
<td>K</td>
<td>44986</td>
<td>0.091</td>
<td>0.098</td>
<td>0.457</td>
<td>0.953</td>
</tr>
<tr>
<td>Total</td>
<td>374730</td>
<td>0.101</td>
<td>0.102</td>
<td>0.454</td>
<td>1.202</td>
</tr>
</tbody>
</table>

In comparison, building B, J, I, K, H, and C (as labeled in the Figure 23) have smaller offset distance compared with the evaluation lidar points, which are less than 0.094 m (as shown in Table 1). The main reason is those six building models have regular facade structures. Experiments show that when the structures of building facade are much simpler or more regular, building models have the lower distance error compared with the evaluation points.

Besides, building A, F, D, G and E (as labeled in the Figure 23) have larger offset distance compared with the evaluation points, which are about 0.109 m to 0.148 m (as shown in Table 1). Among them, Building A and G have complex facade structure, and the extraction of the building facade contours is not as accurate as possible, resulting in certain error between the unrolled building facade points and the fitting wall plane. Building F, D, and E have low level data quality, the model accuracy is influenced by the missing points.

Furthermore, the evaluation is carried out for the entire building and the local parts of Building A (as part A, B, and C in Figure 4) respectively. Table 2 and Figure 25 present the results of this evaluation.

**Table 2 Accuracy evaluation of building model A (m).**

<table>
<thead>
<tr>
<th>Building Model A</th>
<th>Points Number</th>
<th>Average Distance</th>
<th>Standard Deviation</th>
<th>90th Percentile</th>
<th>Maximum Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire</td>
<td>90960</td>
<td>0.109</td>
<td>0.097</td>
<td>0.434</td>
<td>1.119</td>
</tr>
<tr>
<td>Part A</td>
<td>17131</td>
<td>0.168</td>
<td>0.121</td>
<td>0.489</td>
<td>0.757</td>
</tr>
<tr>
<td>Part B</td>
<td>18989</td>
<td>0.054</td>
<td>0.051</td>
<td>0.225</td>
<td>0.553</td>
</tr>
<tr>
<td>Part C</td>
<td>54727</td>
<td>0.109</td>
<td>0.091</td>
<td>0.419</td>
<td>1.119</td>
</tr>
</tbody>
</table>

In partially enlarged drawings of Figure 25, small offsets can be seen in region a. This is caused by the internal wall having a small invagination (about 0.2 m) relative to the boundary of the wall. According to the reconstruction method proposed in this paper, the wall model in this region is generated along the edge points, where the model matches well. In addition, the same situation arises in region b. Small deviations can also be seen in region c. This is because both the inner and outer edges can be detected for the convex structure, and the two layers of edge points lead to an error in the outline fitting. The offset of region d is a result of the complicated structures and noisy points. It is thus difficult to reproduce the building contours exactly, leading to the failure of the reconstruction.

The accuracy was also evaluated within different elevations. The evaluation points of the three building parts (A, B, and C) are individually grouped for each elevation interval (set to 1 m). Then, the average offset distances and standard deviations of the evaluation points in each elevation interval are calculated. The evaluation result is demonstrated in Figure 26, from which we can draw the following conclusions. Part B exhibits a higher accuracy than either part A or C, as already mentioned. The high offset of part B is located on the elevation of the rooftop, where the structure is hollow and there are very few points. Similarly, part C has a large error around the position of the rooftop. In part A, there are clear errors within the [0, 5] elevation intervals, which are mainly caused by the glass doors and windows. In addition, a large amount of fluctuation occurs within the [18, 20] elevation intervals, which are caused by the occlusion from the large structure of part A.
Comparisons with Other Methods

In this section, the comparison of the methods developed by Li et al. (2011) (Method I), Wan and Sharf (2012) (Method II), and Li et al. (2017) (Method III) with our method is presented. These four methods use repetitive structures to consolidate incomplete data and to reconstruct façade models. Method I detected in-plane repetitions with the assistance of 2D photographs, and these 2D images were manually registered with 3D lidar scans. In Method II, candidate repetitive structures were marked manually to complete the missing parts by using a non-local filter. Method III carried out similarity detection procedure after the point cloud was decomposed into several depth planes. Our method can automatically detect repeated structures from multiple façades, some user interactions are required to deal with connection problems and topological errors while merging small tile models into a building façade model. The comparative study is carried out in terms of the quality of the 3D façade models based on the datasets of the Stdh façade (Figure 21a) and Apartment façade (Figure 21b).

Figure 27 compares the original photograph of the Stdh façade and the modeling results obtained using Method I, Method II, Method III, and our method. It can be clearly observed that the façade model reconstructed using our method (Figure 27e) has the highest integrity. As shown in Figure 27b, Method I could not detect the windows and balconies of the façade. Method II could reconstruct some balconies of the façade, but could not reconstruct the structure of the windows (see Figure 27c). As shown in Figure 27(d), Method III generated more accurate boundaries of the balconies, and acquired some detailed structure of the windows, but could not use the repeated structures of the façade. With our method, the self-similarity structures of the façade were detected and applied to acquire the most detailed boundaries and structure of the balconies and windows. Further, a high-integrity façade model could be reconstructed as shown in Figure 27e.

As illustrated in Figure 28, we compared the reconstruction results obtained with our method with those obtained using Methods I, II, and III for the point cloud of the Apartment façade. These four methods can reconstruct most of the elements of the façade. However, Method I (Figure 28b) and Method II (Figure 28c) could not reconstruct the windows of the façade. In comparison, Method III and our method could reconstruct these windows, shown as pink rectangles in Figure 28d and 28e. Unlike Method III, our method could reconstruct some occluded elements by taking advantage of repeated structures, as indicated by some of the windows on the upper and lower part of the façade in Figure 28e.

In addition, the accuracy of our method and the other three methods was evaluated based on the datasets of Stdh façade and Apartment façade. This evaluation was carried out according to the quantitative evaluation conducted by Li et al. (2017), and the results are summarized in Table 3.
Table 3 Accuracy evaluation of the four methods using the two datasets (m).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Methods Tested</th>
<th>Maximum Distance</th>
<th>Minimum Distance</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiDh façade</td>
<td>Method I</td>
<td>1.13</td>
<td>0.02</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Method II</td>
<td>0.48</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Method III</td>
<td>0.45</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Our method</td>
<td>0.39</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Apartment façade</td>
<td>Method I</td>
<td>1.10</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Method III</td>
<td>0.12</td>
<td>0.03</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Our method</td>
<td>0.34</td>
<td>0.01</td>
<td>0.08</td>
</tr>
</tbody>
</table>

As shown in Table 3, the models reconstructed using our method have higher accuracy and lower reconstruction errors than the models reconstructed using the other three methods. It is noteworthy that the Apartment façade model reconstructed using our method has a larger maximum distance than that reconstructed using Method III. This is probably because some missing elements could be reconstructed with our method using repeated structure information from remote regions, which may have influenced the model accuracy.

Conclusions

As mentioned above, there are many challenges to reconstruct high-quality 3D building façade models efficiently from mobile lidar data. In this work, based on the local self-similarities of building façade, we proposed a new modeling method to simplify the complex façade structures, to consolidate the incomplete façade points, and to reconstruct the high-integrity façade models efficiently. Specifically, the dividing and unrolling procedure, as well as the extraction of edge points, can produce a clear and simple flattened point cloud, which reduces the difficulty of data processing for multiple building façades. Besides, a two-direction detection strategy is proposed to discover repetitive structures of building façade, such as adopting a sine wave to fit the periodic structures of building floors along elevation direction. By taking advantage of the repetitions, a restoration procedure is put forward to denoise the sparse outlier points, to complete missing parts using information from remote regions, and to improve the quality of façade points. After that, tile models can be reconstructed in an easy and efficient way based on the restored points, and a high-integrity façade model is reconstructed by combining repeated tile models.

This approach makes use of repeated structures in building’s façades. Thus, the method would not be suitable for the reconstruction of buildings with irregular structures. In addition, façade points may break apart at a deformable position during the unrolling procedure, and would not be recovered until the inverse conversion is applied. Although the edge points are extracted to reduce the amount of data and obtain a clear structure, this may result in a loss of detail. Further work will focus on symmetrical structure detection and a more efficient similarity measurement method for application to the restoration.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (41501456, 41622109), and the National Key R&D Program (2017YFB05004205). Sincere thanks are given for the comments and contributions of the anonymous reviewers and members of the editorial team.

References


Abstract
A primary Landsat constraint is cloud contamination. The goal of this study is to evaluate how increasing Landsat temporal repeat frequency might achieve weekly cloud-free observations. This is a case study which examines eastern United States observations, as a typical representation of the temperate forest biome.

A comparison between simultaneous Landsat and MODIS observations shows that MODIS observations provide a reasonable approximation of observed Landsat cloud cover conditions. Daily MODIS cloud observations are examined, using image compositing, to evaluate how increasing Landsat temporal repeat frequency might reduce weekly Landsat cloud contamination.

Results suggest that for these eastern US locations weekly clear views are best achieved with daily satellite repeat frequency, bi-weekly clear views with 2-day repeat, monthly clear views are achieved with 4-day repeat and seasonal clear views with 8-day repeat. To more fully understand the global impact of cloud contamination on Landsat observations similar studies are needed for the Earth’s other major biomes.

Introduction
Over the last half century, the Earth science community has adopted satellite remote sensing as a means to document land surface dynamics. Landsat and the NOAA Advanced Very High Resolution Radiometer (AVHRR) observatories provided the foundations for this work and stimulated the development of systematic land dynamics monitoring both within and between years (Dethier 1974, MacDonald and Hall 1980, Masek et al. 2013, Townshend 1995). This work continues today with moderate resolution sensors on Landests-7 and -8 and the European Space Agency Sentinels-2a and -2b as well as other similar satellites and sensors (Hagolle et al. 2010, Vermote et al. 2016). These satellite observations have not only revealed the complexity of land surface dynamics, but also how evaluation of these dynamics can be disrupted by other variables including sensor and satellite operations, varying atmospheric conditions and, particularly, the impact of cloud contamination.

There have been many studies which have explored how to detect and mitigate cloud contamination in satellite optical remote sensing measurements (Irish et al. 2006, Platnick et al. 2003, Shenk and Salomonson 1971, Stowe et al. 1987). However, few studies have examined how regional and global cloud cover dynamics interact with satellite observation parameters such as those employed for Landsat-class observations. Such studies have been constrained by the lack of empirical observations suitable for the specifics of a Landsat-type system (i.e., sun-synchronous polar orbit, mid-morning overpass, 16-18 day repeat cycle).

In this study we take advantage of the synergy between Landsat and Earth Observing System (EOS) Terra Moderate Resolution Imaging Spectroradiometer (MODIS) morning overpasses to explore the impact of cloud contamination on Landsat observations.

Background
In 1967, when the US National Research Council “Woods Hole” working group proposed a Global Land Use (GLU) satellite, similar to the Earth Resources Technology Satellite (ERTS aka Landsat) system launched in 1972, they struggled with defining the observation repeat cycle of this system because of the problem of clouds (National Research Council 1969):

The synoptic requirements of GLU pose some problems in the acquisition of imagery for the land surface of the earth, because of cloud cover. Preliminary estimates show that though some parts of the world are cloud covered most or part of the year, 20 to 30 overflights over the course of a year should produce 90 percent or better coverage. Cloud cover becomes a serious problem in obtaining imagery for disaster information, crop reporting, and other uses requiring data at a specific time.

Only a simple understanding of global cloud dynamics over land existed in the early 1960s, including; average global cloud cover is near 50% and different zonal patterns exist in the tropics, mid-latitudes and polar regions (Sellers 1965). In addition, it was generally believed that cloudiness was lowest at sunrise and increased as the sun heated the ground. This led to the placement of the Landsat satellites in a mid-morning orbital pattern – a tradeoff between cloud contamination and solar radiant intensity.

The original ERTS, renamed Landsat-1, was designed with an 18-day repeat cycle which provided ~20 repeat observations per year, at least over the United States.1 This repeat cycle of observations over the continental United States, while other regions of the world were observed less frequently depending upon available power, on-board storage and, occasionally, predicted cloud conditions.

1. From the beginning Landsat only acquired continuous repeat observations over the continental United States, while other regions of the world were observed less frequently depending upon available power, on-board storage and, occasionally, predicted cloud conditions.

Samuel N. Goward, Tatiana V. Loboda and Darrel L. Williams, Chengquan Huang

Photogrammetric Engineering & Remote Sensing
© 2019 American Society for Photogrammetry and Remote Sensing
doi: 10.14358/PERS.85.2.109
cycle was believed to be an effective approach to land monitoring, although based on little available information on regional and global cloud dynamics.

Stoney (1996) made an attempt in the 1990s using newly available meteorological model predictions to evaluate detailed cloud impacts on Landsat observations. His primary conclusion was that 16-day or even 8-day repeat cycles would not be sufficient to acquire clear surface views.

By the early 1990s weather satellite observations made development of the International Satellite Cloud Climatology Project (ISSCP) possible (Rossow and Schiffer 1991). This data set was used to guide development of the Long-Term Acquisition Plan (LTAP) for Landsat-7, an automated procedure focused on the acquisition of less cloud-contaminated observations (Arvidson et al. 2001, Arvidson et al. 2006, Kovalskyy and Roy 2013). A similar approach was adopted for Landsat-8 (Irons et al. 2012).

Since 2001 EOS Terra MODIS has collected high temporal repeat frequency (i.e., daily-at mid-latitude), coarse spatial resolution (i.e., 1 km) similar to the spectral observations of Landsat and in a sun-synchronous polar orbit similar to that used for Landsat satellites since Landsat 4 and near-coincident with Landsats 5 and 7 (Goward et al. 2017, Justice and Townshend 2002). In this study we explore a ten-year record of Terra MODIS imagery to evaluate what Landsat orbital repeat frequency would be needed to achieve varying mission goals such as monitoring agriculture, forests, urbanization and other land use land cover changes.

Study Goal
Satellite-based land-observing systems are designed to acquire “cloud-free” land surface measurements to document land surface conditions for a specific location and over a specific time period. Cloud contamination (i.e., clouds and shadows) prevent such measurements from being acquired. The probability of acquiring a totally cloud-free Landsat image is typically quite low with the exception of desert locations such as the African Sahara. The goal of this study is to evaluate what satellite orbital repeat cycle would be needed to produce a “substantially cloud-free” image over a specific time period. In this study a “substantially cloud-free image” is defined as either a single Landsat image (i.e., 185 by 185 km) with less than 10% cloud cover or a composited Landsat image with less than 10% cloud cover which has been composited from “clear”, non-cloudy pixels that have been acquired over a specified period of observations (e.g. weekly, bi-weekly, monthly, seasonally). A similar approach could be used to assess cloud impacts on the 100 by 100 image tiles used for Sentinel-2 and under consideration for Landsat Level-2 products at USGS EROS (European Space Agency 2018).

We focus this case study on the eastern United States temperate forest ecosystem. The results are used to demonstrate how the MODIS data and similar observations (e.g. SPOT VEGETATION) can be used to evaluate the impact of cloud contamination on collection of Landsat-class “cloud-free” images.

Recent studies by the USGS/NASA Landsat Science Team have identified specific cloud-free temporal repeat frequencies needed to achieve specific thematic goals (see, Table 1) (Roy 2018). These results suggest that cloud-free repeat cycles between one week and one month are sought to meet most science goals. Today understanding of how well Landsat-class operations meet these goals is relatively uncertain.

Approach
As a starting point we evaluated how well MODIS cloud observations can serve as an analog for Landsat cloud observations. To achieve this we compared individual scene Landsat cloud products that were produced in the North America Forest Dynamics (NAFD) project for Landsat-5 and -7 (Huang et al. 2010b, Zhao et al. 2018) with MODIS cloud assessments. The MODIS images were subset to compare directly equivalent area extent with the Landsats and -7 images (see, Figure 1).

The temporal repeat cycles considered were derived from the current Landsat orbital parameters including sun-synchronous near-polar orbit, WRS-2 (i.e., world reference system 2) sixteen-day repeat cycle, a 185 km swath, and a 10 A.M. (±15 min) equatorial crossing. This orbital system provides the opportunity for full global cover every 16-days with minimal orbital swath overlap at the equator and increasing overlap between orbits toward the polar regions. In this study, we assume that the sensor operates continuously over sunlit land. As the increased repeat frequency is simulated in this study via the use of MODIS imagery, we assume additional Landsat satellites would be added and inserted into a constellation of Landsat satellites to optimize temporal spacing between them (see, Table 2). The results reported here are also directly

2. To date Landsat sensor systems typically have been cycled on and off to preserve battery power and minimize heat buildup. Only when the satellite is observing the continental United States has the sensor always been in imaging mode. Elsewhere it is generally only operated on demand, including transmission to international ground stations. For Landsat-7 the LTAP improved global acquisitions but still only “observed always” over all 50 US states. Landsat 8 is now approaching full-time operability globally over sunlit land as it does not require being cycled on and off for heat dissipation as the earlier scanner systems required. (Goward, et.al. 2017)

Table 1. Minimum Landsat-10 observation requirements to enable monitoring of thematic observational areas (Source: Landsat Science Team, August 2018)

<table>
<thead>
<tr>
<th>Monitoring Theme</th>
<th>Temporal resolution (cloud-free)</th>
<th>Spatial resolution</th>
<th>Spectral resolution</th>
<th>Measurement priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>3-5 day</td>
<td>10 meters</td>
<td>VSWIR</td>
<td>Temporal / Spatial</td>
</tr>
<tr>
<td>Albedo</td>
<td>≤ 8-day</td>
<td>10-30 meters</td>
<td>VSWIR</td>
<td>Temporal / Spatial</td>
</tr>
<tr>
<td>Evapotranspiration and Water Use</td>
<td>3-5 day</td>
<td>10 meters</td>
<td>VSWIR</td>
<td>Temporal / Spatial</td>
</tr>
<tr>
<td>Forests</td>
<td>≤ 8-day</td>
<td>10-30 meters</td>
<td>VSWIR</td>
<td>Temporal / Spatial</td>
</tr>
<tr>
<td>LCLU change and vegetation dynamics</td>
<td>3-5 day</td>
<td>10-30 meters</td>
<td>VSWIR</td>
<td>Temporal / Spatial</td>
</tr>
<tr>
<td>Snow and Ice</td>
<td>3-5 day</td>
<td>10 meters</td>
<td>VSWIR (&lt; 1.0 Kelvin)</td>
<td>Spectral / Radiometric</td>
</tr>
<tr>
<td>Urban</td>
<td>≤ 32-day</td>
<td>10 meters</td>
<td>VSWIR</td>
<td>Spatial / Temporal</td>
</tr>
<tr>
<td>Water (inland &amp; coastal)</td>
<td>3-5 day</td>
<td>10-30 meters</td>
<td>VSWIR (&lt; 1.0 Kelvin)</td>
<td>Spectral / Radiometric</td>
</tr>
</tbody>
</table>
applicable to Landsat-8 as its nominal orbital parameters are the same as Landsats-4 – -7.

Table 2. Landsat Repeat Cycle Considerations

<table>
<thead>
<tr>
<th># Satellites</th>
<th>Repeat Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>16 days</td>
</tr>
<tr>
<td>Two</td>
<td>8 days</td>
</tr>
<tr>
<td>Four</td>
<td>4 days</td>
</tr>
<tr>
<td>Eight</td>
<td>2 days</td>
</tr>
<tr>
<td>Sixteen</td>
<td>1 day</td>
</tr>
</tbody>
</table>

Conducting this type of assessment for Sentinels-2a and -2b is more complex. Sentinel-2a has a 10-day repeat cycle but with a swath twice as wide as Landsat’s-4 – -8, the comparative repeat frequency to Landsat is higher, approaching nearly 5 days. This would approximately produce the same results at the 4-day repeat frequency reported in this study. The addition of Sentinel-2b in 2017 increases the Sentinel-2 repeat frequency to nearly 2-3 days relative to Landsat and should produce nearly biweekly (~15 day) clear views. However, evaluating this Sentinel-2 configuration for cloud contamination is not simple. The cloud conditions across the Sentinel-2a and b wider swath are not independent. Assessment of such correlated cloud observations would require considerable further analysis, beyond the scope of this study.

Study Sites

This study examined three Landsat WRS-2 scene locations in the eastern United States; Maryland (p15r33), Pennsylvania (p17r31), and Indiana (p21r32) (see, Figure 2 and Table 3). These sites are representative of humid mid-latitude locations, part of the temperate and mixed forest biome found throughout the mid-latitudes of the Earth where agriculture and forestry land use practices dominate (see, Figure 3) (Olson et al. 2001).

Terra MODIS as an Analog of Landsat Cloud Observations

Terra MODIS observations are used in this study to approximate the scene-level cloud conditions that a Landsat sensor would observe over orbital repeat cycles ranging from 1 to 16 days. The Terra MODIS sensor has been collecting coarse spatial resolution (i.e., 250 – 1000 m) optical imagery approximately 30 minutes after Landsat-7 observations along the same nominal ground track since early 2000 (Williams et al, 1994).
Given the much broader ~2300 km swath of a Terra MODIS sensor in comparison to Landsat’s 185 km imaging swath, Landsat-5 observations are also captured in the same MODIS imaging swath (Nishihama et al. 1997). Interestingly the Landsat WRS-2 orbital pattern results in a “skipping” approach (see, Figure 4) (Freden and Gordon 1983), whereby the adjacent orbital tracks are not temporally sequential (Goward, et al. 2017). Fortuitously, because of this skipping orbital pattern and the fact that Landsat-5’s orbital phasing is 8 days off-set from Landsat-7/EOS Terra MODIS, Landsat-5’s orbital ground track is just one track west of (i.e., adjacent to) the Landsat-7/ EOS Terra MODIS orbital track (see, Figure 4). As an example, Day 8 in the graphic would also represent Day 1 for Landsat-5. Since the MODIS sensor view angle is ± 55° it simultaneously images across seven hours of local solar illumination.

