Spatial Data Authentication Using Mathematical Visualization

Vert, G., Harris, F., Nasser, S.
Dept. of Computer Science, University of Nevada-Reno, Reno, NV 89557
e-mail: gvert@cs.unr.edu

Summary **–** The internet has become an increasingly compromised method to transmit any type of data including spatial data. Due to the criticality of spatial data in decision making processes that range from military targeting to urban planning it is vital that transmission of spatial data be secure. Cryptographic methods can be utilized for this purpose, however they can be relatively slow, especially when encrypting voluminous quantities of data such as is found with spatial data. A new method of low overhead spatially base authentication has been developed that considers the angular and temporal relationships of spatial object data. This method has been initially shown to be extremely fast. Additionally the method can be visualized which makes it easy for users to detect modifications to data and has the added benefit that modified objects are pointed to in the visualization. This allows users to ignore modified objects or selectively have them retransmitted, neither of these properties are found in current cryptographic authentication methods. Additionally a taxonomy of spatial data developed in previous work can be used to improve the authentication process by allowing it to select only extremely relevant objects to utilize in the new method for authentication.

1.0 INTRODUCTION

The importance of the Internet has grown widely in the past couple of decades from being government network to being used by almost everyone at work or at home. Tremendous amounts of data are being transmitted every second. Since the whole world is connected to the same external network through the Internet, securing the transmitted data has become an important issue because some hackers or ill-intentioned people may attempt to manipulate that data for their purposes which may include terrorism, sabotage, political reasons, and other ill-intentioned acts.

Spatial data sets or maps get transmitted over the Internet all the time for planning processes and decision-making support ranging from resource management to urban planning [2]. This highlights the need to create techniques to protect and secure the transmitted spatial data. To give an example of the significance of spatial data, let us take its important role in military targeting operations. Spatial data is used to determine targets to attack during a war. If the spatial data used is not accurate or has been modified, then this will lead to erroneous destruction and killing of innocents. For example, if the coordinates of a military target have been altered to a new map coordinates that may point to a school or a housing complex, by terrorists during transmission, the outcome will be disastrous. This suggests that there is a need to authenticate the data at the receiving end of a network transmission.

Authentication is a method of determining whether an item is genuine or authentic and properly described. It enables computers at the receiving end to verify the contents of the message. [4, 5] Authentication can range from simple functions such as using passwords to very complicated identifiers. Biometric systems use physical characteristics for authentication. For example, a person's fingerprint stays the same throughout his/her life. This triggered the idea of potentially borrowing the concept of biometric systems and applying it to spatial data. However, because spatial data has a temporal component to it, the concepts of biometrics should be extended somehow to include this in any type of authentication process.

In this paper a new authentication technique for spatial data transmission will be introduced. The technique will be based on the Spicule visualization and utilizes a new spatial data taxonomy recently developed in preceding research.

1.1 Background

There are many techniques used to secure data, during both storage time and transmission. Some of those are enforced by Database Management Systems and some are enforced by communication and security protocols. Encryption is the most widely used technique to protect data. There are many encryption techniques available and commonly used such as systematic encryption, and Public-Key

encryption. Encryption is the process of changing data from its clear form into a cipher form to prevent other unauthorized people from reading the data. Encryption is applied on data to assure its privacy and confidentiality. Basically, there are two main types of encryption: a) symmetric encryption, and b) public-key encryption.

Symmetric encryption is also referred to as conventional encryption. It is the type of encryption that uses one single secret-key for the purpose of encryption and decrypting the text. A symmetric encryption has five ingredients: plaintext, encryption algorithm, secret key, cipher-text, and decryption algorithm. The security of symmetric encryption depends on the secrecy of the secret key itself. The second major encryption technique is Public-Key encryption. Public-Key encryption uses two keys instead of one; one for the encryption process, and the other for decryption. Each sender and receiver has a pair of two keys: public-key and private-key. So the sender encrypts the message using the public key (of receiver), and the receiver decrypts the message using its private key (of receiver).