Table 3. Summary of dominant land cover types and topographic attributes for three study sites.

<table>
<thead>
<tr>
<th></th>
<th>MD</th>
<th>PA</th>
<th>IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant land cover types (%) *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest</td>
<td>24.8</td>
<td>62.9</td>
<td>0.6</td>
</tr>
<tr>
<td>cropland and cropland mosaic</td>
<td>49.1</td>
<td>29.9</td>
<td>95.8</td>
</tr>
<tr>
<td>water</td>
<td>14.2</td>
<td>6.3</td>
<td>0.1</td>
</tr>
<tr>
<td>urban and built-up</td>
<td>8.6</td>
<td>0.5</td>
<td>3.4</td>
</tr>
</tbody>
</table>

| Land surface elevation (m)** |       |       |       |
| minimum                 | 0     | 163   | 110   |
| maximum                 | 1052  | 803   | 392   |
| mean                    | 95.4  | 434.0 | 256.5 |

* from MODIS Land Cover product (MCD12Q1)
** from Shuttle Radar Topography Missions Digital Elevation Model (SRTM-DEM)

Figure 2. Location of three Landsat scene equivalents within the MODIS tiles h11v04, h12v04, h11v05, h12v05 in Maryland (WRS-2 p15r33)-, Pennsylvania (WRS-2 p17r31), and Indiana (WRS-2 p21r/32).

Figure 3. Distribution of the Earth’s major terrestrial biomes (Olson, et al. 2001). The eastern US study sites are shown as 3 small red boxes within the ‘green’ Temperate Broadleaf and Mixed Forests biome in the map above. Other biome regions could have quite different cloud dynamics than those represented by the sample locations examined in this study.
Cloud Assessment

Cloud analysis for Landsat was carried out using an automated cloud masking algorithm implemented as part of the Vegetation Change Tracker (VCT) algorithm developed for the NAFD study (Huang et al. 2010a). This approach provides results equivalent to other similar cloud detection approaches (Zhao, et al. 2018). The VCT cloud masking algorithm was designed based on the fact that clouds are in general bright and cold. It uses a set of cloud boundaries in a spectra-temperature space to separate clouds from clear view pixels. It has been used to produce cloud cover estimates for tens of thousands of Landsat images acquired over the United States and globally (Gutman et al. 2013). Schleeweis et al. (2016) demonstrated that cloud estimates derived using this algorithm were comparable to those derived using the Landsat-7 Automated Cloud Cover Assessment algorithm (Irish, et al. 2006). A detailed description of this algorithm has been provided by Huang et al. (2010b).

The MODIS cloud mask used in this study is represented by the 1 km MOD35 component extracted from the daily Terra MODIS surface reflectance collection 5 product MOD09GA (Vermote et al. 2002). This study is based on a ten-year MOD09GA data record primarily including data between 1 January 2003 and 31 December 2012. However, data from December 2002 is added to the record when assessing “seasonal” compositing period that includes December – January – February months for “winter.” MOD35 is a standard dataset based on a suite of radiance (for infrared) and reflectance (for near infrared (NIR) and visible) thresholds that take into account underlying land cover properties, atmospheric column constituents, and sensor viewing geometry (CIMSS 2006, Ackerman et al. 1998). In our initial analysis we explored all cloud masking-related bits of the MOD09 product, including “mixed” clouds from the MOD35 cloud mask and all internal MOD09 cloud bits which are conceptually different from the MOD35 product and thus present a nearly-independent assessment of cloud cover compared to the MOD35 mask (Leienkugel et al. 2013). However, the inter-comparison of the various components of MOD09, MOD35 and their combinations with the Landsat-based cloud masks within our study area has shown that MOD35 cloudy class (MOD09 cloud state bits (0-1) in the “01” bit form corresponding to “cloudy”) has the closest statistical relationship at the WRS-2 scene level.

As noted previously, a “substantially cloud-free image” in this study is either a single Landsat image (i.e., 185 by 185 km) with less than 10% cloud cover or a Landsat image equivalent with less than 10% cloud cover which has been composited from “clear,” (i.e., non-cloudy, pixels that have been acquired over a specified period of observations — e.g. daily, weekly, bi-weekly, monthly, seasonally).

With pixel-based digital observations, the concept of compositing two or more repetitive observations to produce synthesized cloud-free observation has been realized. To conduct such compositing requires multiple, repetitive observations of the same location, as near as possible at the same time of day and solar illumination angle. There are many alternative “compositing” approaches including maximum Normalized Difference Vegetation Index (NDVI), highest surface temperature, minimum red, etc. (Chuvieco et al. 2008, Holben 1986). Compositing protocols of various types have been previously applied to AVHRR and MODIS images to reduce scene cloud contamination (Chuvieco, et al. 2005, Holben 1986). Further the web-enabled Landsat data (WELD) process has applied the same principle to Landsat observations (Roy et al. 2010).

In this study we obtained “composited” cloud-free observations per defined observation period (e.g. daily, weekly, bi-weekly, monthly, and seasonally) by sticking individual cloud masks into a data cube, where x and y represent the geospatial extent and z represents the number of days in the compositing period. A composited pixel is deemed cloud-free if that pixel location has at least one clear view observation during the chosen compositing period (e.g. weekly, bi-weekly, monthly, and seasonally). Otherwise the pixel is considered cloudy. Missing data was excluded from this assessment. This approach is more conservative than simpler one-step compositing approaches in that cloudy pixels are identified and rejected before selecting clear surface observations. This results in a more reliable indicator of “cloud-free” conditions.

Statistical Evaluation

Typical cloud assessments generally consider the average (AVG) and variance of observed cloud conditions over a specific time period (e.g. week, month, season, year). Such an assessment is not suited to evaluate achieving mission goals since the desired goal is to achieve one “cloud-free” image...
Figure 5. (a) For all Landsat scenes examined the fractional cloud cover evaluated in the case-study scenes compared well with the equivalent MODIS subsets. (b) For Landsat-5 scenes (MODIS off-nadir to the west), the relationship is not as strong and more biased than the Landsat-7 (MODIS nadir) scenes.

Figure 6. (a) Ten-year average weekly cloudiness and (b) Ten-year average composited weekly cloudiness for MD, PA and IN study sites.
for a specified time period. In this study we employ a “clear view frequency” (CVF) metric, which categorized the results for a specific location and time period as (a) persistently cloudy, (b) composited clear or (c) clear. When either (b) or (c) are achieved the mission is considered successful for that time period. Examples of AVG cloudiness, AVG composited average cloudiness, composited CVF and clear CVF were compared to demonstrate the importance of composting and assessing mission success.

Results

Landsat and MODIS Cloud Assessment Comparison

For the selected eastern US Landsat WRS-2 sites the comparison of scene-cloud contamination from equivalent Landsat and MODIS observations demonstrated they are strongly related with an overall R² > 0.9 (see, Figure 5a). We anticipated that the relationship between Landsat-5 and MODIS scenes would be weaker because of the time delay between these observations (see, Figure 1 caption). The MODIS nadir observation time is 10:30 A.M. LST but the Landsat-7 overpass is 10:00 A.M. LST (see, Figure 5a). For Landsat-5 images, also acquired at 10:00 A.M. LST, the equivalent Terra MODIS observations are viewed off-nadir, one orbital track to the west with an effective LST of 9:30 A.M. as well as far-off-nadir as noted in Figure 1c (see, Figure 5b). As anticipated the Landsat-5 observations are not as strongly related to the MODIS observations with a R² = 0.89 versus the R² = 0.95 for Landsat-7. Nevertheless, results indicate that Terra MODIS observations provide a good approximation of what a Landsat system would observe over a range of temporal orbital repeat cycles.

Comparison of Site Cloud Statistics: The 10-year average weekly cloudiness derived from single-day, non-composited scenes produces reasonably stable results over a calendar year with a standard deviation of ± 20% (Figure 6a). This scene average cloudiness suggests that the probability of acquiring one completely clear-view per week is low with summer-time averages typically less than 50% and the winter season averages 40% or less. With image composting applied, the potential for acquiring a clear scene substantially improves, typically the improvement is greater than 50% (Figure 6b).

There is no question that some form of image compositing is needed to achieve satellite optical remote sensing goals.

A more detailed assessment of mission success is provided by the CVF assessment (see, Figure 7) results, similar to the composited image average cloud cover. show that as the average cloud cover decreases, the probability of achieving a clear view, either single scene or composited declines even more rapidly (see, Figure 8). At 0% cloud cover the CVF is 100% whereas by 30% to 50% composited cloud cover the CVF is 20%. The CVF metric provides a conservative estimator of mission success.

The weekly average CVF was computed for each orbital repeat cycle being considered (see, Table 2). The Maryland weekly CVF values from 1-day to 16-day repeat cycles are shown in Figure 9. For the 1-day repeat cycle, with the exception of two weeks in late March-early April, all weeks average better
than 90% CVF throughout the year. This indicates that a daily repeat Landsat observatory will generally produce composited weekly images with <10% cloud cover. Two-day repeat reduces CVF to ~75% and four-day to ~50%. For the eight-day repeat cycle (what the community has had fortuitously throughout much of the Landsat era), the CVF values are generally less than 30%, indicating that the chances of getting a clear view each week are low. Finally, the sixteen-day repeat cycle simply halves the eight-day results further as it only occurs ~ every two weeks. With an orbital repeat frequency of 8-days the Maryland CVF values are below 30%. A decrease in the orbital repeat frequency to 16-day (i.e., single satellite) halves the coverage (i.e., 2 weeks versus 1 week) but produces similar (<30%) CVFs.

Similar results are derived for the Pennsylvania and Indiana sites (see, Figure 10). As the satellite repeat frequency decreases, the CVF values decrease. The results from all three sites are similar although the Pennsylvania site is more impacted by residual cloud cover than the other two sites. This appears be a function of the topography in western PA and “lake effect” associated with being predominately downwind of a large water body, Lake Erie.

Interestingly, computing the average annual CVF values of weekly results for 1-, 2-, 4-, 8-, 16-day repeat coverage for the three sites produces similar declines in CVF as a function of orbital repeat frequency values for all sites. The exception is that each site has different initial one-week values (which is likely a function of geography (e.g. topography) and certainly driven by winter month values) (see, Figure 11).

A regional overview of the summer season CVF results from the eastern US MODIS swath clearly points out the topographic influence on cloud probability (Figure 12). This regional map for average conditions observed during the summer reinforces the patterns we observed in weekly detail for the three study sites.

In general these results suggest that for these eastern US locations weekly clear views are best achieved with daily satellite repeat frequency, bi-weekly clear views with 2-day repeat, monthly clear views are achieved with 4-day repeat and seasonal clear views with 8-day repeat (Figure 11). Sixteen day repeat only reliably produces one clear view per year, most likely outside the winter months. However it is important to
Landsat and MODIS, introduce significant residual errors in
services used within and between the two sensors explored here,
study include: Do the variations in cloud detection approach-
approaches which address wider swaths and other observatory
AWiFS as well as to evaluate newly proposed future system
culture contamination in current Landsat sensor configurations.
assess all regions of the globe. This study should be con-
ersters and polar regions will produce quite different outcomes.
other major ecoregions of the Earth. The humid tropics, des-
temperate forest regions elsewhere across the Earth noted in
Figure 3. These conclusions are unlikely to be applicable to
and internationally. There may also be other geographic and
seasonal factors that underlie observed patterns that have not
been discovered in this study.

Summary and Discussion
One of the most important goals that the Landsat observation
system should attempt to achieve in the next decade is a sub-
stantial improvement in temporal frequency of observations
across the globe. The accumulation of Landsat observations
since mid-1972 has come close to achieving annual refresh of
global land dynamics, particularly in less cloud-prone regions
of the Earth. However, monitoring within-year seasonal land
vegetation dynamics is currently not within the purview of
the Landsat system.

The results reported here are most reliably indicative of
conditions found in the eastern US and less so for similar
temperate forest regions elsewhere across the Earth noted in
Figure 3. These conclusions are unlikely to be applicable to
other major ecoregions of the Earth. The humid tropics, des-
erts and polar regions will produce quite different outcomes.
However, the analysis protocol employed here is suitable to
assess all regions of the globe. This study should be consid-
ered a prototype of a more comprehensive assessment of
cloud contamination in current Landsat sensor configurations.

To assess more complex systems such as Sentinel 2 and IRS
AWIFS as well as to evaluate newly proposed future system
designs that consider use of satellite constellations and vari-
able sensor configurations, would require more sophisticated
approaches which address wider swaths and other observatory
innovations (Goward et al. 2012, Hansen et al. 2006, Williams
et al. 2013).

Some interesting issues that have emerged during this
study include: Do the variations in cloud detection approach-
es used within and between the two sensors explored here,
Landsat and MODIS, introduce significant residual errors in
the results? Recent work being carried out by Landsat Sci-
ence Team members may help to answer this question. There
is also the fact that areas with more topography experience
more persistent cloud cover than regions with less relief.
This needs to be evaluated more thoroughly both in the US
and internationally. There may also be other geographic and
seasonal factors that underlie observed patterns that have not
been discovered in this study.

Acknowledgements
This study has been in progress for more than a decade. The
need for this analysis first emerged while Sam Goward and
Darrel Williams served on the USGS Earth Resources Observa-
tions and Science National Satellite Land Remote Sensing
Data Archive (NSLRSDA) Federal Advisory Committee. The
work continued while Drs. Goward and Williams were mem-
bers of the USGS/NASA Landsat Science Team (2006-11). Fund-
ing to complete the work was provided by the NASA Landsat
Project Science Office, under a contract to Global Science &
Technology, Inc. A wide range of USGS and NASA participants
in the Landsat missions provided insights and support over this
time period.

References
from clouds with MODIS. Journal of Geophysical Research
103(D24):32141-32157.
Arvidson, T., J. Gasch and S. N. Goward. 2001. Landsat 7’s long term
acquisition plan - An innovative approach to building a global
Arvidson, T., S. N. Goward and D. J. Williams. Characterizing the
Landsat Global long-term record. In Proceedings of the 13th
Australasian Remote Sensing and Photogrammetry Conference,
held in Canberra, Australia, 21-24 November 2006. Canberra,
ACT: Spatial Sciences Institute.
Cooperative Institute for Meteorological Satellite Studies (CIMSS).
2006. Discriminating clear-sky from cloud with MODIS:
Algorithm Theoretical Basis Document (MOD35). Madison:
University of Wisconsin-Madison.
Assessment of multitemporal compositing techniques of MODIS
and AVHRR images for burned land mapping. Remote Sensing of
Dethier, B. E. 1975. Satellite sensing of phenological events. Ithaca,
N.Y.: Cornell University Agricultural Experiment Station.
sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/
Falls Church, Va.: American Society of Photogrammetry.
Hierarchical Bayesian Model Based on Robust Fixed Rank Filter for Fusing MODIS SST and AMSR-E SST

Yuxin Zhu, Emily Lei Kang, Yanchen Bo, Jinzong Zhang, Yuexiang Wang, Qingxin Tang

Abstract

Spatiotemporal complete sea surface temperature (SST) dataset with higher accuracy and resolution is desirable for many studies in atmospheric science and climate change. The purpose of this study is to establish the spatiotemporal data fusion model, the Hierarchical Bayesian Model (HBM) based on Robust Fixed Rank Filter (R-FRF), that merge Moderate Resolution Imaging Spectroradiometer (MODIS) SST with 4-km resolution and Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) SST with 25-km resolution through their spatiotemporal complementarity to obtain fusion SST with complete coverage, high spatial resolution, and fine spatial pattern. First, a bias correction model was applied to correct satellite SST. Second, a spatiotemporal model called R-FRF was established to model potential spatiotemporal process of SST. Third, the R-FRF model was embedded in the hierarchical Bayesian framework, and the corrected MODIS and AMSR-E SST are merged. Finally, the accuracy, spatial pattern and spatial completeness of the fusion SST were assessed. The results of this study are the following: (a) It is necessary to carry out bias correction before data fusion. (b) The R-FRF model could simulate SST spatiotemporal trend well. (c) Fusion SST has similar accuracy and spatial pattern to MODIS SST. Though the accuracy is lower than that of the AMSR-E SST, the fusion SST has more local detail information. The results indicated that fusion SST with higher accuracy, finer spatial pattern, and complete coverage can be obtained through HBM based on R-FRF.

Index Terms: Hierarchical Bayesian Model based on R-FRF; MODIS SST; AMSR-E SST; scale transformation; local variance.

Introduction

The sea surface temperature (SST) is one of the important parameters in coupling the ocean and atmosphere through exchanges of heat, momentum, moisture, and gases (Donlon et al. 2002, Zhu et al. 2015). Currently, spatiotemporal complete SST dataset with higher accuracy and spatiotemporal resolution is desirable for many studies in atmospheric science and climate change. Though satellite-derived SST products with different spatiotemporal resolutions have a more completely spatial coverage than other observations do, the spatiotemporal resolution and spatial coverage of satellite-derived SST products from a single sensor are limited. The Global Ocean Data Assimilation Experiment has initiated a pilot project to develop SST products with high spatiotemporal resolution through integrating existing satellite products (Donlon 2001).

In recent years, the researchers have developed a series of quantitative fusion methods for the integration of various remote sensing products, such as the integration of SST (Guan et al. 2004, Gao 2010, Li et al. 2013), leaf area index (LAI) (Li et al. 2013, Wang et al. 2011), Normalized Difference Vegetation Index (NDVI) (Busetto et al. 2006, Hwang et al. 2011, Rao et al. 2015), aerosol (Chatterjee et al. 2010, Loyola et al. 2012, Tang et al. 2016), snow water equivalent (Durand et al. 2008, Foster et al. 2011, Gao et al. 2010, Kongoli et al. 2007), ocean color (Kwiatkowska et al. 2002, Maritorena et al. 2010, Pottier et al. 2006), soil moisture (Yilmaz et al. 2012), Surface Reflectance (Gao et al. 2015, Zhu et al. 2016) and so on. The methods that have been applied can mainly be categorized into two groups: data assimilation technique and spatiotemporal fusion technique of multisource data.

Data assimilation technique is used to derive accurate estimations of the current and future states of the system, together with estimations of the uncertainty in the estimated states by observations in combination with a dynamic system model (Nichols 2010). This method has been widely used in the parameter estimation of atmospheric science, Marine science and land surface process. For example, Durand, Molotch and Margulis (2008) reconstructed SWE in the Rio Grande headwaters by combining time series observations from Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat Enhanced Thematic Mapper (ETM+) with a spatially explicit snowmelt model. Data assimilation technique has clear physical mechanism. It can make full use of the satellite observations with different spatiotemporal resolution, and can accurately simulate the evolution of the process with time because it combines remote sensing observations with dynamic process model, such as crop growth model, the ecological
process model, hydrological dynamic model, atmospheric numerical model, the Marine dynamic model and the land surface process model, etc. However, for the data assimilation technique, the dynamic model based on the physical mechanism usually requires many observations as driving parameters. For example, for the soil moisture estimation, the soil type, heterogeneous vegetation, atmospheric dynamic parameters and so on are required as driving parameters. The limitation of parameters availability leads to some difficulties in the application of data assimilation technique.

Spatiotemporal fusion method based on multisource remote sensing products is developed in recent years. This method combines remote sensing data with statistical methods, such as Multi-resolution Tree (MRT) (Wang et al. 2010), Empirical Orthogonal Function (Wang and Liang 2011), wavelet analysis (Zhang 2006), least square method (Yilmaz, Crow, Anderson and Hain 2012), artificial neural network technique (Loyola and Coldewey-Egbers 2012), Bayesian Maximum Entropy (BME) (Li, Bo and Chen 2013, Li, Bo, Zhu, Guo, Bi and He 2013, Tang, Bo and Zhu 2016), spatiotemporal decomposition (Zhu et al. 2018) and so on. The MRT method can fill the missing value, reduce the error and get the results similar to the optimal interpolation, and the efficiency is better than the optimal interpolation, but it is necessary to set strict conditional assumptions in the process of the scale transformation. Although wavelet analysis can keep the details of the high-resolution data, fusion data with complete coverage cannot be obtained by this method. Least square method that does not require any user-defined parameter assumptions is used to obtain improved soil moisture products along with associated uncertainty estimation using triple collocation by merging soil moisture data from numerical model, thermal infrared remote sensing products and microwave remote sensing products. The main limitation of this method is that uncertainty estimation can be obtained only if we do not express the local spatiotemporal variation. BME is a nonlinear spatiotemporal geo-statistical method (Al 2000, Christakos et al. 2001) that can theoretically integrate data with different accuracies from different sources (Li, Bo, Zhu, Guo, Bi and He 2013). Although fusion data with high accuracy and more local details existing in fine resolution data can be achieved, and the problem of scale transformation can also be resolved through BME, the spatiotemporal process model of latent true SST cannot be expressed, and the computing efficiency is low, especially for massive satellite data.

There are many other methods to fill in gaps in satellite-derived SST products, including the optimal interpolation (Chao et al. 2009, Reynolds et al. 1994, Reynolds et al. 2002, Reynolds et al. 2007), objective analysis (Guan and Kawamura 2004, Potier, Garçon, Larnicoul, Sudre, Schaeffer and Le Traon 2006), blended analysis (Reynolds 1988), three-dimensional (3D)-variation assimilation (Zheng et al. 2009). These methods show different feasibility due to a different constraint condition as described by Zhu, Kang, Bo, Tang, Cheng and He (2015). However, none of these methods can simultaneously use a priori knowledge, express uncertainty of the model and parameters quantitatively, match spatiotemporal scale seamlessly or simulate true value of a potential spatiotemporal process of SST. In this paper, we develop a new fusion model, the hierarchical Bayesian fusion model (HBM) based on the robust fixed rank filter (R-FRF), to merge infrared SST and microwave SST by a time series empirical model. This model can use a priori knowledge, express uncertainty quantitatively, match spatiotemporal scale seamlessly and simulate true value of a potential spatiotemporal process of SST simultaneously, and can make full use of the complementarity in time and space of two data. Although spatiotemporal fusion method based on spatiotemporal data-fusion (STDF) developed by Hai et al. (2014) is similar to that in this paper, fusion data obtained by STDF cannot keep the details of the original data. The advantages of the HBM based on R-FRF are the following: (1) it is easy to implement in the complex, high-dimensional spatiotemporal SST settings. (2) It can account for the various types of uncertainty from different sources, such as SST observation, sampling, models, and parameters. (3) It achieves the scale transformation seamlessly between MODIS and AMSR-E. (4) To a certain extent, it can reduce the uncertainty of the fusion data through the constraint of the latent spatiotemporal process of SST.

The hierarchical Bayesian model has existed for a long time in statistics (Berliner et al. 2000, Berliner et al. 2003, Gelfand et al. 2001, Milliff et al. 2011, Song et al. 2014, Vanem et al. 2012, Wikle et al. 1998, Wikle et al. 2001, Wikle 2002, Wikle 2003, Wikle et al. 2003, Wikle et al. 2006, Wu et al. 2013). However, in the remote sensing realm, its application is limited. Guo (2010) used hierarchical Bayesian model to merge MODIS SST with 4-km resolution and AMSR-E SSTs with 25-km resolution. He only achieved fusion data with a coarser resolution by converting fine resolution into coarser resolution. The hierarchy is manifested through the nesting parameters and super parameters rather than a general model of the potential process of SST. Our method emphasizes the seamless fusion of multisource and multiscale remote sensing products. HBM based on R-FRF focuses on how to obtain complete spatiotemporal SST data with fine spatial pattern, how to minimize cloud contamination and land effects by constraint from the potential spatiotemporal process model of SST.

At present, infrared and microwave bands are used to generate satellite SST, such as the Advanced Very High-Resolution Radiometer (AVHRR), the Moderate Resolution Imaging Spectroradiometer (MODIS), the Along-track Scanning Radiometers (ATSR), and the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E). All of these instruments can provide long time series of SST products, but these products have different properties in terms of spatial resolution, accuracy or spatial integrity. Infrared data products have high spatial resolution, but spatial integrity is poor due to cloud contamination. Microwave data products have high spatial continuity due to their cloud penetration but lower spatial resolution (Zhu, et al. 2015). Thus, by borrowing strength over space from infrared and microwave SST products, we are expected to provide a new generation of satellite SST data products with increased accuracy, consistency and resolution by data fusion (Donlon, et al. 2002). Therefore, in our study, we focus on fusion of multiple satellite-derived SST data products, as derived by MODIS and AMSR-E, to obtain spatial complete SSTs with fine spatial pattern, high accuracy and spatial resolution.

The remainder of this article is organized as follows: In Section 2 we describe the study region, data used in this paper and data preprocessing procedure. In Section 3 we outline the method developed in this paper, including bias correction model and the new fusion model, a hierarchical Bayesian model based on Robust Fixed Rank Filtering. We present in Section 4 the sample results of HBM and the assessment of the fusion SSTs including the spatial completeness assessment, accuracy assessment, spatial pattern assessment and the fusion uncertainty assessment. Discussion is presented in Section 5 and Section 6 is conclusions.

Data and Data Preprocessing

Study Region

The geographical area of interest in our study is the rectangular area from 30°S to 46°N and from 30°E to 180°E which covers the joining area of Asia and the Indian-Pacific Ocean
(Guoxiong et al. 2006, Zhu et al. 2013), as illustrated in Figure 1 (grey represents land, white represents ocean). The study region selected in our study is a key area for short-term climate variation and prediction in China (Zhu et al. 2013).

**Figure 1.** Study region (30°S to 45°N, 30°E to 180°E) and spatial distribution of the drifting buoy. Gray and white areas represent land and ocean, respectively; blue points indicate the locations of the drifting buoy.

**Moderate Resolution Imaging Spectroradiometer Sea Surface Temperature (MODIS SST)**

One of the satellite-derived SSTs used in this paper is MODIS Aqua night level-3 mapped products with 8-day time scale and 4-km spatial resolution at local time 1:30 A.M. in 2003. They are provided by Zhu, et al. (2015). Data at 1:30 A.M. make the effect of diurnal warming of the surface ocean minimized. The higher spatial resolution is selected to obtain fusion SST with higher spatial resolution. MODIS SST with 8-day time scale is selected because they provide better coverage than the daily product and their time scale is not long to ignore the local variation with time. Moreover, the weekly or 8-day composite satellite SST products are typically used in related scientific studies (Babin et al. 2004, Barre N et al. 2006, Castelao et al. 2006, Li et al. 2014, Rao et al. 2006, Jorgetti et al. 2014, Zhu, Kang, et al. 2015).

MODIS SSTs are preprocessed by method suggested by Zhu, et al. (2015) which includes transformation of digital number (DN) into real SST and gross quality control. The other satellite-derived SSTs come from AMSR-E. AMSR-E on board the National Aeronautics and Space Administration’s (NASA’s) Aqua satellite, launched on 4 May 2002, was the first microwave radiometer capable of accurately measuring global through-cloud SSTs. We select the AMSR-E SSTs because they have higher spatial coverage and offset the spatial incompletion of the MODIS SSTs, and they can describe the spatial trend of SST at the coarse scale. AMSR-E datasets are provided as daily maps, 3-day mean maps, weekly mean maps, and monthly mean maps. Time series data started from the 152th day, 2002. In our study, we select version-3 Optimal Interpolation (OI) daily SST products in a 0.25° by 0.25° grid (25 km) in 2003, which were produced and provided by the Remote Sensing Systems supported by NASA. Using a simple empirical model of diurnal warming which depends on solar insolation, wind speed and local time of observation (Gentemann et al. 2003), the AMSR-E daily SSTs were ‘normalized’ to a daily minimum SST, defined to occur at approximately local time 8:00 A.M.. Details are available at the Remote Sensing Systems website.

**Advanced Microwave Scanning Radiometer for EOS Sea Surface Temperature (AMSR-E SST)**

AMS-R-E measures the sub-skin temperature at approximately 1 mm depth (Dong et al. 2006). Pixels flagged as 255 represent land, 251, 253 and 254 represent missing data, 252 represent sea ice, and 0 to 250 represent valid SST data. Digital Number is converted into valid SST based on Equation (1):

\[
SST_{AMS-R-E} = 0.15 \times DN - 3.0
\]

where SST \(_{AMS-R-E}\) represents converted temperature (°C), DN represents the Digital Number of the pixel, and the slope coefficient 0.15 and intercept -3 are obtained from specification on ftp://ftp.discover-earth.org/sst/daily/. The valid value of AMSR-E SST is between -3°C and 34.5°C (http://www.ssmi.com/amsr/amsr_data_description.html). So pixels less than -3°C and more than 34.5°C are screened out.

Drifting Buoy Sea Surface Temperature SST

In our study, we use the drifting buoy SSTs in 2003 which are also provided by Zhu et al. (2015) as the benchmark to correct satellite SSTs and to validate the fusion data. We select the preprocessing method suggested by Zhu, et al. (2015) which includes a simple gross error quality check and the time composite method. For quality-controlled SSTs, to minimize the effects of the ocean diurnal warm layer, we choose the Coordinated Universal Time (UTC) 0:00 data over regions 30°E – 105°E while over regions 105°E – 180°E we choose the UTC 18:00 data. The number of matching points at the MODIS scale is 41,843, at the AMSR-E scale is 32,880. Considering there is no obvious visual difference in spatial distribution between 4-km scale and 25-km scale, an example of the 8-day averaged drifting buoy SST with 4-km spatial resolution in 2003 is only presented in Figure 1.

**Methods**

The data fusion procedure involves filling in the gaps in data and a combination of two data by their complementarity in spatial resolution and completeness to improve their availability. We develop an HBM based on R-FRF to merge MODIS SSTs and AMSR-E SSTs. Seamless scale transformation is introduced into fusion model. HBM based on R-FRF allows the information incorporation from diverse sources and can account for statistical uncertainty which is expressed by the conditional probability. R-FRF process model which describes the spatial trend of SST achieves the computation of massive data by spatial dimension reduction and makes full use of spatiotemporal dependence of data. We can obtain spatial complete data through R-FRF. The estimation by R-FRF is put into an hierarchical Bayesian model as a prior mean. However, through the fusion method only the systematic and stochastic errors of satellite data cannot be reduced. So we applied a data-driven
bias-correction model to correct satellite SSTs before fusion implementation (Zhu, et al. 2015). Then HBM based on R-FRF is applied to the bias-corrected MODIS and AMSR-E SSTs to obtain spatial complete fusion data. The general process is shown in Figure 2

**Bias Correction**

Original Satellite-derived SST is biased. These biases may be caused by instability of sensors and contamination of cloud, water vapor, aerosols (Dong, et al. 2006, Jones et al. 1996). At the same time, they are spatially heterogeneous (Zhu, et al. 2015). In addition, there is another bias caused by differences in observation depth between satellite-derived SST and buoy drifting SST. HBM itself cannot reduce these biases. So a bias-correction procedure suggested by (Zhu, et al. 2015) is applied to the original MODIS SST and AMSR-E SST before the HBM is implemented. First, the relationship between MODIS SST, AMSR-E SST and the aggregated drifting buoy SST is investigated by a scatter diagram, respectively (Figure 3 (a) and Figure 4 (b)).
(a)). MODIS SSTs and AMSR-E SSTs show different characteristics. Over cool regions, MODIS SSTs are generally lower than drifting buoy SSTs while AMSR-E SSTs are higher than drifting buoy SSTs. Over warm regions, satellite-derived SSTs and drifting buoy SSTs are distributed more symmetrically along the 1:1 line. Then, the bias correction model is built over two regions, cool and warm, respectively. The critical value between cool and warm regions is determined using an iterative procedure (Zhu, et al. 2015): first, an empirical temperature based on the scatter diagram as an initial critical value is determined and the bias-correction equation based on this empirical value is built. Then the mean bias of bias-corrected MODIS and AMSR-E SSTs is calculated. Second, the initial critical value plus or minus 0.1 as the new critical value is used to build the new bias correction equation. The bias of the satellite SSTs corrected by the model based on the new critical value is calculated. This procedure is iteratively repeated over an empirical critical value of plus or minus 2°C until the critical value that comes up with the minimum bias of the corrected satellite SSTs is obtained (Zhu, et al. 2018). Based on the iterative procedure, the selected critical values in this paper are as follows: MODIS SST is 11°C and AMSR-E SST is 13°C. Part of regression coefficients is shown in Table 1.

Table 1. Partial weekly regression coefficients between MODIS SST, AMSR-E SST and Drifting Buoy SST.

<table>
<thead>
<tr>
<th></th>
<th>MODIS 11°C</th>
<th>MODIS &lt;11°C</th>
<th>MODIS ≥13°C</th>
<th>MODIS &lt;13°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st week</td>
<td>slope: 1.006</td>
<td>0.874</td>
<td>0.999</td>
<td>1.068</td>
</tr>
<tr>
<td></td>
<td>intercept: 0.005</td>
<td>1.316</td>
<td>−0.059</td>
<td>−0.876</td>
</tr>
<tr>
<td>12th week</td>
<td>slope: 1.006</td>
<td>0.875</td>
<td>0.999</td>
<td>1.084</td>
</tr>
<tr>
<td></td>
<td>intercept: 0.005</td>
<td>1.306</td>
<td>−0.068</td>
<td>−1.057</td>
</tr>
<tr>
<td>25th week</td>
<td>slope: 1.006</td>
<td>0.872</td>
<td>0.999</td>
<td>1.077</td>
</tr>
<tr>
<td></td>
<td>intercept: 0.022</td>
<td>1.325</td>
<td>−0.053</td>
<td>−0.971</td>
</tr>
<tr>
<td>36th week</td>
<td>slope: 1.007</td>
<td>0.872</td>
<td>0.999</td>
<td>1.077</td>
</tr>
<tr>
<td></td>
<td>intercept: −0.011</td>
<td>1.325</td>
<td>−0.053</td>
<td>−0.971</td>
</tr>
</tbody>
</table>

The scatter plots of the original and the bias-corrected satellite SSTs against the aggregated drifting buoy SSTs are shown in Figure 3 and 4. The results of accuracy assessment are shown in Table 2.

Table 2. Accuracy assessment results of original and corrected satellite SSTs.

<table>
<thead>
<tr>
<th>Matched Points</th>
<th>Bias</th>
<th>Std. Dev</th>
<th>RMSE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original MODIS SST</td>
<td>43.195</td>
<td>0.168</td>
<td>0.585</td>
<td>0.609</td>
</tr>
<tr>
<td>Corrected MODIS SST</td>
<td>43.195</td>
<td>0.000</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>Original AMSR-E SST</td>
<td>30.697</td>
<td>0.086</td>
<td>0.452</td>
<td>0.460</td>
</tr>
<tr>
<td>Corrected AMSR-E SST</td>
<td>30.697</td>
<td>−0.001</td>
<td>0.452</td>
<td>0.452</td>
</tr>
</tbody>
</table>

From Figure 3, Figure 4 and Table 2 for original MODIS and AMSR-E SSTs, the bias over cool regions is more obvious than that over warm regions. The mean bias is decreased greatly by correction model. The correction effects over cool regions are more obvious than over warm regions.

Hierarchical Bayesian Model

A hierarchical Bayesian model, including three sub-models, data model, process model and parameter model, is constructed. During the data model stage, the conditional distribution of MODIS SSTs is created. The seamless scale transformation between MODIS and AMSR-E SSTs achieves through one intermediate variable, the mean of normal distribution of data with the same scale as AMSR-E SSTs. During the process model stage, the R-FRF model is used to simulate the true value of a potential spatiotemporal process of SST. The simulation values which describe the trend of SST are used as a priori values of the intermediate variable which link MODIS and AMSR-E SST. In the parameter model stage, the a priori value is assigned to every parameter of the data model and process model. The a posteriori distribution of the average value of potential spatiotemporal process of SST varied in one-year period is estimated through HBM based on R-FRF. Considering the SST variation with latitude is larger than that with longitude, we divide the geographical region into 25 sub-regions per three latitudes except the 26th sub-region. The 26th sub-region covers only one latitude, 29°S to 30°S. Parallel computation based on three longitudes is implemented to improve the computational performance, of which the pixel number of MODIS is 72×3600 = 259200.

Hierarchical Structure of the True SST Spatiotemporal Process

We think of SST as a spatiotemporal random field. We will estimate the a posteriori distribution of the average value of the potential spatiotemporal process of SST updated by the observation via the Bayesian theorem (Equation 2):

\[ \hat{Z}(s,t) = Y(s,t) + \epsilon(s,t). \]

Where \( \text{process} | \text{data} \) denotes the posteriori distribution of the average value of the potential spatiotemporal process of SST. \( \text{data} | \text{process}, \text{parameters} \) denotes the measurement model, where the measurement is thought of as an observation with errors. \( \text{process} | \text{parameters} \) denotes the potential spatiotemporal process of SST that cannot be observed but can be modeled during the process model stage, and \( \text{parameters} \) is the parameter model including all parameters of the data model and process model.

We assume that satellite SST is an observation from an observable spatiotemporal process \( \{Z(s,t) : s \in D = [0, \ldots, t] \text{ with measurement error. The potential spatiotemporal process of SST, } \{Y(s,t)\}, \text{ cannot be observed but can be modeled in the frame of HBM. The relationship between satellite observations and potential true values can be expressed by the Equation 3.} \]

\[ Z(s,t) = Y(s,t) + \epsilon(s,t). \]

Where \( Y(s,t) \) is potential spatiotemporal process of SST described by R-FRF model. \( \epsilon(s,t) \) is a white-noise Gaussian process in space and time, with mean zero and variance \( \text{var}(\epsilon(s,t)) = \sigma^2 \).

Data Model and Scale Transformation

Now, we have only a series of incomplete satellite observations with errors at time \( t \), \( Z(t) = [Z(s_{t,1}, t), \ldots, Z(s_{t,n}, t)]' \). Applying HBM, the conditional distribution of the satellite-derived SSTs must be defined. The MODIS SSTs are treated as a normal distribution with mean, \( Y(s_{mt}, t) \). Recall that the spatial resolution is different between MODIS and AMSR-E SSTs; thus, there is a problem of scale transformation. Comparing these two kinds of data, there is a nested structure between MODIS and AMSR-E pixels. Thirty-six-pixel of MODIS are nested in one pixel of AMSR-E at the same spatial location. So, we develop the nested scale transformation model. We link the two scales together through a normal distribution, \( Y(s_{mt}, t) - N(Y(s_{mt}, t), \sigma^2) \). That is, per 6×6=36 pixels of MODIS corresponding to one pixel of AMSR-E is treated as the normal distribution with mean value, \( \text{AMSR-E SSTs, } Y(s_{mt}, t) \) and variance \( \sigma^2 \). (Guo 2010, Hooten et al. 2008, Wikle 2002). The AMSR-E SSTs, \( Y(s_{mt}, t) \), is also treated as the normal distribution with mean value, a priori means, \( \hat{Y}(s_{mt}, t) \) which is estimated by R-FRF based on corrected
AMS-R SST. In this manner, the spatial resolution match problem of the two products is resolved seamlessly. Therefore, the data model can be expressed as Equation 4 through 6:

\[ Z_{st}(s_{m,t}) \sim N(Y_s(s_{m,t}), \sigma^2_{s_{m,t}}) \]  

\[ Y_s(s_{m,t}) \sim N(Y_s(s_{m,t}), \mathbf{\sigma}_{s_{m,t}}^2) \]  

\[ Y_s(s_{m,t}) \sim N(\hat{Y}_s(s_{m,t}), \mathbf{\tau}_{s_{m,t}}^2) \]  

Where \( \sigma^2_{s_{m,t}}, \mathbf{\sigma}_{s_{m,t}}^2 \) and \( \mathbf{\tau}_{s_{m,t}}^2 \) represent variances; \( N \) denotes the normal distribution; the letters A and a represent AMSR-E; the letters M and m represent MODIS SST.

Equation 4 is the error model of MODIS observations. Equation 5 expresses the scale transformation. Equation 6 expressed the \textit{a priori} value of the middle variable \( Y_s(s_{m,t}) \).

**Process Model – R-RRF Model**

To implement HBM model, the \textit{a priori} mean in Equation 6, \( \hat{Y}_s(s_{m,t}) \), must be estimated at each pixel by potential spatiotemporal process of SST, Equation 7. Because SST has obvious spatial trend with latitude change, after removing the spatial trend the residual still has spatial pattern and the error expression of model decomposition is also considered, Equation 7 includes three sub-process modeling, large-scale variation, small-scale variation, \( \nu(s_{m,t}) \) and fine-scale variation, \( \xi(s_{m,t}) \) (Kang et al. 2010). Considering the completeness and resolution of datasets used in this paper we select the AMSR-E SSTs to estimate the \textit{a priori} mean, \( \hat{Y}_s(s_{m,t}) \).

\[ \hat{Y}_s(s_{m,t}) = \mu(s_{m,t}) + \nu(s_{m,t}) + \xi(s_{m,t}) \]  

**Large-Scale Variation**

In Equation 7, \( \mu(s_{m,t}) \) represents the trend of SST by deterministic spatiotemporal average function. We model it using a linear function to account for a stable estimator of SST trend as shown in Equation 8.

\[ \mu(s_{m,t}) = X_s(s_{m,t}) \beta_t \]  

Where \( X_s(\cdot) = (X_{s,1}(\cdot), ..., X_{s,r}(\cdot))' \) is the vector process of known covariates, and \( \beta_t = (\beta_{t,1}, ..., \beta_{t,r}) \) is the unknown coefficients which we will estimate by detrended AMSR-E SSTs.