Encryption can be used for the purpose of digital signatures and thus authentication where the sender signs the message using its own private key (of sender), and the receiver verifies the signature by decrypting the signature using the public key (of sender). A public-key encryption has six major ingredients: plaintext, encryption algorithm, public key, cipher text, decryption algorithm, and private key.

When one considers the application of such methods to spatial data there are several questions that must be considered:

- cryptographic algorithms tend to be designed to work on relatively small amounts of data and thus can be computationally expensive and slow to encrypt and decrypt
- when considering the application to spatial data, the question becomes which data needs to be encrypted. Should all of the data be encrypted or can some of the data, and if such, which data should be encrypted for transmission

As an example, our research ran a test encryption of a three page document, sparsely populated with text utilizing PGP on a dedicated computer (2.4 Mhz) and found that on average it took around two seconds to encrypt this very small amount of data. Due to large amounts of spatial data that can be part of a spatial data set, encryption of all of this data unacceptably slow, especially as one approach near real time transmission of spatial data over a network. A reduction in encryption time might be found if one only encrypted part of the data in a spatial data set. However, this leads to the second question which is determining what data in a spatial data set is truly significant and should be encrypted.

Very little work appears to have been done on the development of authentication methods based on properties found in spatial data. This has became the motivation for the development of a new method for doing spatial data authentication inspired from the concepts of biometrics. This approach utilizes a taxonomy to select data based on its temporality and a visualized mathematical model to generate a geometric signature for the data sets that can be authentication and will point to the modified objects in a spatial dataset.

2.0 THE PROBLEM

Creating a reliable authentication signature need to address which spatial objects and how many of them need to be included to reliably authenticate a spatial data set. Secondly, there is a need to design an fast, intuitive authentication algorithm or method that can then be used to analyze spatial data and ideally point to modified objects.

The first question of how to identify spatial objects for authentication signatures is based on research similar to the classification categories of Peuquet [2]. Our research extended this previous work classify spatial object based on the effect time has on. That means every object was studied with respect to time, and what changes can occur to that object due to time. We define the term degree of temporality as being how long it takes an object to change its spatial geometry and define this concept as:

$$Degree\ of\ temporalit\ y = \frac{\Delta\ Spatial\ \ Geometry}{Time}$$

From this definition, we derive from Peuquets work to define the following classifications for objects:

- temporal continuous (TC) – an object whose degree of temporality and attributes change continuously
- temporal sporadic (TS) – an object whose degree of temporality and attributes change in an unpredictable fashion
- static (S) – an object that has no change in degree of temporality or attributes
- Static-temporal (ST) – an object that is most of the time static, but may under certain situations may have changes in degree of temporality and attributes

Previous to this work current work we have developed a taxonomy of spatial objects. Table 1 presents some objects from this taxonomy. As shown in Table 2, static objects possess attributes that are least affected by time. Static-temporal objects also may naturally be classified as static but may have certain or special cases that are temporal. Temporal-continuous objects contain attributes that change continuously over any period of time. For example, rivers are categorized as temporal-continuous because rivers are continuously changing. The river may become wider or narrower during different seasons. Finally, temporal-sporadic objects may have their attributes change at a certain point in time but not continuously.

Group	***Class***	***Innate***	***Relative Extent***	***Seasonal***	***Continual***	***Singular***	***C.***
Body of Water							
	Ocean	Yes	Large				**S**
	Ice Mass		Large	Yes			**TC**
	Sea	Yes	Vary				**ST**
	Lake		Vary	Yes			**TC**
	River		Vary	Yes			**TC**
	Desert Water		Vary	Yes			**TC**
Surrounding Areas of Water Surface							
	Island	Yes					**S**
	Shore	Yes			Yes		**TC**
	Port					Yes	**TS**
	Dam/Weir					Yes	**TS**
Soil							
	Sand	Yes			Yes		**TC**
	Silt			Yes			**TC**
	Clay			Yes			**TC**
	Rocks	Yes					**S**
Vegetation							
	Land		Vary			Yes	**TS**
	Forest	Yes	Large				**S**
	Farm		Vary			Yes	**TS**