**Small-Scale Variation**

In Equation 7, the small-scale variation of SST, \( \nu(s_{m,t}) \), is known as a spatiotemporal stochastic process which is simulated by spatiotemporal random effect model (STRE) as shown in Equation 9 (Cressie et al. 2010). It is described by spatiotemporal effect model and specifically at any fixed time \( t, \nu(\cdot,t) \) has zero mean (Kang et al. 2010).

\[ \nu(s_{m,t}) = S_s(s_{m,t}) \eta_t \]  

Where \( S_s(\cdot) = (S_{s,1}(\cdot), ..., S_{s,r}(\cdot))' \) denotes the vector of deterministic known spatial basis functions with the fixed ranks \( r \), and they are not necessarily orthogonal. Through the fixed ranks \( \{r\} \), the spatial dimension of the computation is decreased greatly. Of course, the number of \( r \) changes with the time \( t \). \( \eta_t = (\eta_{t,1}, ..., \eta_{t,r})' \) denotes a zero-mean Gaussian random vector with \( r \times r \) covariance matrix given by cross-covariances \( K_\nu \). As time \( t \) progresses, the \( \eta_t \) evolves as a vector autoregressive process of order 1:

\[ \eta_{t+1} = H_\nu \eta_t + \zeta_{t+1}, \quad t=1,2,3,... \]  

Where \( H_\nu \) is a \( r \times r \) first-order autoregressive matrix, and \( \zeta_{t+1} \) is \( d \)-dimensional Gaussian innovation vector with zero-mean and innovation variance, \( \text{var}(\zeta_{t+1}) = U_{t+1} \). The related equations are shown as follows:

The cross-covariances:

\[ K_{t+1,t} = \text{cov}(\eta_t, \eta_{t+1}), \quad t = 1,2,... \]  

Based on Equation (9), for \( t_{1} < t_{2} \):

\[ K_{t+1,t} = K_{1} (H_{t+1}, H_{t+1}, ..., H_{t+1})' \]  

And

\[ K_{t+1} = H_{t+1}K_{t+1}H_{t+1}^t + U_{t+1} \]  

Specially, if the time lag is 1 the cross-covariances is described as:

\[ L_{t+1} = K_{t+1,t} = K_{1}H_{t+1}, \quad t = 1,2,... \]  

**Fine-Scale Variation**

In Equation 7, \( \xi(s_{m,t}) \) is error model from decomposition of process model, \( \hat{Y}_s(s_{m,t}) \). It describes the random variation and is modeled as a mean-zero Gaussian white-noise process independent of \( \eta_t \) and has temporally dynamical structure. This random component captures a pixel-scale spatial structure and satisfies \( \xi(s,t)=0 \) and

\[ E(\xi(s, t), \xi(r, u)) = \left\{ \begin{array}{ll} \sigma^2, & \text{if } s = r \text{ and } t = u \\ 0, & \text{otherwise} \end{array} \right. \]  

**Robust Feature**

The FRF method has been successfully applied into global satellite aerosol optical depth (AOD) data (Kang et al. 2010). The selection strategy of basis function in the FRF framework works well for satellite AOD in a regular region. But it cannot be used straightforwadly to the satellite SST data in an irregular region with a large number of missing values. We need a Robust FRF model. The robust feature of FRF model is described by the selection of the multi-resolution spatial basis functions. There are two states during the selection. For the first stage, though the selection of the multi-resolution spatial basis function in this paper is similar to that suggested by Kang, et al. (2010), our fixed ranks \( \{r\} \) depend on time \( t \). Two ratios suggested by Zhu, et al. (2015), \( R_{f}(\cdot)(s) \) and \( R_{f}(\cdot)(f) \), are introduced into the second stage of the basis function selection. \( R_{f}(\cdot)(f) \) is the ratio of the number of ocean pixels where wavelet is nonzero to the number of all pixels where the wavelet is nonzero in the study region. \( R_{f}(\cdot)(s) \) is the ratio of the number of ocean pixels with valid observations as well as nonzero values of wavelet to the number of all pixels where the wavelet is nonzero in the study region. We divide the data of all year into 8 groups, and the maximum of \( r \) in each group is selected.

**R-RRF Predictions**

To obtain the optimal predictor of \( \hat{Y}_s(s_{m,t}) \) given the data \( Z_{st} = (Z_{s,1}(\cdot), ..., Z_{s,l}(\cdot))' \), two steps are needed: forecasting and updating. The forecasting based on one-step-ahead is as follows:

\[ \hat{h}_{t+1,1} = (\eta_{t} | Z_{s,t+1}) = H_{\nu} \hat{h}_{t+1,t} \]  

\[ \hat{P}_{t+1,1} = E(\hat{h}_{t+1,1} - \eta_{t})|\hat{h}_{t+1,1} - \eta_{t})' = H_{\nu} P_{t+1,1} H_{\nu} + U \]
With input of current data, $Z(t)$, prediction is updated as follows:

$$
\hat{y}_{t|t-1} = E[y|Z(t)] = \hat{y}_{t|t-1} - G_t[Z(t) - \mu(t) - S_t \hat{y}_{t|t-1}], \quad t = 1,2,\ldots
$$

$$
P_{t|t-1} = E[(\hat{y}_{t|t-1} - \eta)(\hat{y}_{t|t-1} - \eta)'] = P_{t|t-1} - G_tS_tG_t'
$$

Where $\eta$ is Kalman gain matrix and expresses as follows:

$$
G_t = P_{t|t-1}S_t(S_tP_{t|t-1} + D_t)^{-1}S_t'
$$

Where $D_t = \sigma_{\epsilon}^2I_s + \sigma_f^2V_s$ and $V_s = \text{diag}(v_1(s_1), \ldots, v_s(s_n))$

That the final part of the filtering procedure is on $\tilde{\zeta}_t(s)$ is shown as Equation 21

$$
\tilde{\zeta}_t(s) = c(s)_t(S_tP_{t|t-1} + D_t)^{-1}(Z(t) - \mu(t) - S_t \hat{y}_{t|t-1})
$$

Where $c(s)_t = \text{cov}(Z(t), \tilde{\zeta}_t(s))$.

Then the optimal predictor of $Y(s,t)$ based on the observation $Z_t$ expresses as Equation 22

$$
Y(s,t) = \mu(s,t) + S_t(s,t)\hat{y}_{t|t-1} + \tilde{\zeta}_t(s)
$$


**Parameter Model**

In an HBM framework, for each SST pixel the $a$ priori mean of the potential process of SST, $Y_a(s,m,t)$, is simulated by the R-FRF model; the SST process at each pixel, $Y_a(s,m,t)$, depends on the potential SST process, $Y(s,m,t)$, conditionally. The variance of the MODIS SSTs and the AMSR-E SSTs describes the random variable that the $a$ priori distribution belongs to conjugate prior as Equation 23.

$$
\sigma^2_t \sim IG(q_t, r_t)
$$

Where IG denotes the inverse Gamma distribution, $A$ denotes the different satellite observation. For AMSR-E SSTs the shape parameter $q_r$ and scale parameter $r_r$ are both 0.1 (Xu et al. 2005).

**Results**

This section presents the spatiotemporal simulation of SSTs based on the R-FRF and the sample results after applying HBM to the bias-corrected MODIS and AMSR-E SSTs. The spatial completeness of the fusion SSTs is assessed in terms of the availability of ocean pixels, the ratio of the number of valid ocean pixels to the number of all ocean pixels. Using the drifting buoy SSTs as reference, the accuracy of the fusion SSTs is assessed in terms of mean bias ($\tilde{B}$), error standard deviation ($\text{Std.Dev}$), root mean square errors ($\text{RMSE}$) and correlation coefficient ($R$). The accuracy assessment is implemented over all study regions and four sub-regions, respectively. These four sub-regions are: region with MODIS SSTs and AMSR-E SSTs, region with MODIS SSTs and without AMSR-E SSTs, region without MODIS SSTs and with AMSR-E SSTs and region without MODIS and AMSR-E SSTs.

**Sample Results and Fusion SSTs**

The estimation of SSTs based on R-FRF, regarded as the latent variable, is embedded in the conceptual model of the HBM. Then the biased-corrected MODIS SSTs and time-averaged AMSR-E SSTs are merged by the HBM. HBM is implemented using a sketch of the Markov Chain Monte Carlo (MCMC) algorithm. MCMC is a popular tool for dealing with complex statistical problems, especially in Bayesian analysis, which often requires complex high-dimensional integrals. MCMC introduces the Markov process of random process into Monte Carlo simulation to realize dynamic simulation. Essentially, MCMC method is the Monte Carlo integral using Markov chains. If we know historical data, actually, we know the posterior distribution of known variables, the posterior variables such as mean, variance, quantile, etc. are achieved through integrating the posterior distribution of higher dimension of the variables. In this paper, the MCMC method is based on Gibbs sampling.

For the results presented here, the MCMC runs for 1500 iterations from a single chain until the chain reached convergence with the first 1000 considered burn-in. After the Markov chain converges, the parameters are reliably estimated and the goodness-of-fit of the model is verified. We tried other iterative numbers and found that with the increase of the iterative number, the estimation of the posterior distribution of the average value of the potential true SST process varied very negligible. Therefore, considering the relationship between prediction accuracy and the iterative number we selected 1500 iterations. Table 3 provides part of results of MCMC algorithm for one sample of the 5th week.

From Table 3 the MC error is small, and the chain reached convergence; thus, the estimation of the posterior distribution of the average value of the potential true SST process is believable.

Spatial distribution of the merged, original MODIS and AMSR-E SSTs is shown in Figure 5.

From Figure 5, the original MODIS and AMSR-E SSTs are both incomplete in space. For AMSR-E SSTs, the missing pixels are near the land and for MODIS SSTs the missing pixels are scattered over all study regions, especially over high latitudes and equatorial regions. However, the fusion SSTs are complete in space.

**Assessment of Spatial Completeness**

We assess the spatial completeness of the original MODIS, AMSR-E and fusion SSTs using the availability of ocean pixels. The annual-mean availability of ocean pixels of the fusion SSTs, the original MODIS and AMSR-E SSTs is 100%, 70.08%, and 87.53%, respectively. Figure 6 shows the variation with time of weekly mean availability.

---

**Table 3. Part of results of MCMC algorithm for one sample**

<table>
<thead>
<tr>
<th>Node</th>
<th>Mean</th>
<th>Std.Dev</th>
<th>MC Error*</th>
<th>2.5%</th>
<th>Median</th>
<th>97.5%</th>
<th>Start</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>mu1[38,2166]</td>
<td>29.35</td>
<td>0.9741</td>
<td>0.04592</td>
<td>27.55</td>
<td>29.29</td>
<td>31.3</td>
<td>1001</td>
<td>500</td>
</tr>
<tr>
<td>mu1[38,2167]</td>
<td>29.36</td>
<td>0.9787</td>
<td>0.04589</td>
<td>27.48</td>
<td>29.34</td>
<td>31.38</td>
<td>1001</td>
<td>500</td>
</tr>
<tr>
<td>mu1[38,2168]</td>
<td>29.35</td>
<td>0.9717</td>
<td>0.04769</td>
<td>27.43</td>
<td>29.31</td>
<td>31.27</td>
<td>1001</td>
<td>500</td>
</tr>
<tr>
<td>mu1[38,2169]</td>
<td>29.35</td>
<td>0.9827</td>
<td>0.04767</td>
<td>27.45</td>
<td>29.35</td>
<td>31.41</td>
<td>1001</td>
<td>500</td>
</tr>
<tr>
<td>mu1[38,2170]</td>
<td>29.37</td>
<td>0.9884</td>
<td>0.05298</td>
<td>27.46</td>
<td>29.34</td>
<td>31.43</td>
<td>1001</td>
<td>500</td>
</tr>
<tr>
<td>mu1[38,2171]</td>
<td>29.36</td>
<td>0.961</td>
<td>0.04882</td>
<td>27.57</td>
<td>29.33</td>
<td>31.27</td>
<td>1001</td>
<td>500</td>
</tr>
<tr>
<td>mu1[38,2172]</td>
<td>29.35</td>
<td>0.9658</td>
<td>0.05329</td>
<td>27.53</td>
<td>29.36</td>
<td>31.4</td>
<td>1001</td>
<td>500</td>
</tr>
</tbody>
</table>

*MC error is the standard error of Monte Carlo simulation.
Figure 5. Spatial distribution of the merged, original MODIS and AMSR-E SSTs of 12th, 24th, 36th and 46th week (a)-(d) the merged SSTs; (e)-(h) the original MODIS SSTs; (i)-(l) the original AMSR-E SSTs. White represents no data and gray represents land.

Figure 6. Variation with time of weekly mean availability.
From Figure 6, the availability of ocean pixels of MODIS SSTs is the lowest and the fluctuation is obvious. The availability of ocean pixels of AMSR-E SSTs is high and stable. The fusion SSTs with completeness in spatial domain are derived by the HBM.

Assessment of Accuracy

Taken the drifting buoy SSTs as the reference, the accuracy of the original MODIS, AMSR-E and fusion SSTs is evaluated in terms of mean bias ($B$), error standard deviation (Std.Dev), root mean square error (RMSE) and correlation coefficient ($R$). To assess the overall accuracy and local accuracy of fusion data, we divide our evaluation procedure into two stages (i.e., stage I and stage II) suggested by Zhu, et al. (2018). In stage I, we evaluate the overall accuracy of the fusion data, and in stage II, we evaluate the local accuracy of the fusion data. In each validation stage, we also compare the accuracy of the fusion SSTs with that of the original MODIS and AMSR-E SSTs.

Figure 7 shows the scatter graphs of the fusion, original MODIS and AMSR-E SSTs against the spatially and temporally collocated drifting buoy SSTs. Table 4 shows the evaluation results.

For overall accuracy, the mean bias of the fusion SSTs is smaller than that of the original MODIS and AMSR-E SSTs, but it is a little larger than that of the corrected MODIS and AMSR-E SSTs. The error standard deviation, RMSE and R of the fusion SSTs is similar to the MODIS but a little larger than that of the original and corrected AMSR-E SSTs. Comparing with the results from BME method suggested by Aihua Li (2013), the mean bias of our fusion data is lower than that of Aihua Li (2013). They are 0.045°C and 0.146°C, respectively. The RMSE of our fusion data is also a little smaller than that of Li (2013). They are 0.621 and 0.653 respectively. For the assess results from the STHBM based on spatiotemporal decomposition model suggested by Zhu, et al. (2018), the mean bias, error standard deviation, RMSE and R is -0.1605, 0.6860, 0.7045.

Table 4. Accuracy comparison of the fusion, MODIS and AMSR-E SSTs.

<table>
<thead>
<tr>
<th>Matched Points</th>
<th>B</th>
<th>Std. Dev</th>
<th>RMSE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion SSTs</td>
<td>60370</td>
<td>0.045</td>
<td>0.620</td>
<td>0.621</td>
</tr>
<tr>
<td>Fusion SST with MODIS and AMSR-E SSTs</td>
<td>39761</td>
<td>-0.001</td>
<td>0.533</td>
<td>0.533</td>
</tr>
<tr>
<td>Fusion SST with MODIS SSTs and without AMSR-E SSTs</td>
<td>3239</td>
<td>-0.027</td>
<td>0.657</td>
<td>0.658</td>
</tr>
<tr>
<td>Fusion SST without MODIS SSTs and with AMSR-E SSTs</td>
<td>15775</td>
<td>0.172</td>
<td>0.752</td>
<td>0.771</td>
</tr>
<tr>
<td>Original MODIS SSTs</td>
<td>43195</td>
<td>0.168</td>
<td>0.585</td>
<td>0.609</td>
</tr>
<tr>
<td>Corrected MODIS SSTs</td>
<td>43195</td>
<td>0</td>
<td>0.584</td>
<td>0.584</td>
</tr>
<tr>
<td>Original AMSR-E SSTs</td>
<td>30697</td>
<td>0.086</td>
<td>0.452</td>
<td>0.460</td>
</tr>
<tr>
<td>Corrected AMSR-E SSTs</td>
<td>30697</td>
<td>-0.001</td>
<td>0.452</td>
<td>0.452</td>
</tr>
</tbody>
</table>

Figure 7. Scatter graphs of the fusion, MODIS and AMSR-E SSTs against collocated drifting buoy SSTs (a) fusion SSTs; (b) original MODIS SSTs; (c) corrected MODIS SSTs; (d) original AMSR-E SSTs; (e) corrected AMSR-E SSTs; (f) fusion SSTs with MODIS and AMSR-E SSTs; (g) fusion SSTs with MODIS SSTs and without AMSR-E SSTs; (h) fusion SSTs without MODIS SSTs and with AMSR-E SSTs; (i) fusion SSTs without MODIS and AMSR-E SSTs.
0.9826 respectively, while using method in this paper they are 0.045, 0.620, 0.621, 0.976. Therefore, the overall accuracy of these two types’ fusion data is similar. The results also show that the difference of performance between the HBM based on R-FRF and the STHBM based on spatiotemporal decomposition model is not obvious though there is a little difference in the partitioned scope of the reference data (critical value is 105E and 120E, respectively) and partitioned computing range (critical value is two latitudes and three latitudes, respectively).

In stage II, we evaluate the local accuracy using the same assessment index. The whole study region is divided into four sub-regions (i.e., region with MODIS and AMSR-E SSTs), region with MODIS SSTs and without AMSR-E SSTs, region without MODIS SSTs and with AMSR-E SSTs, and region without MODIS and AMSR-E SSTs. From Table 4 over region without MODIS SSTs and with AMSR-E SSTs and over region without MODIS and AMSR-E SSTs the absolute value of mean bias, error standard deviation and RMSE of the fusion SSTs are a little larger than other regions. The mean bias over the region without MODIS and with AMSR-E SSTs is similar to that of the original MODIS SSTs. The mean bias over the region without MODIS and AMSR-E SSTs is smaller than that of the original MODIS SSTs. Over the region with MODIS and AMSR-E SSTs, the accuracy of fusion data is higher than that over region with MODIS SSTs and without AMSR-E SSTs. Therefore, accuracy of fusion data over region with MODIS SSTs is higher than that of the original MODIS. Comparing with the results in Li (2013), over region without MODIS the absolute mean bias of our fusion data is lower than that of Li (2013). They are 0.172°C and 0.261°C, respectively. The RMSE of our fusion data is similar to that of Li (2013), and they are 0.771 and 0.769 respectively. Over region without AMSR-E the absolute mean bias of our fusion data is smaller than that of Aihua Li (2013), and they are 0.027°C and 0.138°C respectively. The RMSE of our fusion data is smaller than that of fusion data from BME, and they are 0.658 and 0.744 respectively. Over region without MODIS and AMSR-E the absolute mean bias of our fusion data is smaller than that of fusion data from BME. They are 0.122°C and 0.255°C respectively. The RMSE is a little larger than that of fusion data from BME, and they are 0.887 and 0.836 respectively. Over region with MODIS the absolute mean bias of our fusion data is smaller than that of Li (2013). They are 0.001°C and 0.101°C respectively. The RMSE is a little smaller than that of fusion data from BME, and they are 0.533 and 0.590 respectively. Comparing with the results from the STHBM based on spatiotemporal decomposition, over region with MODIS and AMSR-E SSTs, the mean bias, RMSE, Std. Dev., and R is -0.1084, 0.6335, 0.6241, and 0.9827, respectively. Over region with MODIS SSTs and without AMSR-E SSTs it is 0.172, 0.7537, 0.7537, and 0.9920, respectively; over region without MODIS and with AMSR-E SSTs it is -0.3497, 1.0374, 0.9774, and 0.8000, respectively. The accuracy of fusion data from F-FRF-based STHBM is a little higher than that from spatiotemporal decomposition-based STHBM, but the difference is very small.

Although the effect of the restraint of the latent true SST process on the accuracy of the fusion data is obvious, the effect of the valid MODIS SST on the accuracy of the fusion data is more obvious. Complementarity between the MODIS SSTs and the restraint of latent true SST process has great contributions to the data fusion.

Assessment of Spatial Pattern
Using local variance suggested by Li (2013), we evaluate the ability of the fusion SSTs to keep spatial pattern lying in the MODIS SSTs at 4 km. Local variance is a scene texture statistic which has been shown to characterize the relationship between spatial resolution and objects in the scene (Coops et al. 2000, Woodcock et al. 1987). Therefore, Local variance can express the pattern information in the scene, and it can be used to describe the information richness of the fusion SSTs. A high local variance indicates that the data has large variability and fine pattern. The larger the local variance is, the finer spatial pattern lies in SST images (Li, et al. 2013b). Local variance is defined as Equation 24.

$$IVar = \frac{\sum_{i=1}^{N} \left(1 - \frac{1}{n} \sum_{j=1}^{n} (x_{ij} - \bar{x}_j)^2 \right)}{N}$$

Where $IVar$ is local variance, $x_i$ denotes the $i$th pixel in the $j$th window, $\bar{x}_j$ denotes the number of valid ocean pixels in a moving window and $N$ denotes the number of the window.

Different from Li (2013), we reject the moving window in which the number of valid ocean pixels is less than 30 for the fusion SSTs and the original MODIS SSTs, and less than two for the original AMSR-E SSTs to keep computation stability. Figure 8 presents the weekly local variance of the fusion, original MODIS and original AMSR-E SSTs. From Figure 8, the local variance of fusion SSTs is similar to that of original MODIS SSTs. It is much larger than that of original AMSR-E SSTs. From the 19th to 46th week, it is larger than that of original MODIS. So, the fusion SSTs keep the spatial pattern well at 4 km resolution.

Assessment of Fusion Uncertainty
The uncertainty of data, process model and parameter model are expressed well in HBM. HBM also provides the posterior standard deviation to evaluate the uncertainty of the fusion SSTs quantitatively. Figure 9 shows the spatial distribution of the posterior standard deviation of part week (12th, 24th, 36th and 46th week).

Comparing Figure 9 with Figure 5, the distribution trend of posterior standard deviation is that over regions without MODIS SSTs the posterior standard deviation is large. For example, there are lots of missing pixels over high latitudes and equatorial regions at 24th week, correspondingly, the posterior standard deviation is larger than other regions. Over regions with MODIS SSTs, the posterior standard deviation is small. These results indicate that the effect of valid MODIS SSTs on the accuracy of the fusion data is more important than that of the restraint of latent true SST process based on R-FRF.

Discussion
This study explores a new fusion approach, HBM based on R-FRF, to combine infrared and microwave SST to obtain the fusion data with high resolution, complete coverage and fine spatial pattern. Although the Hierarchical Bayesian Model has been applied in Environmental Science and Ecological Science, it is rarely applied in satellite-derived data fusion, especially in SST data fusion.

In our study, all inferences are made according to a Hierarchical Bayesian framework. That is, all parameters are regarded as stochastic variables. R-FRF process model accounts for the trend of the latent true SST process over irregular regions. The estimations of SST based on R-FRF, which are regarded as $a priori$ mean, are embedded into HBM. The preliminary results demonstrate the practicality and advantages of the Hierarchical Bayesian Model in combining satellite data. In our study, we also attempt to regard the error standard deviation as a deterministic value, which is estimated by R-FRF, to replace one of the error random variables in HBM, but we obtain the worse result. Therefore, HBM demonstrates the flexibility in addressing uncertainty, including the uncertainty of the observation data, process model, and parameter...
model. We consider the stochastic term of the time series process as a spatial random effect model instead of simple Gaussian distribution with mean zero. The model discussed here is useful in reconstructing a spatiotemporal series and is not useful in predicting information for a future time. This method can be expanded to other combinations of satellite products. Comparing with BME method, the main difference is that there is a restraint from potential spatiotemporal process of SST in HBM framework, which can reduce the uncertainty of fusion data to some extent. HBM developed in this paper is superior to computational efficiency and accuracy of fusion SST. Comparing with spatiotemporal decomposition-based STHBM, the accuracy of two types fusion data with the restraint from different process model is similar. This also shows that these two process models can both describe the potential spatiotemporal process of SST well.

Conclusions

Fusion results indicate that the accuracy of fusion SST over all study regions is better than that of the original MODIS SST. Although the accuracy of fusion data is lower than that of the original AMSR-E SST, the fusion SST is complete spatially, the spatial resolution is much higher than that of the original AMSR-E SST, and the fusion SST provide a finer spatial pattern than the original AMSR-E SST. Although over regions without MODIS SST, the random and systematic errors of the fusion SST are slightly larger than other regions, the accuracy is better than that of the original MODIS SST. In addition, the fusion SST has the characteristic of the two types of satellite-derived data. The effect of the valid MODIS data on the accuracy of the fusion SST is larger than that of the restraint of the potential spatiotemporal process of SST based on R-FRF. Although, the potential spatiotemporal process of SST as a restraint can be reduce uncertainty of fusion data, the bias correction is necessary before fusion. Finally, we obtain fusion SST with high spatial resolution, spatial completeness, fine spatial pattern and high accuracy.

![Weekly Mean Local Variance](image)

**Figure 8.** Weekly local variances of the fusion, original AMSR-E and original MODIS SSTs.

![Spatial Distribution of Posterior Standard Deviation](image)

**Figure 9.** Spatial distribution of posterior standard deviation of part week.
In this paper, during validation and bias correction we ignore the depth difference of skin SST and sub-skin SST between satellite and drifting buoy SST. We also ignore the spatial representativeness of drifting buoy SST. In order to minimize the impact of the difference in the observed depth we choose the night MODIS SST for analysis in this paper. It is still a challenging to take account for the complicating effect of diurnal stratification in validating satellite SST as we suggested before in (Zhu, et al. 2015). We will implement research on the spatial representativeness of drifting buoy point observations in the near future.

Acknowledgments
This work was supported by the Natural Science Foundation of China under Grant 41401405, the Statistics Bureau of China under Grant 2016LY32, the Natural Science Foundation of Shandong Province under Grants ZR2013DL002 and ZR2017MD017, the China Postdoctoral Science Foundation under Grant 2014M561039, Huainan key laboratory for geographic information technology and application under Grant HAP201405. We would like to express our sincere gratitude to the people who provided excellent comments and valuable suggestions in the preparation of this manuscript.

References


Donlon, C. 2001. The global ocean data assimilation experiment high resolution sea surface temperature pilot project strategy and implementation plan. Marine Environment Unit, Joint Research Centre, Ispra, Italy:58.


SUSTAINING MEMBERS

Aerial Services, Inc.
Cedar Falls, Iowa
www.AerialServicesInc.com
Member Since: 5/2001

Applanix
a Trimble Company
Richmond Hill, Ontario, Canada
www.applanix.com
Member Since: 7/1997

Ayres Associates, Inc.
Madison, Wisconsin
www.AyresAssociates.com
Member Since: 1/1953

Bohannan Huston, Inc.
Albuquerque, New Mexico
www.bhinc.com
Member Since: 11/1992

Cardinal Systems, LLC
Flagler Beach, Florida
www.cardinalsystems.net
Member Since: 1/2001

Certainty 3D LLC
Orlando, Florida
www.certainty3d.com
Member Since: 3/2011

DAT/EM Systems International
Anchorage, Alaska
www.datem.com
Member Since: 1/1974

Dewberry
Fairfax, Virginia
www.dewberry.com
Member Since: 1/1985

DigitalGlobe, Inc.
Longmont, Colorado
www.digitalglobe.com
Member Since: 6/1996

Environmental Research Incorporated
Linden, Virginia
www.eri.us.com
Member Since: 8/2008

Esri
Redlands, California
www.esri.com
Member Since: 1/1987

GeoBC
Victoria, Canada
www.geobc.gov.bc.ca
Member Since: 12/2008

GeoCue Group
Madison, Alabama
info@geocue.com
Member Since: 10/2003

GeoMni, Inc.
Lehi, Utah
GeoMni.net/psm
Member Since: 03/2018

GeoWing Mapping Inc.
Oakland, California
www.geowingmapping.com
Member Since: 1/2017

Global Science & Technology, Inc.
Greenbelt, Maryland
www.gst.com
Member Since: 10/2010

GPI Geospatial Inc.
formerly Aerial Cartographics of America, Inc. (ACA)
Orlando, Florida
www.aca-net.com
Member Since: 10/1994

Keystone Aerial Surveys, Inc.
Philadelphia, Pennsylvania
www.keystoneaerialsurveys.com
Member Since: 1/1985

Kucera International
Willoughby, Ohio
www.kucerainternational.com
Member Since: 1/1992

LizardTech
Portland, Oregon
www.lizardtech.com
Member Since: 9/1997

Martinez Geospatial, Inc. (MTZ)
Eagan, Minnesota
www.mtzgeo.com
Member Since: 1/1979

Merrick & Company
Greenwood Village, Colorado
www.merrick.com/gis
Member Since: 4/1995

Microsoft UltraCam Team
(Vexcel Imaging, GmbH)
Graz, Austria
www.microsoft.com/ultracam
Member Since: 6/2001

Michael Baker Jr., Inc.
Beaver, Pennsylvania
www.mbakerinc.com
Member Since: 1/1950

PCI Geomatics
Richmond Hill, Ontario, Canada
www.pcigeomatics.com
Member Since: 1/1989

Pickett and Associates, Inc.
Bartow, Florida
www.pickettusa.com
Member Since: 4/2007

Office of Surface Mining
Denver, Colorado
www.tips.osmre.gov
Member Since: 8/2017

Quantum Spatial, Inc.
Sheboygan Falls, Wisconsin
www.quantumspatial.com
Member Since: 1/1974

Robinson Aerial Survey, Inc. (RAS)
Hackettstown, New Jersey
www.robinsonaerial.com
Member Since: 1/1954

Routescene, Inc.
Durango, Colorado
www.routescene.com/
Member Since: 12/2007

Sanborn Map Company
Colorado Springs, Colorado
www.sanborn.com
Member Since: 10/1984

Sidewell Company
St. Charles, Illinois
www.sidwellco.com
Member Since: 11/1992

SkyIMD - Imaging Mapping & Data
Richmond, California
www.skyimd.com
Member Since: 1/2017

Surveying And Mapping, LLC (SAM)
Austin, Texas
www.sam.biz
Member Since: 12/2005

Teledyne Optech
Toronto, Canada
www.teledyneoptech.com
Member Since: 1/1999

Terra Remote Sensing (USA) Inc.
Bellevue, Washington
www.terrarremote.com
Member Since: 10/2016

The Airborne Sensing Corporation
Toronto, Canada
www.airsensing.com
Member Since: 7/2003

Towill, Inc.
San Francisco, California
www.towill.com
Member Since: 1/1952

University of Twente/Faculty ITC
Enschede, Netherlands
www.itc.nl
Member Since: 1/1992

U.S. Geological Survey
Reston, Virginia
www.usgs.gov
Member Since: 4/2002

Woolpert LLP
Dayton, Ohio
www.woolpert.com
Member Since: 1/1985
Quantification of Airborne Lidar Accuracy in Coastal Dunes (Fire Island, New York)

William J. Schmelz and Norbert P. Psuty

Abstract
To establish a basis for the utilization of lidar topography as a data source for coastal geomorphological analyses, this study generated statistical metrics of lidar error through the comparison of a June 2014 USGS collection of airborne lidar with a concurrently collected high-accuracy GPS topographic survey collected within the beach and dunes of a portion of Fire Island National Seashore. The examination of bare earth lidar error within the experiment site revealed a complex association between accuracy and environment within the coastal landscape. Accuracy was constrained to better than 50 cm RMSE in areas with vegetated dune topography and, overall, a 38.9 cm RMSE was measured. Higher accuracies were achieved in the flat, non-vegetated beach. A three-dimensional minimization of residuals between the lidar and GPS surveys reduced the total RMSE to 25.2 cm, indicating a correctable systematic offset between the surface generated from the lidar and the true ground surface.

Introduction
Topographical datasets are the foundation of methodologies created to assess and quantify coastal geomorphological change (e.g., Brown and Arbogast, 1999; Morton et al., 1993; Thom, 1991; Woolard and Colby, 2002; Brenner et al., 2017). Moreover, airborne lidar (light detection and ranging) has evolved into a valuable source of topographical data. This is particularly true within the trend towards higher-resolution data that are the product of technological advancement since the turn of the 21st century (Brock and Purkis, 2009). Considering the significant socioeconomic importance of coastal geomorphology and its influence within coupled physical (Leatherman, 1979; Nordstrom and Psuty, 1980), biological (Roman and Nordstrom, 1988), and anthropogenic (Carapuco, 2016) systems (Fitzgerald et al., 2008), the collection and application of lidar data in the coastal realm provides opportunities to better understand the evolution of coastal systems and benefit coastal communities. However, with modern airborne lidar (herein referred to as lidar) systems capable of collecting a high density of data, upwards of 100,000 points per second and covering hundreds of square kilometer in a single day, errors in the data can be reasonably expected (Leigh et al., 2009). Understanding the accuracy of this source of coastal topographical data is critical to its utilization in geomorphological analyses.

Prior research has addressed general error and limitations of lidar systems (Leigh et al., 2009 and references therein), as well as the error and utility of lidar under a variety of terrains and land cover (Hodgson and Bresnahan, 2004; Liu, 2008; Su and Bork, 2006), including the coastal landscape (Gesch, 2009; Krabill et al., 2000; Mitasova et al., 2009; Nayegandhi et al., 2009; Woolard and Colby, 2002). Aspects of the data acquisition, bare earth processing, geodetic transformations, and topographical modeling procedures contribute to error in lidar-derived elevation datasets. The total error in the dataset is a sum of these discrete contributions. The relative magnitude of each of these partial contributions is, in part, a function of variables within the physical environment, ranging from satellite geometries to densities of ground vegetation; and, as a result, the physical environment is the primary factor contributing to the scale of the error. The error reported along with most publicly available lidar is a metric that minimizes a number of contributing factors, such as the presence of variable slope and land cover characteristics. This approach makes it a useful measure of error for the lidar data acquisition in ideal conditions and the lidar system in general, but it is not a metric of error applicable to data collected in physical environments with complex configurations.

The coastal environment is a challenging setting to measure and model topography, in part, due to the spatially variable relief, slope, and vegetation density of the beaches and dunes. These physical characteristics influence the accuracy of the ground surface elevations derived from lidar. However, a comparative analysis and evaluation of the lidar end products to a concurrently collected high accuracy control surface has not been undertaken within the beach/dune environment. This site-specific appraisal is of significance because: (1) coastal geomorphology is expected to constitute a significant impact of climate change and global mean sea-level rise (Wong et al., 2014) with societally important implications (Fitzgerald et al., 2008; Wong et al., 2014); (2) lidar can be a powerful data source for the measurement of coastal geomorphological change (Brock and Purkis, 2009); and (3) the error of lidar collected in the beach/dune setting needs to be evaluated to validate examinations of coastal geomorphology using lidar as a data source. Therefore, a test of lidar error within the coastal beach/dune system was undertaken through the comparison of high accuracy ground-based GPS survey data to contemporaneously collected airborne lidar to contribute to an understanding of the distribution and magnitude of variables influencing the error.

Background: General Limitations of Lidar Accuracy
The accuracy of a topographical lidar DEM is dependent upon a number of factors resulting from a combination of issues encountered in data acquisition and processing (Leigh et al., 2009). A measure of accuracy is often provided as metadata...
alongside lidar datasets, as the originator will typically publish the range of horizontal and vertical error. The vertical accuracy is generally stated as a root mean square error (RMSE) ranging from 10 cm to 20 cm, and the horizontal error typically ranges from 20 cm to 1 m at the 95% confidence interval (e.g., NYSDEC, 2012; USACE, 2010; USGS, 2013; and Wright et al. 2014). This published error accompanying lidar datasets specifies the accuracy of the calibrated lidar system, but does not usually specify the types of terrain wherein this measure of accuracy can be expected to be applicable (Leigh et al., 2009). It also does not account for error potentially introduced by DEM interpolation or the artifacts left behind in bare earth processing routines, a significant challenge to achieving accurate ground elevation portrayal.

In terms of data acquisition, the main sources of error are well described by Leigh et al. (2009), and references therein, as: (1) the accuracy of the laser range measurement; (2) the accuracy of the measured position of the aircraft obtained utilizing differential GPS receivers; (3) the attitude accuracy provided by the inertial measurement unit; and (4) the potential recorded time offsets between these instruments. Further, issues of system calibration can arise and they may include offsets within components, such as the laser scanner and gyroscope, to significantly affect accuracy of the observations (Shrestha et al., 2007). Calibration parameters can change on a flight to flight basis, and it is recommended to design lidar data collection in a manner that reveals such calibration errors (Shrestha et al., 2007). Procedures to calibrate collected data to a well-surveyed reference site, such as an aircraft runway, or analyzing the fit of overlapping swaths are undertaken within both the NASA ATM and USGS EAARL collection programs (Bonisteel et al., 2009; Brock et al., 2002).

Separate from the data collection, one of the largest sources of error within lidar-derived ground surface elevations is the challenging bare earth processing procedure that separates laser pulse returns attributable to the ground surface from those that are associated with other surfaces and objects, such as vegetation, construction, and/or animals. Assigning lidar points to either a ground or a non-ground classification has been referenced as both the most important and the most challenging task in creating a ground surface model from lidar elevation data, and is difficult for areas of variable terrain (Liu, 2008). This is pertinent to the coastal environment because it is often an irregularly vegetated landscape comprised of terrain features with variable slopes and relief. The large number of lidar data points to be processed requires automation in addition to manual interactive processing due to the complexity of coastal topography. The selection of the appropriate filtering algorithm for a particular locality has been described as more of an art than a science (Shrestha, 2007). Naturally, if non-ground points are left in the dataset or if, conversely, ground points are removed, the resulting derivative products of the dataset will provide a less accurate representation of the true ground surface.

Additionally, it has been demonstrated that vertical error can be introduced in the conversion of elevation from ellipsoidal heights to those of a local geoid (Daniels, 2001). For use in a local coordinate system such as Universal Transverse Mercator (UTM), with a geoid based vertical datum such as the standard North American Vertical Datum 1988 (NAVD88), lidar data need to be converted from an ellipsoidal height to an orthometric value utilizing a geoid model (Bonisteel et al., 2009). This is necessary because lidar data are collected in the International Terrestrial Reference Frame (ITRF) or World Geodetic System 1984 (WGS84), systems that utilize ellipsoidal heights. Conversion of the horizontal coordinates is not problematic, whereas any spatial variation of height between the geoid and the ellipsoid will be incorporated into the dataset. Past analyses have noted systematic and uniform vertical error in lidar datasets and have attempted to calibrate the collected observations to a reference surface in order to improve accuracy (Mitrova et al., 2009; Daniels, 2001; Shrestha, 2007).

Further, the irregularly spaced lidar topographical data are typically modeled and stored as a grid-cell matrix DEM that is generated through interpolation. Comparisons of DEM values to original point data have established a negligible bias (mean error), regardless of interpolation algorithm (Bater and Coops, 2009). However, the RMSE in DEMs has been found to be considerably higher on steep slopes, in vegetated areas, or in DEMs with coarser spatial resolutions (Bater and Coops, 2009). This is an important finding relating to the measurement of coastal geomorphological change because DEMs are the typical data storage format used in this sort of spatial analysis. Regarding the assignment of an appropriate spatial resolution for DEMs generated for coastal topographical inquiry, Woolard and Colby (2002) concluded that consideration should be given to the resolution required to portray the relevant terrain features accurately. The extent of detail (resolution) required relates to the local terrain type. However, the resolution of the DEM is also limited by the density of the original point data. Ideally, the resolution should be determined by examining the objectives of the application, the terrain characteristics, and the spatial density of the lidar point cloud.

**Methods**

**Site**

A 70 m stretch of coastline at Fire Island, a 50 km long barrier island on the south shore of Long Island, was chosen for this experiment (Figure 1). The survey site, located near Bellport Beach (UTM 18N 675000 E 4508840 N), contained prototypical beach/dune topography for a wave-dominated shoreline, including a sand beach, a foredune ridge, abandoned foredunes,
and swales (Figure 2). The most prominent feature within the site was a large active foredune ridge that ran parallel to the shoreline landward of a sand beach berm. Landward of this active dune, was a small abandoned foredune ridge, with a small flat swale between the two dune features. The beach, active foredune, and a portion of the swale inland of the dune ridge were examined in this experiment. The abandoned foredune was just outside of the study area. This combination of coastal features provided an ideal natural laboratory because it contained a number of transitions between geomorphological formations in a small area (beach to dune to swale), as well as vegetated and non-vegetated sub-areas situated on a variety of slope magnitudes oriented with differing aspects (Figure 3).

GPS Field Survey Dataset

On 27 June 2014, a grid network of elevation control points was collected utilizing geodetic GPS survey equipment, providing a high accuracy representation of the ground surface for comparison with concurrently collected lidar data. The topographic survey covered 4,300 m² and contained 734 points. The collection of the GPS survey data was conducted utilizing Two Trimble R10 and one Leica GS12 geodetic GPS receivers configured for Real Time Kinematic (RTK) differential correction. These receivers were mounted on 2 m rover poles, with controllers that facilitated navigation and review of the collected data in the field. To maintain a relatively even distribution of data points, survey points were collected every 2.5 m along 26 transects with a north-south orientation that were spaced 2.5 m apart, forming a 2.5 m x 2.5 m grid of points (Figure 4). The coordinates for the survey points were created prior to the survey and loaded onto the survey controllers. The GPS equipment was used to navigate to the survey point coordinates. Once the equipment was positioned at a survey point location, the topographical point was collected. At some of the pre-selected locations, it was not possible to collect data. For example, a small portion of the survey consisted of thick brush measuring higher than 2 m (See Supplemental Materials Figure 1) that made the area impossible to survey due to a lack of satellite signal (Figure 4).
During the survey, the two Trimble R10 geodetic GPS receivers were configured as a base/rover pair for the RTK differential correction. The “base” receiver provided the RTK correction to the “rover” receiver collecting the survey points. The base station was set up over National Geodetic Survey (NGS) benchmark “U 374” (PID KU0206) that was approximately 75 m from the survey site. Trimble, Inc. (2017) identifies the vertical accuracy (RMS) of a single baseline RTK survey for the R10 receiver as 0.015 m + 1.0 ppm (parts-per-million), referenced to the distance from the base station. The distance to the base station was less than 110 m for all survey points. The Leica GS12 geodetic GPS receiver had cellular internet functionality and received RTK differential corrections from the New York State Department of Transportation (NYSDOT) Real-Time Network (RTN) of Continuously Operating Reference Stations (CORS) during the survey. Leica Geosystems (2011) identifies the network RTK survey vertical accuracy (RMS) of the Leica GS12 receiver as 0.015 m + 0.5 ppm, referenced to the distance from the nearest network base station. The nearest CORS network location was the NYCI station 23 km away from the survey site. Seventy-one of the GPS survey points were reoccupied several hours after they were first obtained to empirically measure the accuracy of the GPS topographic survey.

For the Trimble base station set-up, the NGS “U 374” benchmark position had not been recently surveyed, so temporary coordinates for the base station were obtained on-site using the Leica GS12 initialized with NYSDOT RTN differential correction. This temporary position was used to configure the Trimble base/rover pair for RTK differential correction in the field and provide reasonably accurate navigation to the target survey point locations in the survey shapefile. The base station location was refined after the survey using the NGS’s Online Position User Service (OPUS). This OPUS solution was derived using data collected by the base receiver during the 7.5-hour survey at the site. The refined base station coordinates were incorporated as the control point for differential correction through survey post-processing.

The survey download and post-processing for the data collected with the Trimble survey equipment was undertaken with Trimble Business Center software, and exported to a text file that contained the easting and northing coordinates of each collected survey point in the UTM 18N coordinate system and their corresponding orthometric heights relative to GEOID12A. The survey download for the data collected using Leica equipment was completed using Leica Geo Office software, and the data were similarly exported to a text file in the UTM 18N coordinate system with orthometric heights referenced to GEOID12A.

### Lidar Dataset and Processing

The lidar dataset provided by the U.S. Geological Survey (USGS) for use in this research was flown over this stretch of the Fire Island coastline on 27 June and 30 June 2014, collected with the Experimental Advanced Airborne Research Lidar (EAARL) sensor. This sensor has the capacity to measure accuracies to a few centimeters for open canopies and has been shown to achieve 16 to 20 cm accuracy in terms of root mean square error (RMSE) for sub-canopy topography, with accuracy dependent upon environmental factors such as GPS satellite configurations and physical parameters such as scan angle (Nayegandhi et al., 2009; Bonisteel et al., 2009). The EAARL configuration utilizes a full-waveform sensor that provides a sample rate of up to 10,000 pulses per second with a footprint of 20 cm in diameter (Nayegandhi et al., 2009). The dataset was provided by the USGS as a LAS (log ASCII standard) file that contained a bare earth classified point cloud. The points, provided in the UTM 18N coordinate system and with elevations relative to GEOID12A, were extracted to a text file using LAStools software (Isenburg, 2014). The dataset contained an average point spacing of approximately 1.24 m or density of 0.65 points/m².