	Park		Vary			Yes	**TS**
	Bushes		Vary	Yes			**TC**
Elevation							
	Summit	Yes					**S**
	Mountain	Yes	Large				**S**
	Hill	Yes	Vary				**S**
	Valley	Yes	Vary				**S**
	Mine		Vary			Yes	**TS**
Urban							
	County		Large			Yes	**TS**
	City		Large			Yes	**TS**
	University		Vary			Yes	**TS**
	Building		Vary			Yes	**TS**
	National Monument					Yes	**TS**
	Voting District		Large			Yes	**TS**
	Parcel Data		Large			Yes	**TS**
	Sewer and Water Systems					Yes	**TS**
	Communication and Electricity						
Rural							
	County		Large			Yes	**TS**
	Village		Large			Yes	**TS**
	Town		Large			Yes	**TS**
	Building		Vary			Yes	**TS**
	Corral		Vary			Yes	**TS**
	Voting District		Large			Yes	**TS**
	Parcel Data		Large			Yes	**TS**
	Septic Systems and Wells					Yes	**TS**
Transportation							
	Road					Yes	**TS**
	Trail					Yes	**TS**
	Crossing					Yes	**TS**
	Railroad					Yes	**TS**
	Airport		Vary			Yes	**TS**
	Bypass					Yes	**TS**
	Service Facility		Vary			Yes	**TS**

Table 1. Temporal classification of spatial data classes based on attributes from the taxonomy developed in previous work

Static	***Static-Temporal***	***Temporal-Continuous***	***Temporal-Sporadic***
Ocean	Sea	Ice Mass	Port
Island		Desert water	Land
Rocks		Shore	Farm
Forest		Sand	Park
Summit		Silt	University
Mountain		Clay	Parcel data
Hill		Bushes	Corral
Valley		Lake	Dam/Weir
		River	Mines
			County
			Building
			National Monument
			Voting Districts
			Road
			Crossing
			Railroad
			Bypass
			Service Facility
			Septic sys/ Wells
			City
			Airport
			Village
			Town
			Trail
			Sewer sys/ water sys.
			Communications & Electricity

Table 2. Temporality classification of the taxonomy

3.0 AUTHENTICATION OF SPATIAL DATA

The spatial taxonomy provide solutions to the first two problems of authentication mentioned previously. The first of these was that of the naming of spatial objects and secondly the selection of which objects could provide a unique signature for a spatial data set considering time and spatial extent. Specifically when we understood the classification of spatial objects we decided that authentication could and should initially be done utilizing static objects. However future work does point to an investigation of objects with temporal properties in authentication. The following section therefore examines the development of an new authentication method based on the use of static spatial objects.

3.1 Spicule Visualization Tool

After selecting a representative set of spatial data objects, that set of objects was be used to build a mathematical signature for the authentication process. One method of creating the signature is to utilize the Spicule, which is a tool developed for the mathematical analysis and visualization of 3D data in which vectors of the spicule are mapped to objects in 3D spaces [6]. Originally, Spicule was developed as a possible tool for conducting intrusion detection utilizing the visual intuitiveness of computer graphics. The Spicule's mathematics are based in vector algebra, and thus there is an algebra that exists for comparing two Spicules. Specifically, if the mathematical representation of two Spicules is subtracted a "change form" is created. The change form can be visualized which then results in a smooth featureless 3D ball if the two versions of the Spicule authentication signature are similar. The advantage of this is that it is simple and visually intuitive to recognize change with out having to conduct analysis or inspection of the underlying mathematica1 data. Figure 1 shows an example of the spicule.