The flight altitude is presumed to be 300 m, the suggested flight altitude in the EAARL data processing manual (Bonisteel et al., 2009). According to the timestamps associated with each of the points, the site was flown over four times during the two lidar survey days, once on 27 June and three times on 30 June (Figure 5). Each lidar swath covered a varying amount of the ground survey area. Swath 1 was flown on June 27 from a position seaward of the foredune and covered the beach portion of the ground survey, contributing 664 data points within 5 m of the GPS data points. Swath 2, the first swath flown on 30 June, covered the majority of the survey area, from a flight line seaward of the foredune. This swath contributed 1,218 bare earth data points within the ground survey area. This segment of the full dataset covered portions of both the beach and the dune. Considering the seaward vantage point and resulting scan angles, there was a noticeable dearth of bare earth points on the landward side of the foredune, a condition possibly attributable to a shadow effect. Swath 3, also flown on 30 June, covered the entire ground survey area, contributing 2,263 points to the dataset within the survey area. The point density for this swath was higher than the others. Swath 4, the final swath collected on 30 June 30, contributed a very small amount of data within the study area. The flight line was likely landward of the survey area and the northwestern corner of the back dune area contained 100 points from this swath. It is unclear whether or not points at the lateral margins of each swath were previously removed from the lidar dataset.

![Figure 5. Distribution of lidar survey data collected on each of the four passes flown over the GPS survey area on 27 June and 30 June 2014.](image-url)
Overview of Data Comparisons
A first step in the process of comparison was to establish a general metric of accuracy for the GPS survey. This involved calculating differences in elevation between 71 GPS survey points that were collected as a reoccupation of a counterpart within the set of 734 GPS survey points collected for comparison to the lidar. The vertical RMSE for elevations provided by the GPS survey was derived from this comparison. This test of GPS survey accuracy provided a foundation for the larger experiment testing lidar accuracy because it determined the GPS survey data to be about an order of magnitude more accurate than typical lidar derived elevations. Therefore, the difference in elevation obtained through a comparison of these datasets is a measure of lidar accuracy.

The lidar data were then compared to the GPS survey data with two primary objectives: (1) calculating the accuracy of this lidar collection within the beach/dune environment; and (2) characterizing the nature and magnitude of partial contributions to the error from aspects of the lidar collection and physical characteristics of the beach/dune environment. The first objective was accomplished through the subtraction of the reference elevations provided by the GPS survey from the lidar elevations and the calculation of the mean difference, absolute mean difference, and the standard deviation of the differences. The second objective was accomplished by: (1) budgeting predictable partial contributions to the total error based on assumptions of normal distributions of published error associated with the lidar system coupled with the physical characteristics of the landscape and artifacts of the two data collections, such as local slope and small differences in the horizontal position of compared elevation points; and (2) assuming the remaining “unaccounted” error is likely introduced by lidar system calibration, geoid transformation, bare earth processing, land cover characteristics, and/or interpolation (for a DEM).

Calculation of Lidar Accuracy
The comparison of the lidar and the reference dataset through subtraction was carried out by directly subtracting the GPS survey point elevations from bare earth lidar point elevations. The GPS survey points were programmatically compared to the lidar points by cycling through each GPS point and subtracting its elevation value from lidar elevation returns that fell within 0.5 m of the GPS point location. Therefore, the error associated with an individual lidar point elevation follows Equation 1, where the elevation of the lidar point is given by \( Z_{\text{lidar}} \), and the elevation of the GPS point is provided by \( Z_{\text{GPS}} \).

\[
\text{Lidar Error} = Z_{\text{lidar}} - Z_{\text{GPS}}
\] (1)

The differences in elevation between each of the lidar returns and their corresponding GPS survey points were summarized by the calculation of mean error (ME), mean absolute error (MAE), and RMSE between the sets of lidar and GPS elevation points.

The error associated with a lidar DEM was also examined through comparison to the GPS survey points. First, a 2.5 m spatial resolution DEM was created from the bare earth lidar dataset provided by the USGS. The DEM was generated utilizing the Delaunay triangulation with linear interpolation method in Golden Software’s Surfer software, with the interpolated value corresponding to the GPS survey point locations. The GPS point elevations were then subtracted from the resultant DEM elevations. Similar to the point-to-point comparison, the difference in elevation between each DEM return and the corresponding GPS survey point was recorded, and ME, MAE, and RMSE between the sets of lidar DEM elevations and GPS elevation points were calculated.

Characterization of Lidar Error
Because the horizontal and vertical components of lidar error are well defined for a well calibrated lidar system utilized in ideal conditions, some partial contributions to composite error for data collected over variable terrain should be predictable, such as the error resulting from normally distributed horizontal error and slope (Hodgson and Bresnahan, 2004). Removing the predictable components from the resultant error, the remainder likely represents error introduced by components such as system calibration, land cover characteristics, bare earth processing, geoid transformation, and DEM interpolation. From this “remaining” error, quantitative and qualitative characterizations of the lidar error can be made. Lidar system calibration and geoid transformation errors are likely to be systematically manifested, quantifiable, and removable. Error persistent beyond these manipulations is likely to correspond to: (1) bare earth processing; (2) the physical obstruction of laser pulses to the ground surface during acquisition; and (3) the introduction of error through DEM interpolation.

The statistical model and associated equations presented in Hodgson and Bresnahan (2004) were used to segregate and budget sources of error within the lidar dataset. They identified the partial contributors to the composite error of lidar as: (1) normally distributed horizontal error; (2) error associated with the inherent accuracy of the lidar system; (3) error introduced by interpolation; and (4) the error associated with the ground survey. The formulas were applied utilizing the assumption that the aircraft was flying at an altitude of 300 m during the survey resulting in a horizontal RMSE of 30 cm, while also conservatively assuming that this lidar system contains a vertical RMSE of 15 cm (Brock et al., 2002; Leigh et al., 2009). The results are metrics of variance representing the contribution to the total observed error from the lidar system (\( \text{RMSE}_{\text{lidar}} \)), ground survey error (\( \text{RMSE}_{\text{gps}} \)), and normally distributed horizontal error (\( \text{RMSE}_{\text{H}} \)). As a modification, the contribution of error introduced by the distance of control points from the lidar comparison points was also incorporated. This was accomplished utilizing the RMS of all expected errors attributable to slope and distance from control points (\( \text{RMSE}_{\text{SD}} \)). Utilizing the adapted Hodgson and Bresnahan (2004) formula created to determine the RMSE due to slope and normally distributed horizontal error (\( \text{RMSE}_{\text{SD}} \)), all variables were accounted for except the lidar system error (\( \text{RMSE}_{\text{lidar}} \)). Solving for lidar system error (\( \text{RMSE}_{\text{lidar}} \)) followed Equation 2:

\[
\text{RMSE}_{\text{lidar}} = \sqrt{\text{RMSE}^2 - \text{RMSE}^2_{\text{gps}} - \text{RMSE}^2_{\text{H}} - \text{RMSE}^2_{\text{SD}}}
\] (2)

The method for calculating \( \text{RMSE}_{\text{lidar}} \) involved calculating the slope of the topography at each ground control point/lidar elevation point comparison. This was undertaken by calculating the plane of all triangles that: (1) incorporate the comparison survey point as a vertex; (2) incorporate two other GPS survey points within a 4 m radius as vertices; and (3) contain the lidar elevation point (\( Z_l \)) within the triangle boundary. From these triangles, the triangle containing the non-GPS survey point vertices with the smallest distance between them was chosen to interpolate an expected comparison elevation based on the location of the lidar survey point relative to the control GPS point (Figure 6). Essentially, this method produced the planar equation for the 45° arc segment of the 50 cm search radius circle surrounding the ground survey point that had a lidar point within it, and a linear interpolation of the ground surface between the ground survey point and the nearest GPS measured elevations in the direction of the lidar point. Utilizing this planar equation, the slope adjusted GPS survey elevation (\( Z_{\text{GPS, adjusted}} \)) at the lidar point’s horizontal position was calculated. The slope adjusted error follows Equation 3:
The vertical shift in the data was evaluated by displacing the lidar survey points vertically at centimeter intervals outward from the original position (in centimeter intervals) while rotating between 0 and 358° in 2° intervals. The points were also displaced vertically from 0 to 30 cm in 1 cm intervals. Metrics of error were re-calculated utilizing points that landed within 50 cm of a GPS data point for each iteration. So, 30 iterations were required for the analysis of the vertical shift following the model of Equation 4; where \( n \) is the number of GPS survey points, \( m_i \) is the number of lidar points within 50 cm of GPS point \( i \) with horizontal coordinates of \((x_i, y_i)\), \( Z_{GPS_i} \) is the elevation of GPS point \( i \), \( Z_i \) is the elevation of a lidar point within the subset \( m_i \) that contains horizontal coordinates \((x_i, y_i)\) that are within 50 cm of \( Z_{GPS_i} \), and \( k \) is the vertical adjustment of the lidar elevation (0 to 30 cm). The examination of a three-dimensional shift required 108,000 iterations following the model of Equation 5, where the coordinates \((x_j, y_j)\) of the lidar point \( Z_i \) have been shifted with a radius of \( r \) (0 – 2 m) and an angle \( \theta \) (0 to 358°) to coordinates \((x_i', y_i')\) that are within 50 cm of \( Z_{GPS_i} \). The shift in horizontal coordinates was executed according to Equation 6.

\[
RMSE(k) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} m_i \sum_{j=1}^{m_i} (Z_{GPS_i} + k - Z_i)^2}
\]

\[
RMSE(r, \theta, k) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Z_{GPS_i} + k - Z_i)^2}
\]

\[
\begin{bmatrix}
x_i' \\
y_i'
\end{bmatrix} + r \cdot \cos(\theta) = \begin{bmatrix}
x_i' \\
y_i'
\end{bmatrix}
\]

**Evaluation of Spatial Variations in Error**

The derived metrics were considered spatially and with respect to various components of the physical environment. This was approached quantitatively by separating the error for the beach and dune geomorphological classifications, vegetated and non-vegetated land cover characteristics, varying magnitudes of topographical slope, and individual flight swathes. The spatial boundaries of the beach and dune were defined by the geomorphological map of Psuty et al. (2015). Presence or absence of vegetation was identified from an orthophoto (NOAA, 2012) (Figure 4). The topographical slope across the survey area was calculated through a planar regression of GPS survey points. Lidar points associated with individual flyovers were identified using the GPS timestamps in the LAS data file.

To quantify the spatial variations in slope statistically, a Pearson’s r coefficient and associated p-values were calculated to test the correlation between slope magnitude and absolute error (AE).

**Results**

**GPS Survey Accuracy**

The initial calculation, a comparison of the subset of points from the GPS survey collected to assess its accuracy to the subset of GPS points collected for comparison to the lidar data, revealed a vertical RMSE of 2.61 cm. This confirmed the accuracy of the GPS equipment and demonstrated that the topographical survey was a valid source of ground elevations for comparison to the lidar.

**Lidar Point to GPS Survey Point Comparison**

Metrics of error were calculated through the comparison of the lidar point elevations to GPS survey point elevations.
The results were subdivided by geomorphological classification, beach or dune. The RMSE and MAE were lower for points that fell on the gently sloping and vegetation free beach than for points located within the dune area. The adjustment for local slope did not improve the metrics of mean, absolute, or RMS error for this comparison. Lidar system error ($RMSE_{LS}$) was solved for utilizing Equation 2 and inputs derived from the survey data ($RMSE_{GPS}$, $RMSE_{ls}$, $RMSE_{d}$) and the results of the initial comparison ($RMSE$). The resulting estimation of $RMSE_{LS}$ was 36.4 cm. Presuming that the calibrated lidar system used for this flight was approximately 15 cm RMSE, an additional square root of variance contribution ($RMSE_{OTHER}$) of approximately 35 cm would be required to reduce the $RMSE_{LS}$ to this value. The ME for the vegetated dune area indicated an overestimation of the topographical surface (lidar elevations higher than the ground surface) from the lidar data contrasted with a significant underestimation of the non-vegetated dune surface (lidar elevations lower than the ground surface). Aside from ME, the metrics for the vegetated dune and non-vegetated dune were similar.

Shifting the points 12 cm higher to account for underestimation of the ground surface dataset resulted in a RMSE$_{LS}$ of 33.7 cm. This 12 cm shift was the best-fit of all vertical displacements. It required an additional $RMSE_{OTHER}$ of ~33 cm to achieve the established expectation of 15 cm RMSE for the calibrated sensor. Results of this best-fit and vertically-shifted comparison were provided for beach, dune, and combined topographical features, in addition to the vegetated and non-vegetated surfaces of the dune (Table 2). This shift resulted in no ME, and reduced the total RMSE for the entire dataset. The RMSE for the low-slope and non-vegetated beach that was originally underestimated was also improved, whereas the RMSE values for the dune were increased. Within the dune topography, the RMSE of the originally underestimated non-vegetated portion was reduced, whereas RMSE values increased in the vegetated portion. Similar to the original unshifted comparison, the adjustment for local slope did not significantly improve the metrics of MAE nor RMSE. The differences between the “Slope Adjusted RMSE” and “RMSE” amounted to 4 mm and 7 mm for the unshifted and the vertically-shifted comparisons, respectively.

The best-fit three-dimensional shift of the lidar data to the GPS survey (Table 3) came from a 130 cm horizontal displacement at a 260° angle (counter-clockwise from the east) with 9 cm of elevation added (Figure 7). The $RMSE_{LS}$ was 25.5 cm and required only 20.6 cm from $RMSE_{OTHER}$ to account for the originally-determined calculated error in addition to the 15 cm RMSE published for the lidar system. The RMSE of the low-slope, non-vegetated beach was only 18.5 cm, whereas the RMSE of the dune was 32.9 cm. The RMSE of the vegetated portion of the dune was significantly higher than the non-vegetated area in the three-dimensionally shifted comparison. Whereas the dune as a whole was overestimated by 9.2 cm, the non-vegetated portion of the dune was overestimated by only 1.5 cm. Additionally, the adjustment for local slope improved the metrics of MAE and RMSE by 0.7 and 1.0 cm, respectively.

### Influence of Slope and Individual Flight Swaths

The error of both the unshifted and three-dimensionally shifted lidar datasets is positively correlated with the slopes calculated from the GPS survey data (Figure 8A). The Pearson’s r coefficient is slightly higher in the three-dimensionally shifted data (0.36) than it is in the unshifted data (0.30). The correlation is significant and p-values are less than 0.01 in both cases. The lower Pearson’s r coefficient for the unshifted data is exemplified by the higher unshifted RMSE value on slopes of 0 to 1.5 degrees than the RMSE value for slopes of 1.5 to 4 degrees (Figure 8D). In contrast, the MAE and RMSE of the three-dimensionally shifted lidar data increases within each slope bin (Figure 8C and 8D). Slopes greater than 4° were calculated

---

**Table 1. Results of lidar point to GPS survey point comparison, subdivided by beach and dune areas.**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>ME (cm)</th>
<th>MAE (cm)</th>
<th>RMSE (cm)</th>
<th>Slope Adjusted ME (cm)</th>
<th>Slope Adjusted MAE (cm)</th>
<th>Slope Adjusted RMSE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>-11.6</td>
<td>30.5</td>
<td>38.9</td>
<td>-11.8</td>
<td>30.5</td>
<td>38.5</td>
</tr>
<tr>
<td>Beach</td>
<td>-23.1</td>
<td>24.0</td>
<td>27.1</td>
<td>-24.0</td>
<td>24.8</td>
<td>28.1</td>
</tr>
<tr>
<td>Dune</td>
<td>0.0</td>
<td>36.7</td>
<td>47.8</td>
<td>0.0</td>
<td>36.7</td>
<td>47.1</td>
</tr>
<tr>
<td>Vegetated dune</td>
<td>24.2</td>
<td>37.3</td>
<td>49.2</td>
<td>24.3</td>
<td>35.0</td>
<td>46.8</td>
</tr>
<tr>
<td>Non-vegetated dune</td>
<td>-31.1</td>
<td>36.1</td>
<td>46.0</td>
<td>-34.1</td>
<td>39.6</td>
<td>50.4</td>
</tr>
</tbody>
</table>

**Table 2. Results of lidar point to GPS survey point comparison, subdivided by beach and dune areas, with the lidar elevation points elevated 12 cm vertically.**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>ME (cm)</th>
<th>MAE (cm)</th>
<th>RMSE (cm)</th>
<th>Slope Adjusted ME (cm)</th>
<th>Slope Adjusted MAE (cm)</th>
<th>Slope Adjusted RMSE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.0</td>
<td>26.5</td>
<td>37.2</td>
<td>0.0</td>
<td>26.4</td>
<td>36.5</td>
</tr>
<tr>
<td>Beach</td>
<td>-11.1</td>
<td>14.4</td>
<td>18.1</td>
<td>-11.9</td>
<td>15.2</td>
<td>18.9</td>
</tr>
<tr>
<td>Dune</td>
<td>11.6</td>
<td>38.3</td>
<td>49.2</td>
<td>11.6</td>
<td>38.2</td>
<td>48.5</td>
</tr>
<tr>
<td>Vegetated dune</td>
<td>36.2</td>
<td>44.0</td>
<td>56.1</td>
<td>36.3</td>
<td>42.2</td>
<td>54.0</td>
</tr>
<tr>
<td>Non-vegetated dune</td>
<td>-19.1</td>
<td>31.1</td>
<td>38.9</td>
<td>-22.7</td>
<td>35.1</td>
<td>43.3</td>
</tr>
</tbody>
</table>

**Table 3. Results of lidar point to GPS survey point comparison, subdivided by beach and dune areas, with the lidar elevation points shifted 1.3 m at an angle of 260° and 9 cm vertically.**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>ME (cm)</th>
<th>MAE (cm)</th>
<th>RMSE (cm)</th>
<th>Slope Adjusted ME (cm)</th>
<th>Slope Adjusted MAE (cm)</th>
<th>Slope Adjusted RMSE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.0</td>
<td>19.8</td>
<td>26.2</td>
<td>0.9</td>
<td>19.1</td>
<td>25.2</td>
</tr>
<tr>
<td>Beach</td>
<td>-8.5</td>
<td>15.3</td>
<td>18.5</td>
<td>-8.3</td>
<td>15.8</td>
<td>19.1</td>
</tr>
<tr>
<td>Dune</td>
<td>9.2</td>
<td>25.3</td>
<td>32.9</td>
<td>8.6</td>
<td>23.6</td>
<td>31.2</td>
</tr>
<tr>
<td>Vegetated dune</td>
<td>21.1</td>
<td>30.9</td>
<td>39.2</td>
<td>22.1</td>
<td>30.1</td>
<td>38.2</td>
</tr>
<tr>
<td>Non-vegetated dune</td>
<td>1.5</td>
<td>21.6</td>
<td>27.8</td>
<td>2.1</td>
<td>19.7</td>
<td>25.4</td>
</tr>
</tbody>
</table>
to have 34.2 cm RMSE, whereas slopes of 0 to 1.5 degrees had an RMSE value of 18.5 cm after the three-dimensional shift.

These comparisons were calculated with a lidar dataset comprised of points from four separate swaths of data, flown over two days. The error for each swath was unique and is provided in Table 4, unadjusted for systematic error.

**Lidar DEM to GPS Survey Point Comparisons**

The error associated with lidar data interpolated into a DEM (Figure 9A) was also calculated considering its widespread use in spatial analyses of coastal change. The results of the comparison between the DEM derived from the un-displaced lidar data had a negative ME for the entire study area (Table 5). Spatially, the ME is negative for the beach and positive for the dune area (Figure 9B), whereas the MAE grades from a lower value in the beach to a higher value in the dune. Unlike the point-to-point comparison, the RMSE values of the vegetated portion of the dune were significantly higher than that of the non-vegetated areas in the unshifted DEM-to-point comparison.

The results of the comparison of the DEM derived from the three-dimensionally shifted data indicated lower RMSE values than the unshifted comparison (Table 6). Additionally, lower values of MAE were observed in the beach relative to the dune (Figure 9C). Similar to the three-dimensionally shifted lidar point to GPS point comparison, there was greater error in the vegetated portion of the dune than in the non-vegetated area.

### Table 4. Results of lidar point to GPS survey point comparison, subdivided by beach and dune areas for each of the four swaths flown over the surveyed area.