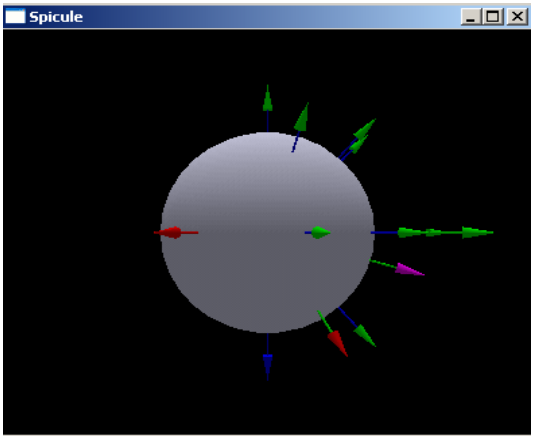

Figure 1. Sample picture of the Spicule

Takeyama and Couclelis have shown that GIS layering abstraction of a location is equivalent to a set of multiple attributes [9]. So, the map can be looked at as a 3D-set of layers on top of each other. In this 3D paradigm of layered spatial data, the

spicule can be utilized to create a mathematical signature for authenticating spatial data by mapping the tips of vectors on the spicule to the unique spatial objects identified from the taxonomy. The signature that can be generated using this approach becomes an n tuple which can be visually subtracted using Spicule to detect changes in the spatial data. This n tuple consists of information about a specific spatial objects vector consisting of a unique set of attributes such as magnitude, angular orientation and location of a vector on the 3D central ball. The vectors can be mapped to objects in various layers of spatial data objects (mentioned above), thus creating vectors that are not tied to the objects in a given layer, increasing the uniqueness of the signature. The number of vectors going from the spicule will be equal to the number of selected objects from the taxonomy in the spatial dataset being authenticated. The collection of these vectors for a given set can be used to describe a unique signature for a particular GIS data set.

The idea behind the proposed authentication process is to utilize the spicule tool to create a geometrical vector for each of several spatial objects selected from the spatial temporal plots. Vectors be point from the center of the Spicule to the (x, y, z) coordinates of a spatial object. Each vector is unique and has three attributes that are represented as follows:

$$V_i = (\text{degreesVertical,degreesEquator,magnitude})$$

In this scheme there is a vector pointing from the center of the spicule, at the origin, to each point or spatial object selected from the taxonomic spatial temporal plot for signature. The three data layers are initially proposed to be placed at one vertical unit apart from the spicule layer. So, the first layer points will have coordinates of (x, y, 1), the second layer points' coordinates will be (x, y, 2), and the third layer points' coordinates will be (x, y, 3). Based on this the vector attributes for each authentication point in the three layers will be:

$$Mag_i = \sqrt{x^2 + y^2 + i^2} \qquad (1)$$

where:

i is the data layer number.

x, y are point original coordinates.

Mag_i is the magnitude of the vector from (0,0,0) to a point in layer i.

$$Sin\theta_{ei} = \frac{x}{\sqrt{x^2 + y^2}} \Rightarrow \theta_{ei} = Sin^{-1}\frac{x}{\sqrt{x^2 + y^2}} \qquad (2)$$

$$Sin\theta_{vi} = \frac{i}{\sqrt{i^2 + y^2}} \Rightarrow \theta_{vi} = Sin^{-1}\frac{i}{\sqrt{i^2 + y^2}} \qquad (3)$$

Equations (2) and (3) are used to calculate the equator and the vertical angles respectively,

where:

i is the data layer number.

θ_{vi} is the vertical angle degrees for a vector from (0,0,0) to a point in layer i.

θ_{ei} is the equator angle degrees for a vector from (0,0,0) to a point in layer i.

The collection of attributes and angles for all authentication vectors forms a two-dimensional matrix that is used as for the authentication signature and the Spicule visualization authentication process (figure 5).

The signature calculation process is done when a spatial dataset is requested to be transmitted over the internet. Table 3 shows a sample calculated vector matrix.

Object ID	Layer	Mag_i	θ_{vi}	θ_{ei}
1	3	7.68	66.8	18.43
2	2	16.31	42.51	4.76
		. .	. .	. .
n	i	29.22	51.95	3.18

Table 3. Sample calculated vector matrix

At the receiving end, the same process to create a signature matrix from the received spatial dataset was applied. By visualizing the mathematical difference between the received spatial data sets matrix and the transmitted matrix, it can be determined if the dataset has been intercepted or altered during transmission. This process may be described by:

IF Visual Mathematical Difference = Zero THEN
No Interception or Alternation.