<table>
<thead>
<tr>
<th>Swath</th>
<th>Comparison</th>
<th>ME  (cm)</th>
<th>MAE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Total</td>
<td>-12.0</td>
<td>16.6</td>
</tr>
<tr>
<td></td>
<td>Beach</td>
<td>-12.0</td>
<td>16.6</td>
</tr>
<tr>
<td></td>
<td>Dune</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Total</td>
<td>-8.0</td>
<td>23.6</td>
</tr>
<tr>
<td></td>
<td>Beach</td>
<td>-19.4</td>
<td>19.8</td>
</tr>
<tr>
<td></td>
<td>Dune</td>
<td>3.7</td>
<td>27.6</td>
</tr>
<tr>
<td>3</td>
<td>Total</td>
<td>-13.1</td>
<td>36.1</td>
</tr>
<tr>
<td></td>
<td>Beach</td>
<td>-30.9</td>
<td>31.1</td>
</tr>
<tr>
<td></td>
<td>Dune</td>
<td>-0.6</td>
<td>39.6</td>
</tr>
<tr>
<td>4</td>
<td>Total</td>
<td>-22.9</td>
<td>34.2</td>
</tr>
<tr>
<td></td>
<td>Beach</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dune</td>
<td>-22.9</td>
<td>34.2</td>
</tr>
</tbody>
</table>

### Table 5. Results of lidar DEM to GPS survey point comparison, subdivided by beach and dune areas.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>ME  (cm)</th>
<th>MAE (cm)</th>
<th>RMSE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>-5.7</td>
<td>31.9</td>
<td>41.2</td>
</tr>
<tr>
<td>Beach</td>
<td>-24.2</td>
<td>24.6</td>
<td>27.1</td>
</tr>
<tr>
<td>Dune</td>
<td>4.3</td>
<td>35.8</td>
<td>47.0</td>
</tr>
<tr>
<td>Vegetated</td>
<td>33.5</td>
<td>41.2</td>
<td>53.8</td>
</tr>
<tr>
<td>Non-vegetated</td>
<td>-25.1</td>
<td>29.0</td>
<td>36.9</td>
</tr>
</tbody>
</table>

### Table 6. Results of lidar DEM created from shifted data to GPS survey point comparison, subdivided by beach and dune areas.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>ME  (cm)</th>
<th>MAE (cm)</th>
<th>RMSE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>7.3</td>
<td>23.0</td>
<td>32.7</td>
</tr>
<tr>
<td>Beach</td>
<td>-7.9</td>
<td>12.0</td>
<td>14.5</td>
</tr>
<tr>
<td>Dune</td>
<td>15.6</td>
<td>28.9</td>
<td>39.2</td>
</tr>
<tr>
<td>Vegetated</td>
<td>27.1</td>
<td>35.5</td>
<td>46.2</td>
</tr>
<tr>
<td>Non-vegetated</td>
<td>7.0</td>
<td>21.5</td>
<td>29.2</td>
</tr>
</tbody>
</table>
Discussion

Systematic Shift of the Lidar Data

The comparison of the lidar data to the GPS survey revealed relatively large variations between the two datasets. Spatial patterns in the spatial distribution of error showed consistent overestimation of landward facing slopes, combined with an underestimation of seaward facing slopes (Figure 9B). The configuration of over- and under-estimation on slopes of opposing direction indicated a consistently shifted misrepresentation of the topography. Moreover, the subset of the lidar dataset covering the beach surface was underestimated by 23.1 cm (27.1 cm RMSE), and there was a ME of 0.0 cm for the dune surface (47.6 cm RMSE), resulting in an underestimation of 11.6 cm for the entire dataset (38.9 cm RMSE). It would be more consistent with expectations if there were an overestimation of the elevation of the dune area, because of vegetation intercepting returns prior to contact with the ground surface, and if there were a negligible ME for the beach surface. Considering these observed spatial patterns of error, the fit of both elevation and slope of a shifted lidar surface to that created using the GPS ground survey, and the high error associated with the lidar collection, it is likely that the lidar dataset contained a systematic spatial displacement in both the horizontal (x,y) and vertical (z) dimensions.

Metrics of error were improved through the applied vertical adjustment, but the technique still provided a relatively high estimation of RMSE$_{z}$ of 33.7 cm. The results when the lidar data were shifted both horizontally and vertically fit the ground surface data better (Figure 7; 25.5 cm RMSE$_{z}$) and were therefore a more accurate representation of the three-dimensional ground surface. Further, metrics accounting for slope and distance between the GPS and lidar comparison point pairs improved in the best-fit three-dimensionally shifted comparison, indicating a better fit of slopes in addition to a better fit of elevations. This signifies that shifting each lidar point in a three-dimensional direction defined by the slopes of the GPS survey and the distance between the lidar and GPS reference point improves the statistical metrics of error. This occurrence necessitates that the lidar slope and GPS survey slope are correlated, otherwise the correction would produce random results. Likely for this reason, improvement correcting for slope was not observed in the original and vertically shifted point-to-point comparisons, implying that the slopes of the lidar survey and the GPS survey were not correlated using data that were not shifted horizontally. This evidence supports the premise that the entire surface was shifted systematically while largely maintaining the true, relative dimensions of the surface features.

It has been well established that it is reasonable and appropriate to correct for any apparent systematic sources of error within the dataset, and correcting for vertical offsets are a common topic of prior research (Mitrovica et al., 2009; Daniels, 2001). Whereas there are well-established causes for the potential vertical displacement of lidar data converted into a local coordinate system (Daniels, 2001), it is not well documented whether or not issues may arise that can result in the introduction of systematic horizontal error. We suggest that the empirical results herein would support applying a three-dimensional shift to correct lidar data, if possible, despite the lack of a proposed mechanism.

Influence of the Physical Environment

In all comparisons, the dunes within the survey area were less accurately measured than the beach. Moreover, the vegetated portions of the dune areas were less accurately measured than those that were not vegetated. Slope was statistically correlated to error and the largest errors occurred in the areas with the steepest slopes (Figure 8). Additionally, the aerial imagery and observations from the field connected the largest overestimations with the presence of dense vegetation, particularly in the area immediately landward of the dune crest (Figure 9A and 9C). The results showed that the presence of high slopes or dense vegetation can reduce the accuracy of lidar collected within beach/dune environments.

Moreover, the lidar dataset consisted of four swaths of collected data, one collected on a different day than the others. Each swath possessed unique metrics of error and contributed to the error in the final dataset in varied proportions. For example, swath 4 only contributed 8 points that fell within 50 cm of a GPS survey point, whereas swath 3 contributed 206 points. The swath that contributed the most points in this one locale had the largest error. Considering that each swath only took 1 to 2 seconds to pass through the GPS survey area, an important question is whether or not the issues with error are attributable to environmental variables occurring in the couple seconds of lidar data collection that we have attempted to verify. Many errors are unique to individual datasets (Leigh et al., 2009) and occur as a result of variable conditions during the collection of a single dataset itself. Every dataset is collected in different environmental situations, with issues including but not limited to atmospheric conditions, satellite constellation configuration, mechanical calibration, and/or the sampled

![Figure 8](image)

Figure 8. (A) The slope of the terrain, as recorded by the GPS survey data, correlated positively with lidar error, but there was not a strong correlation; (B), (C), and (D) Statistical metrics of lidar error (ME, MAE, RMSE) for three slope magnitude bins (low, moderate, high) that correlated with the beach, beach/dune transition, and dune face components of the geomorphology.
reflection surface. As is common practice, collecting overlapping swaths of data would help control for any unexpected environmental variables that could result in erroneous data, and it is well established that ground control surveys can be utilized to remove systematic error from a dataset. However, correcting data by extrapolating results outside of the ground control survey area should be approached cautiously.

DEMs and Interpolation Error
In this instance, the RMSE of the DEM was only a few centimeters less accurate than that for the point data itself. The reduced accuracy is expected considering the variable surface that, along with other errors such as incorrect ground classification, could facilitate the introduction of error through interpolation. The small scale of the introduced error supports the use of the raster DEM data structure to represent coastal topography from lidar. Moreover, spatially segregated DEM error within the dune shows similar results relative to the point data. The RMSE value for the beach was negligibly changed; it decreased to within the margin of error expected for the GPS survey itself.

Overall, the results of this experiment provide the values of accuracy that are approximately representative of the error in modeled lidar data stored in grid-cell format collected in a beach/dune system acquired from the provider with no additional processing (Table 5) and processed to remove systematic error (Table 6). These RMSE metrics for each subdivision (i.e., total, beach, dune, non-vegetated dune, and vegetated dune) provide statistical constraints for the accuracy of lidar data collected within each environment. These metrics also form the basis for confidence intervals constraining calculated dimensions of geomorphological change derived from comparison of lidar DEMs collected at different points in time.

Utility in Geomorphological Analyses
Despite the relatively large errors identified through this experiment, we view lidar data as a functional source of topographical elevation that is suitable to quantify geomorphological changes occurring in the coastal setting. The error we identify, less than half of a meter within dunes, provides a vertical resolution fine enough to capture the erosion of a foredune from a single storm event. Volumetric changes to the beach itself can be captured because the volume is largely dependent upon movement of the beach face (topographical feature with >1 m of relief) that can oscillate tens of meters a year landward or seaward. Considering the powerful nature of this dataset that combines: (1) the ability to mesh practically instantaneous data collection with hundred-kilometer-scale coverage; and (2) the horizontal and vertical resolution to capture geomorphologically relevant changes to the coastal landscape, we would recommend that airborne lidar is ideally, collected over sensitive coastal natural resources both frequently and systematically. Practically, cost will limit the amount of lidar collected, but this does not concern the utility of the dataset. Although we recommend the use of lidar for quantitative analyses of coastal geomorphology, we do suggest that the metrics of accuracy provided herein are used to: (1) delineate geomorphological changes that can, and cannot, be currently resolved using lidar (e.g., cm scale variations cannot be distinguished with an error of 38.7 cm RMSE); and (2) to provide statistical confidence limits on dimensions of geomorphological changes derived from comparisons of lidar elevation datasets.

We also note that the foredune that we measured herein is an active foredune that has evolved in tandem with the beach and in close proximity to the land/water interface. Other foredune classifications (i.e., the abandoned and ancestral classifications identified by Psuty et al., 2015) that invoke evolution in different morphodynamic settings or for longer periods of elapsed time might possess topographical and land cover characteristics different from active dunes. Additional work could more completely assess the influence of a greater diversity of coastal topographies on lidar accuracy.

Figure 9. (A) The topography of the GPS survey area portrayed by elevation isolines derived from a DEM generated from the lidar elevation data collected in June 2014; (B) Spatial distribution of differences in elevation between the lidar-derived DEM and the GPS survey; (C) Spatial distribution of differences in elevation between the lidar-derived DEM – shifted 130 cm, 160° to the south, and 9 cm higher - and the GPS survey.
Conclusions
Modern lidar is a highly useful, continually advancing technology that allows for the collection of valuable geospatial data documenting the dimensions of large expanses of the coastal landscape. The utilization of lidar topographical data provides the benefits of a high resolution view of system-wide coastal topographical configurations collected at a “single” moment in time. These data can facilitate high resolution analyses of coastal geomorphological change. However, the tradeoff for this expansive coverage is decreased accuracy compared with traditional survey methods, such as ground-based GPS surveys. The examination of the evolution of coastal topography requires knowledge of the accuracy of the data as a basis for confidence in quantitative results. This information that is not readily available because the metrics of lidar error provided with lidar datasets minimize sources of error associated with the collection of lidar in the coastal environment. This experiment provides metrics of error for spot elevations obtained from airborne lidar elevations of prototypical beach and foredune topography. We find that lidar provides more accurate measurements of flat non-vegetated surfaces, the beach, than it does for slope and/or vegetated terrain, the dunes. The magnitude of error is statistically correlated with the higher slopes of the dune area, and the vegetated portions of the dune are less accurately measured than the non-vegetated dune areas. This indicates that both slope and presence of vegetation contribute to decreased lidar accuracy. The elevations of the vegetated portions of the dune are overestimated by lidar, likely because lidar returns are intercepted by plants prior to contact with the ground surface. Additionally, the results show that lidar error within this environment can be a result of both a systematic horizontal and vertical shift of topographical surface dimensions. This component of the error can be corrected for if accurate ground surface elevations are available.

Acknowledgments

References


US Army Corps of Engineers (USACE), 2010. 2010 US Army Corps of Engineers Topo-Lidar bathymetric survey due to a lack of satellite signal.


Supplemental Material

Detailed Site Description

The southeastern portion of the study area was composed of a relatively flat beach berm that had no vegetation cover. The beach had a very gentle slope, 2° to 8°, dipping to the southeast. The steepest slopes of the beach area within the study site were close to the beach/dune boundary, where the relatively steep face of the dune transitioned into the flatter berm surface (SM Figure 1). The beach face was not incorporated into the experiment, and the intersection of the beach face and NAVD88 was situated approximately 35 to 50 m seaward of the study site boundary.

Measuring 35 to 50 m wide, the prominent, 8.6 m (relative to NAVD88) primary foredune ridge was oriented in a southwest-northeast direction parallel to the shoreline (Figure 3B). Its beach facing slope ranged from approximately 8° to 45° with an aspect of approximately 295° from east. Relative relief along the dune face varied from 4.5 to 6 m from the berm surface to the dune crest. The substrate of the entire study area was sand, but the sand surface of the dune was firmer than the sand surface of beach because of the stability provided by vegetative roots and rhizome structures. Towards the eastern portion of the survey area, the crest possessed a knife-edge configuration, sloping steeply on the landward side of the dune crest. For the majority of the survey area, the sharp slope of the dune face transitioned to a gentler slope closer to the crest. The crest itself was a hardly distinguishable feature because of the gentle slopes on its landward and seaward sides. The landward facing side of the dune beyond the crest area transitioned to a steeper slope leading down to the swale at the landward base of the dune. The landward facing slope was not as steep as the seaward face of the dune, reaching a maximum slope of 35°. At the widest segment of the foredune feature, slopes as high as 20° were oriented 60 to 70 degrees from the strike of the ridge itself, at approximately 75° and 225°, from east. A number of secondary dune features, relatively stable forms largely sheltered from modification by oceanographic processes such as waves and currents, were situated landward of the primary foredune ridge.

Located just seaward of the foredune crest for the majority of the survey area, the vegetation line sat relatively high on the profile. Sparse dune grass was prevalent within a few meters of the dune crest and a small portion of the scarped dune face contained the roots of older dune vegetation. On the landward side of the dune, the sparse dune grass transitioned to dense vegetation, particularly on the sharp slope between the crest line and the swale behind it. The vegetation consisted of thick brush measuring higher than 2 meters in some areas (SM Figure 1). The vegetation was so dense and high, it made a small portion of the survey area impossible to survey due to a lack of satellite signal.

2 m above ground surface

SM Figure 1. Rover pole height of 2 m with receiver attached is located in the swale inland of foredune crest. Vegetation height in portions of the GPS survey area exceeded the 2 m height of the rover pole.
Ground-Level Ozone Concentration and Landscape Patterns in China’s Urban Areas

Jiayi Li and Xin Huang

Abstract
We monitored the spatio-temporal distribution of urban population-weighted ozone for 2014–2017 to investigate ground-level pollution in China. During the study period, the national average was 88.68±10.4 μg/m³ for O₃-8h, and 5.27% of the days that exceed the 160 μg/m³ standard. Pollution hotspots in the Tibetan Plateau are mainly attributed to natural factors, while those in the North China Plain (NCP), Yangtze River Delta (YRD), and Pearl River Delta (PRD) are more closely related to anthropogenic activities. The results indicate that the ozone pollution and its correlation with the landscape are influenced by the ozone regime and climatic factors. In the Yangtze Plain and its southern urban areas in transitional ozone regimes, sprawl, contiguity, compactness, and the size of the urban area also facilitate the accumulation of pollution.

Introduction
During the last few decades in China, the rapid industrialization and urbanization have been accompanied by elevated air pollution at a national scale. Recently, much attention has been paid to ground-level ozone, which is an important risk factor for both human health and vegetation (Feng et al., 2015; Lefohn et al., 2017; Lelieveld et al., 2015; Liu et al., 2018). It has been reported that, in summer, most of the developed and populated city clusters in China, such as the Yangtze River Delta (YRD) (Shao et al., 2016; Shu et al., 2016), the Pearl River Delta (PRD) (GPEMC, 2017), and the Beijing-Tianjin-Hebei (BTH) region (Wang et al., 2017), are facing serious photochemical pollution. As a ground-level secondary pollutant, in addition to the downward transport of stratospheric air and horizontal wind transport, ground-level ozone is generated through chemical reactions, which is a process that requires the presence of sunlight and the participation of emitted precursors. With an abundant diversity of precursors in the different ozone formation regimes and the complicated synoptic meteorological conditions in China, it is a challenging task to analyze the influence of anthropogenic activities on urban ozone pollution.

Since the mid-2000s, several studies have been made of surface photochemical pollution in populated areas in China, including model simulations (Ou et al., 2016), satellite data analyses (Huang et al., 2013; Jin et al., 2017; Jin and Holloway, 2015), and field experiments (Jia et al., 2016; Shu et al., 2016). Due to the coarse spatial resolution of the data, the observation of ozone in the first two approaches has mainly focused on global- or regional-scale studies (Li et al., 2017; Liu et al., 2018). Meanwhile, the most recent satellite remote sensing data can provide us with the precursor distribution at a finer spatial resolution (Jin et al., 2017), which can be used to estimate the formation mode of ground-level ozone. On the other hand, some studies have focused on the dynamic differences between urban and rural sites, based on large cities (Jia et al., 2016) or agglomerations (GPEMC, 2017), using field observations from several ground stations. However, the conclusions from these studies may not be comprehensive or may even be contradictory, partly because of the limitations of the regional climate conditions, the geographical locations, and the development levels. For instance, Huang et al. (2013) argued that tropospheric ozone was insensitive to urbanization in three urban agglomerations in east China, while Li et al. (2017) suggested that the intensified urban heat island phenomenon during urbanization aggravated the ozone pollution and the emission of its precursors in the YRD urban cluster. It is also notable that relatively small cities in China account for a substantial proportion of the population, and the ground-level ozone situation in these areas should thus be monitored. Since 2013, a national-scale network, which tracks and regulates the Environmental Protection Agency (EPA)’s six criteria pollutants (including ground-level ozone), has been gradually constructed by the Chinese Ministry of Ecology and Environment (MEE), to facilitate the monitoring of near real-time in-situ, ground-level ozone. This context provides us with an opportunity to investigate the spatio-temporal characteristics of urban ground-level ozone and its relationships with urbanization in multiple cities.

Urban landscape indicators are metrics that can be used to reveal linkages between urbanization and environmental consequences (Bai et al., 2011; Fang et al., 2016; Pontius and Gilmore, 2017). The built environment in urban landscapes, relating to the surrounding site environment, exhibits unique radiative, thermal, moisture, and aerodynamic properties (Frank and Engelke, 2005; Oke, 1982). The demographic distribution in urban landscapes, influencing the metabolism of the surrounding districts, also portrays the magnitude and extent of human activity (Zhu et al., 2015). In this context, a number of studies have reported that urban landscapes have a significant impact on local air pollution, such as nitrogen oxides (NOx) (Bechle et al., 2011; Larkin et al., 2016), sulfur dioxide (SO₂) (Zou et al., 2007), and fine particles with a diameter of 2.5 μm or less (PM₂.₅) (Larkin et al., 2016; Li et al., 2016). In recent years, the effects of urban planning and spatial optimization methods on ground-level ozone concentration have also been investigated. For instance, taking US urban areas as the study region, population centrality has
been associated with lower concentrations of ground-level ozone (Clark et al., 2011; McCarty and Kaza, 2015), while the urban sprawl level of large metropolitan regions has been positively linked to a greater number of ozone exceedances (Stone, 2008). It is therefore remarkable that so few studies have engaged in the task of estimating the influence of urban form on ground-level ozone. Furthermore, very few studies have paid attention to the distinct functions of the urban form under different ozone formation regimes, especially in China. Hence, there is an urgent need to investigate the relationship between the urban landscape and ground-level ozone pollution, considering the ozone formation regimes, for the different cities in the diverse climate zones of China.

In this study, we built one of the first long-term urban ground-level ozone datasets for China, which includes the ground-level ozone concentration (referred to as “ozone concentration” hereafter) measured during three years of continuous monitoring in the urban areas of all the prefecture-level (or above) cities in China. About 1,563 monitoring stations in 339 cities were operating between June 2014 and May 2017. In this paper, an overview of the spatio-temporal variations of the urban ozone pollution is documented. In the different ozone formation regimes and climate zones, the relationships between several typical urban characteristics (including the built environment and demographic factors) and ozone concentration over the study period are investigated. By studying this extensive dataset, our understanding of the urban ozone pollution at a fine spatio-temporal resolution should be enhanced. The findings of this paper will provide valuable advice for the development of pollution reduction policies for high-pollution days and for the improvement of urban air quality from the perspective of urban landscape design, accounting for the climate conditions and the ozone formation regimes throughout China.

Materials and Methods
Figure 1 shows the framework of the method used to estimate the urban ozone status and reveal its relationship with urban characteristics. The four steps marked in Figure 1 are described in detail in the following.

Study Region and Urban Areas
All the prefecture-level cities in China, except for those in some autonomous regions, were included in this study. The ~30 m gridded summary of national land cover (the China Land Use/Cover Dataset [CLUD]), which includes water bodies, built-up land, etc.) for the year circa 2015 (Liu et al., 2014), and a elevation dataset (URL: http://earthexplorer.usgs.gov/), and the administrative boundaries (URL: http://www.gadm.org) were utilized for the urban area extraction. Details of the urban boundary calculation procedure are shown in Figure 2. Pixels of water bodies or those pixels with elevations exceeding the highest point in the urban area by 50 m were excluded from this analysis.

Climate Zones and Ozone Formation Regimes
According to the Köppen-Geiger climate classification system (Rubel and Kottek, 2010), the mainland of China can be divided into five climate zones by the different climatic conditions (i.e., rainfall and temperature, which are important natural environmental factors related to ozone). EW stands for equatorial climate and warm and fully humid temperate climate; W stands for warm temperate climate with dry winter; A stands for the climate of arid steppe and desert; S stands for snow climate with dry winter; and TS, at the Qinghai-Tibet Plateau, stands for tundra climate and snow climate with cool summer and cold winter.

The Ozone Monitoring Instrument (OMI) onboard NASA’s Earth Observing System Aura satellite can provide us with the major proxies of the ozone precursors, including nitrogen dioxide (NO2) and formaldehyde (HCHO) (Ziemke et al., 2006), which are widely used for determining the ozone formation regime. For the study period, the concentration of monthly NO2 was obtained with a resolution of 0.125° (DOMINO–OMI–NO2v2) and an uncertainty of 1×1015 mol/cm2, and that of the monthly HCHO was acquired with a resolution of 0.25° (BIRA–OMI–HCHOv14) and an overall error of 7×1015 mol/cm2 (De Smedt et al., 2015). Firstly, following the criteria applied in East Asia in a previous study (Jin et al., 2017), the ozone regime for each grid cell was classified as either null, a volatile organic compound (VOC)-limited regime, a NOx-limited regime, or a transitional regime. A grid cell where the average concentration of NOx was less than 2.5×1013 mol/cm2 was set as null and regarded as an insensitive area for anthropogenic activities. We then recorded the percentage of the ozone regime classes within each urban area. For each urban area whose dominant regime was not null, it was assigned as the regime class with the highest percentage when this percentage was larger than 50%; otherwise, it was regarded as a mixed region.

Population-Weighted Ground-Level Ozone Concentration Estimation
Since 2013, a national network of 1,563 monitoring stations has been gradually established by the MEE. In this network, at least one station is located in each prefecture-level city of China. Quality control of the utilized data in this study (from 01 June 2014, to 31 May 2017) was based on the National Ambient Air Quality Standards (NAAQS) (GB 3095–2012) (Chinese MEP, 2012).
As suggested by Clark et al. (2011), the population-weighted (PW) ozone concentration for each urban area was estimated as follows. Firstly, for each hour, the available concentrations from the 12 nearest monitoring stations were fed into the inverse distance weighting based spatial interpolation method to simulate the ozone grid. We estimated the model performance by comparing the modeled results with observations from 1,241 monitoring sites of urban areas in China. In general, the simulated ozone agreed well with the observations, with correlation coefficients greater than 0.99 (p < 0.001). We then used the latest gridded population dataset (POP) (1-km spatial resolution, year 2010), which was obtained from the Data Center for Resources and Environmental Sciences (RESDC) of the Chinese Academy of Sciences (Fu et al., 2014). PW ozone was then calculated as follows:

\[
PW - ozone = \frac{\sum_{i=1}^{n} (ozone_i \times POP_i)}{\sum_{i=1}^{n} POP_i}.
\] (1)

For the urban area of each city, PM, and ozone are the interpolated concentration and the population within each prefecture, respectively, in each 1 km grid-cell center \(i\), where \(n\) is the number of 1 km grid-cell centers within the current urban area. Finally, the annual dynamics for 2014–2017 were calculated, where June to the next May was regarded as one year. Considering the seasonal and diurnal factors, we calculated the daily, daytime, and nighttime average ozone concentrations, and O\(_3\)-8h, for all year round and the ozone episode periods, respectively.

**Urban Characteristics**
The spatial pattern of the built-up areas and population was used to understand and quantify the anthropogenic activity impact on the concentration of ozone (Figure 3). The built-up related metrics were extracted from the CLUD circa 2015 and the POP dataset circa 2010 (Fu et al., 2014). Roads were extracted from Map World (URL: http://www.tianditu.cn/). Detailed descriptions of the urban landscape metrics are provided in Table 2, and the descriptive statistics are listed in Table 3.

**Results**
The pollution situations of the 339 urban areas in the mainland of China, grouped by the five climate zones, were analyzed. Both the spatio-temporal distribution and the seasonal and diurnal variation of the pollution in China were also analyzed. Spearman's rank correlation coefficients were then calculated to investigate the relationships between ozone pollution during the ozone season and the urban landscape metrics.

**Overview and Spatial Distribution of Ozone Pollution in China**
Figure 4 presents the spatial distribution of the triennial average concentrations of the 339 cities in China. Taking all the cities as a whole, the triennial average was 55.34 ± 7.77 μg/m\(^3\) for daily ozone and 88.68 ± 10.41 μg/m\(^3\) for O\(_3\)-8h over the study period. The triennial average O\(_3\) -8h concentration ranged from 60.75 ± 33.50 μg/m\(^3\) (Urumqi in the A climate zone) to 119.46 ± 55.47 μg/m\(^3\) (Weifang in the W climate zone). As can be seen in Figure 4, the highest average concentrations of all four metrics (i.e., daily average, daytime, nighttime, and O\(_3\)-8h) were found in the TS climate zone, the North China Plain (NCP) (in the W zone), and the YRD (in the EW zone). In the meantime, cities in the PRD (in the EW zone) also suffered from unignorable ozone pollution in both daytime and nighttime. Ozone pollution in the TS urban areas is less related to anthropogenic sources, as the ozone formation regimes of the

![Figure 3](image-url)

**Figure 3.** Elongation and contiguity of the built-up patches within the urban areas, illustrated by (a) Zunyi, in the EW climate zone (0.85, 9, and 88 for RCC, LSI, and the three-year average concentration of PW O3-8h, respectively); (b) Liuzhou in the EW climate zone (0.64, 8, 99); (c) Shiyan in the W climate zone (0.55, 17, 95); and (d) Changzhou in the W climate zone (0.55, 16, 125).
14 urban areas (Table 3, 14/15 = 93.3%) are assigned as null. Similar nocturnal ozone concentration results are seen in the high-altitude urban areas (Figure 4d).

### Temporal Variation of Ozone Pollution in China

Fine weather with intense sunlight, high temperature, and low wind velocity is an ideal breeding ground for ozone episodes. In view of the diurnal variation, the lower ozone concentration during the nighttime can be explained by the low temperature, as well as dry deposition. As can be seen in Figure 5, for all the metrics except for nocturnal ozone, the W and S zones

<table>
<thead>
<tr>
<th>Table 2. Descriptive statistics of the urban landscape metrics grouped by climate zones and ozone formation regimes.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regime</strong></td>
</tr>
<tr>
<td><strong>POP</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>POP</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>POP</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>CA</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

* POPD is the abbreviation for population density, and is equal to POPI divided by CA

<table>
<thead>
<tr>
<th>Table 4. Non-attainment days and exceedance rates for the Grade II standard (160 µg/m³) and Grade I standard (100 µg/m³) for each climate zone and China as a whole.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade standard</strong></td>
</tr>
<tr>
<td><strong>Jan.</strong></td>
</tr>
<tr>
<td><strong>Feb.</strong></td>
</tr>
<tr>
<td><strong>Mar.</strong></td>
</tr>
<tr>
<td><strong>Apr.</strong></td>
</tr>
<tr>
<td><strong>May</strong></td>
</tr>
<tr>
<td><strong>Jun.</strong></td>
</tr>
<tr>
<td><strong>Jul.</strong></td>
</tr>
<tr>
<td><strong>Aug.</strong></td>
</tr>
<tr>
<td><strong>Sep.</strong></td>
</tr>
<tr>
<td><strong>Oct.</strong></td>
</tr>
<tr>
<td><strong>Nov.</strong></td>
</tr>
<tr>
<td><strong>Dec.</strong></td>
</tr>
<tr>
<td><strong>All</strong></td>
</tr>
<tr>
<td><strong>Ozone season</strong></td>
</tr>
<tr>
<td><strong>China</strong></td>
</tr>
</tbody>
</table>

* Note: N: n mean the number of non-attainment days and the exceedance rate (unit: %) for all cities in each zone

**Table 3. Number of urban areas assigned by the ozone formation regime across the five climate zones.**

<table>
<thead>
<tr>
<th><strong>NOx-limited</strong></th>
<th><strong>Transitional</strong></th>
<th><strong>VOC-limited</strong></th>
<th><strong>Null</strong></th>
<th><strong>Mixed</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EW</strong></td>
<td>37</td>
<td>27</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>W</strong></td>
<td>31</td>
<td>41</td>
<td>2</td>
<td>43</td>
</tr>
<tr>
<td><strong>A</strong></td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td><strong>S</strong></td>
<td>14</td>
<td>19</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td><strong>TS</strong></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td><strong>China</strong></td>
<td>86</td>
<td>97</td>
<td>5</td>
<td>150</td>
</tr>
</tbody>
</table>
present the highest average concentrations, followed by the A, EW, and TS zones in the growing season (spring and summer), while the TS and EW zones show a higher concentration than the others in winter. For nocturnal ozone, the TS zone shows a higher concentration than the other regions. With regard to seasonality, significant variability can be seen, with the highest concentration in the growing season and the lowest in the winter. Taking China as a whole, all the ozone measurements are presented as the approximated unimodal distribution by season (see the green bars in Figure 5). Besides May as the most severe period for all the climate zones, the other seriously affected months are September in the EW zone and July for both the A and S zones. According to the NAAQS of China, the non-attainment days that exceeded the O$_3$-8h limit (i.e., 160 µg/m$^3$ for urban areas) for each zone are listed in Table 4. For O$_3$-8h, daily ozone, and daytime ozone, the most severe exceedance rates for the EW, W, and S climate zones are from March to October, while the most severe exceedance rates for the A zone are from May to July, and the most severe exceedance rates for the S zone are from May to August. The ozone

Figure 4. Triennial average PW concentrations of daily O$_3$, O$_3$-8h, daytime O$_3$, and nighttime O$_3$ for each urban area of the 339 cities. The shading color for each city relates to the climate zone.

Figure 5. Monthly averages of the four ozone metrics in China for the five climate zones
season for each zone in this study is defined by this result. The exceedance rate during the ozone season (3.18%–11.32%) is approximately twice that of all the available days (1.04%–
6.80%), and pollution in the W and EW zones is more frequent. For all the climate zones, the non-attainment days begin to sharply increase in April, and the most frequent ozone pollution occurs in May, while several of the urban areas in the S and TS climate zones still suffer from ozone pollution in July.

The Relationships Between Urban Ozone Pollution and Landscape Patterns

Spearman’s rank correlation coefficients were applied to quantify the relationships between ozone concentrations during the ozone season and the urban landscape metrics across both climate zones and ozone formation regimes. According to the regime criteria (Table 3), the most polluted urban areas are located in the EW and W zones, while there are also 10 instances in the A zone and 33 instances in the S zone. As shown in Table 3, more than half the urban areas are assigned as precursor-controlled regions, most of which are located in the EW and W zones. The correlations for all the pollutant measurements are listed in Table 5, as represented by the O3-8h results detailed below.

For the demographic distribution, only the polluted urban areas assigned as transitional regimes indicate some correlation to O3-8h (Table 5). In detail, population in the EW region is moderately and positively associated with O3-8h concentration (\( r = 0.44, p < 0.05 \)). In addition, population heterogeneity (POPV) in the W climate zone facilitates the alleviation of O3-8h concentration (\( r = -0.40, p < 0.05 \)). In addition, the POPV in the NOx-limited EW areas is positively associated with ozone concentration (\( r_p < 0.05 \)). In addition, the POPV in the NOx-limited EW areas is positively associated with ozone concentration (\( r = 0.44, p < 0.05 \)).

Table 5. Spearman’s rank correlation coefficients in the different ozone regimes between the O3-8h during the ozone season and the landscape metrics.

<table>
<thead>
<tr>
<th>Ozone regime</th>
<th>Climate zone</th>
<th>EW</th>
<th>W</th>
<th>A</th>
<th>S</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. 37</td>
<td>27</td>
<td>31</td>
<td>41</td>
<td>10</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>POPF</td>
<td>-0.06</td>
<td>-0.19</td>
<td>0.28</td>
<td>-0.08</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>POPV</td>
<td>-0.05</td>
<td>-0.23</td>
<td>0.30</td>
<td>-0.15</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>CA</td>
<td>-0.16</td>
<td>0.48</td>
<td>0.28</td>
<td>-0.08</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>LS1</td>
<td>0.05</td>
<td>0.42</td>
<td>0.28</td>
<td>-0.08</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>RCC</td>
<td>-0.17</td>
<td>-0.25</td>
<td>0.03</td>
<td>0.31</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>DIST</td>
<td>-0.04</td>
<td>-0.42</td>
<td>0.30</td>
<td>-0.15</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>POPF</td>
<td>-0.09</td>
<td>0.34</td>
<td>0.07</td>
<td>0.13</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>POPV</td>
<td>-0.38</td>
<td>-0.23</td>
<td>0.13</td>
<td>0.31</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>CA</td>
<td>-0.21</td>
<td>0.47</td>
<td>0.13</td>
<td>0.33</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>NP</td>
<td>0.07</td>
<td>0.54</td>
<td>0.15</td>
<td>0.32</td>
<td>0.27</td>
<td>0.15</td>
</tr>
<tr>
<td>LS1</td>
<td>-0.02</td>
<td>0.51</td>
<td>-0.12</td>
<td>0.18</td>
<td>0.54</td>
<td>0.17</td>
</tr>
<tr>
<td>RCC</td>
<td>-0.11</td>
<td>-0.30</td>
<td>-0.15</td>
<td>0.49</td>
<td>0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>DIST</td>
<td>0.30</td>
<td>-0.65</td>
<td>0.17</td>
<td>-0.69</td>
<td>0.53</td>
<td>0.49</td>
</tr>
<tr>
<td>RD</td>
<td>-0.10</td>
<td>0.38</td>
<td>-0.21</td>
<td>-0.17</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td>POPF</td>
<td>-0.16</td>
<td>0.38</td>
<td>0.33</td>
<td>0.15</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>POPV</td>
<td>-0.01</td>
<td>-0.22</td>
<td>-0.12</td>
<td>-0.43</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>CA</td>
<td>-0.08</td>
<td>0.52</td>
<td>0.42</td>
<td>0.31</td>
<td>0.08</td>
<td>0.25</td>
</tr>
<tr>
<td>NP</td>
<td>0.16</td>
<td>0.59</td>
<td>0.43</td>
<td>0.29</td>
<td>0.19</td>
<td>0.30</td>
</tr>
<tr>
<td>LS1</td>
<td>-0.14</td>
<td>0.55</td>
<td>0.13</td>
<td>0.19</td>
<td>0.54</td>
<td>0.01</td>
</tr>
<tr>
<td>RCC</td>
<td>-0.05</td>
<td>-0.36</td>
<td>-0.27</td>
<td>-0.17</td>
<td>0.61</td>
<td>0.14</td>
</tr>
<tr>
<td>DIST</td>
<td>-0.06</td>
<td>-0.69</td>
<td>-0.45</td>
<td>-0.72</td>
<td>-0.48</td>
<td>-0.37</td>
</tr>
<tr>
<td>RD</td>
<td>-0.10</td>
<td>0.38</td>
<td>-0.21</td>
<td>-0.17</td>
<td>0.14</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Discussion

The rapid urbanization in China has focused public attention on both environmental protection and urban planning. Against a backdrop of rising ground-level ozone pollution, subject to the measurement limitations and the complicated chemical process, the ozone pollution situation of urban areas in China is still unclear. In this paper, we have analyzed the ozone concentration data collected from the newest national air quality monitoring network in 339 prefecture-level (or above) cities from June 2014 to May 2017. By reporting the situation of the PW ground-level ozone pollution in urban areas nationwide and describing the associations with urban characteristics, this research will help to expand our scientific understanding, and
has important theoretical and management implications. To mitigate the impact of urban characteristics on ozone, urban planners will be able to gain insights into optimizing the spatial distribution of both demographic and physical layouts.

In contrast to the recent studies that have mainly focused on developed regions (Shao et al., 2016; Shu et al., 2016), this paper has reported on the ozone status in the urban areas of multiple cities. As an addition to the survey of ozone and the proportion of days violating the NAAQS standard (160 µg/m³) of 15 typical cities in 2015 (Liu et al., 2018), this research further reports on the most recent three-year situation in all the urban areas of China and the associated non-attainment rates. During the study period, the national triennial average was 88.68±10.41 µg/m³ for O₃-8h, and around 5.27% of the days were non-attainment days that exceeded the new NAAQS standard of China. However, this rate is much lower than that reported in Liu et al. (2018), as the typical cities in this previous study are metropolises located in heavily polluted regions (Figure 4). In addition, it is also demonstrated that the stringent guideline (i.e., 100 µg/m³ for O₃-8h) is still hard to meet, as more than 30% of the days for each climate zone exceeded this standard between June 2014 to May 2017 (Table 4).

In view of the spatial distribution, the high ozone concentrations to the west of the Aihui-Tengchong Line are mainly due to natural activity (such as the strong UV and vertical transportation), while those in the EW and W climate zones are more closely associated with human activity (Table 3). In these zones, NCP, YRD, and PRD are pollution hot spots, and local industry and transportation are the dominant anthropogenic sources (Wang et al., 2017). In view of the temporal distribution, the urban ozone concentrations show a remarkable seasonal and diurnal variability, which is strongly affected by the synoptic meteorological conditions (i.e., high temperature and stagnant weather). Overall, the spatio-temporal results suggest that specific control measures and mitigation strategies should be targeted in the different regions of China, in terms of the different features of local/regional emissions and meteorology. The Spearman’s rank bivariate correlation analysis was conducted by controlling the ozone sensitivity and weather conditions, and the associations with the spatial distribution of population and built-up land were investigated during the ozone season. Urban residents, who are not only the victims of pollution but also the source of the pollutant precursors, can be expected to have a great effect on urban ozone concentrations. As the proportion of urban residents is expected to reach 70–80% by the year 2030, this effect is worthy of attention. At a certain level of population scale, large variation is often associated with resource inequity and suburban expansion, which can intensify air pollution in urban areas. As expected, in the EW climate zone, the ozone concentration shows a concomitant increase in population size for the transitional regime areas and heterogeneity (POPOP) for the NOX-limited urban areas. However, the negative relationship between POPOP and ozone concentration for the transitional regime areas in the W climate zone is counterintuitive. A potential reason for this may be the large variation of population but modest population density (see POPI and population density (POPOP) in Table 2) in this region.

At the same time, spatial patterns of emissions and pollutant concentrations can also be altered by city planning; for instance, the spatial distribution of built-up patches. Firstly, in the NOX-limited regions, the insignificant correlations indicate that the contribution to pollution from anthropogenic VOC emissions (in the built environment) is minor. It is also noted that no urbanization-induced factor is associated with ozone pollution in the S zone, which is partly due to the limited sample size and the decrease of concentration with the increase of latitude (Figure 4). With regard to the precursor-controlled areas, increasing the total area of built-up land, urban sprawl in physical expansion, and built-up contiguity can lead to aggravation of the ozone concentration, which can be partly explained by increased vehicle exhaust emissions and fuel volatilization. In addition, the positive associations between DEST and ozone concentration suggest that the sea-land breeze acts as important meteorological assistance to distribute ozone in coastal cities, while the anticyclones and tropical cyclones in the western Pacific promote pollution production, recirculation, and accumulation. Several studies have noted this phenomenon in developed cities in both the YRD and PRD (Wang et al., 2017), and we have further proved the existence nationwide. As the sea-land breeze can be intensified by the coastal urban heat island phenomenon (Li et al., 2017), further attention to ozone and its precursors for the coastal cities is needed.

It is of interest to link the emerging theory of urban air environment protection and the management of urban landscapes, as it provides a new idea for stringent ozone pollution control measures. Encouraging residents to move out of urban areas has long been viewed as an effective way to mitigate ozone pollution. Essentially, large population variations can be mainly attributed to the uneven distribution of resources, as households vote with their feet to achieve a balance between the cost of living and local public goods. Many first-tier cities, such as Beijing, Shanghai, and Guangzhou, have established goals and implemented strategies to re-home millions of downtown residents. Our results should inspire a similar approach for the small cities. Meanwhile, with the private car annual growth rate at 22.76% in Chinese cities (Zheng and Kahn, 2013), this could further aggravate ozone pollution. In addition to fuel taxes, congestion tolls, vehicles with low emission, etc., planning the built-up environment to facilitate the dilution ability and reduce travel time could also be taken into consideration. The results from this study will provide additional evidence for promoting the management and implementation of related population policies and improving the urban ecology and air environment, particularly for locations where urbanization is still in process.

This study does have some limitations. The unbalanced spatial distribution of the in-situ stations and the inverse distance weighted interpolation suggested in previous studies (Clark et al., 2011; Stone, 2008) in developed nations, whereas the static ozone pollutant situation was addressed in this study. It is also noted that the most recent three-year average was utilized to avoid the specificity from a given year.

Conclusions

This paper has presented the ground-level ozone concentrations measured during three years of continuous monitoring in 339 prefecture-level cities in China between 2014 and 2017, and we have also described the relationships between human activity and ozone concentration in the most polluted cities. During the study period, the national triennial average was 55.34±7.77 µg/m³ for daily ozone and 88.68±10.41 µg/m³ for O₃-8h, and 5.27% of the days were non-attainment days that exceeded the new NAAQS standard (160 µg/m³) of China. The urban ozone concentrations showed remarkable seasonal, diurnal, and spatial variability, which suggests that specific control measures and mitigation strategies should be targeted in the different regions of China, in terms of the different features of local/regional emissions and meteorology. It was found that both the spatial arrangement (including composition and configuration) of built-up areas and population density affect the concentration of ozone, which could provide us with important insights into the ecological design and management of urban areas. This
suggests that the impact of urbanization on ozone could be mitigated by balancing the spatial distribution of urban residents and optimizing the built-up area configuration in China.

Acknowledgements
The research was supported by the National Natural Science Foundation of China under Grants 41522110, 4176382 and 41771360, the National Program for Support of Top-notch Young Professionals, the Hubei Provincial Natural Science Foundation of China under Grant 2017CFA029, Hubei Provincial Natural Science Foundation Project under Grant 2017CGB188, Open Fund of Ministry of Education key Laboratory under Grant 2016LSDMIS04, and the National Key R&D Program of China under Grant 2016YFB0501403 and 2017YFB0504103.

References


ASPRS Offers

» Cutting-edge conference programs
» Professional development workshops
» Accredited professional certifications
» Scholarships and awards
» Career advancing mentoring programs
» *PE&RS*, the scientific journal of ASPRS

asprs.org
IN THE AERIAL IMAGERY WORLD THIS IS SANTA’S WORKSHOP

FREE TRIAL OFFER
Check out our massive cloud-based library of 2-3” imagery through our viewing/measuring tools, API or our ArcGIS plug in. Take advantage now.

Geomni
www.geomni.net/psm