In the above method if the visual mathematical difference between the two matrices does not equal to zero, it is assumed that the spatial dataset has been intercepted and altered. However,

we can not determine the extent and the type of change that have been made because removal, addition, or movement of a given spatial object or point may result in the change of sequence for many vectors in the matrix after the point of modification in the matrix. Figure 5 shows detection of alternation, thus authentication failure by use of the difference operator. Additionally the vector in the change form points to the data that was modified.

4.0 COMPARATIVE AUTHENTICATION SIGNATURE GENERATION PERFORMANCE

Spatial data may be protected for transmission by encryption or by the generation of a signature using MD5, SHA or RIPEMD. In order to compare the performance of the spatial signature approach to that of above traditional methods a test suite was set up on a PC running at 2.4ghz with a P4 processor. The Crypto++ package was utilized for comparison with timing figures measured down to the millisecond. Crypto++ has a program call Cryptest that may be called with command line switch to encrypt symmetrically, decrypt and generate SHA, MD5 and RIPEMD160 digests. The command line interface was invoked from a command line shell generated with Visual Studio. Because Cryptest was being called using a system command from inside the compiled test program, the first part of the test suite called the operating system shell to load a simple C program. This allowed us to measure the effect on performance of just loading a simple program. Of note in the spatial signature generation test, this test selects increasingly more and more static spatial objects from the test data which are part of the objects from the previous work with taxonomies mentioned above. The above test was run thirty times for each part of the above test program with the following results:

Test Type	Pass 1 (10x)	Pass 2 (10x)	Pass 3 (10x)
Shell	63.00	58.00	57.00
Encrypt (symmetric)	126.60	123.4	121.90
Decrypt (symmetric)	115.60	123.5	121.90
MD5/SHA/RIPEMD	67.20	67.20	64.00
Spatial Authentication	< .01 millisecond	< .01 millisecond	< .01 millisecond

Table 4 Average performance comparison of Spatial Authentication versus Symmetric encryption, SHA, MD5, RIPED (milli seconds) on test data

4.1 Visualization of Authentication Signature and Change Detection Algebra

The Spicule has a unique vector algebra that has two very distant benefits [6]:

- a user does not have to scan the generated signature arrays shown in previous examples
- if changes have been made to the spatial signature, vectors point to the object in the signature that was modified.
- change detection is extremely fast because it is visually intuitive
- allow a user to visually quantify the amount of modification.

The property off having the modified object pointed to can be utilized by a user to:

- have just the modified object retransmitted
- decide the modified object is not being utilized and skip retransmission
- decide that the object is critical and have all data retransmitted

Change detection in Spicule is implemented by the definition of an a visual algebra and is performed through application of the difference operator ("-"). When this operator is applied to two exactly similar vectors in a spatial authentication signature, the result is that the vectors cancel each other. As an example, in Figure 5.8 Spicule's difference operator is applied to determine if the two instances of the 3D data spicules are the same. In this case, a stored spicule signature is subtracted from a Spicule representing the signature of the "modified" spatial dataset. The result, if the two forms are the same is referred to as a change form, a relatively featureless 3D ball. The number of objects left on the change form indicates the degree of mismatch (altered spatial data), where a totally featureless ball indicates that the two sets of data are exactly the same. A perfect match indicates that the signatures of the received spicule spatial data and what it is being compared to are the same, thus

the GIS data is determined to not have been modified and is authenticated.

In Figure 5, the modified spatial signature is detected by the change form having a vector (right image) that should not be present. The advantage of this approach is that it is fast, mathematically rigorous and is visually intuitive (user friendly). Additionally, the degree and amount of alteration that the spatial data may have is indicated through the quantity and presence or absence of vectors on the change form.

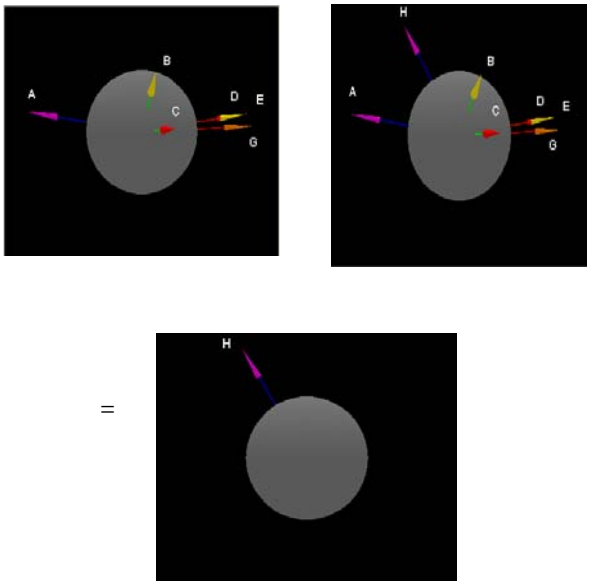

Figure 5. Visualization of authentication signature: Template form (left) – Authentication form(middle) = Change form indication H has been added to the transmitted data set generating the Authentication form

In the above example an object has been added to the Spicule data that was transmitted, which makes it differ from the authentication signature. Consequently when the signature of the transmitted data is differenced "-" with the transmitted signature and then visualized, the Change form has a vector pointing to the added object. The user can easily see that a single object has been added to the transmitted data and remove it.

5.0 CONCLUSION AND FUTURE WORK

The taxonomy developed as part of previous research classifies spatial data objects into their basic categories. This taxonomy used in the process of selection of spatial objects for generation of authentication signatures. In addition, the taxonomy may have broad application for future research that aims at standardizing spatial object's names across the different geographical information systems. This could allow spatial data to be more widely shared.

The spatial data authentication signature process developed in this research has the following benefits:

- extremely fast due to its mathematical model compared to current authentication methods
- lends itself easily to visualization for intuitive and quick change detection in data
- points to modified data using Spicule vectors which allow users to request retransmission of a single object, or ignore modified objects if they are unimportant. Current methods of encryption and authentication can not support identification of what was modified and thus require full retransmission of all data.
- can be utilized to authenticate any type of spatially organized information eg. bioinformatic, images, audio, etc utilizing the same methods presented in this paper.

Additionally, the use of the signature matrix, is very small in size compared to that of a complete spatial data set. While our performance measures looked at full authentication of all objects in the dataset, future work will refine the concept of using the taxonomy to send selected smaller sets of information. Future work will consider but is not limited to several areas :

- how are objects selected and what is the role of the taxonomy in that selection. This may be accomplished through the use of Hilbert Curves, Z curves, overlays of grids and selection from cells.
- what other types of visual algebraic operations may be useful in Spicule based authentication and what types of visual information may be added to the Spicule to assist in that process
- how does the storage size of a Spicule authentication signature compare to that of standard cryptographic methods.
- what is the relationship of Spicules origin to spatial data in the role of generation of change indication vectors. For example the further away from a plane of spatial data is from the origin should affect how much of a change vector is generated
- finally how can this approach be applied to authentication for other types of spatial organized data

Much work remains to be done based on this initial study, however the benefits could be dramatic for the field of information integrity and transmission.

REFERENCES

[1] ESRI Data & Maps, Media Kit. 2002. Esri ArcGIS. www.esri.com

[2] "An Intorduction to Geographical Information Systems", by Ian Heywood, Sarah Cornelius, and Steve Carver. Second Edition, 2002. Prentice Hall.

[3] Environmental Modeling Systems, Inc. WMS 7.1 Overview. http://www.ems-i.com/WMS/WMS_Overview/wms_overview.html

[4] William Stallings. 2003. Network Security Essentials, Applications and Standards. Prentice Hall.

[5] Charlie Kaufman, Radia Perlman, Mike Speciner. 2002. Network Security, Private Communication in a Public World. Prentice Hall PTR.

[6] Vert, G. Yuan, B. Cole, N. A Visual Algebra for Detecting Port Attacks on Computer Systems, Proceedings of the Intl. Conf. on Computer Applications in Industry and Engineering (CAINE-2003), November 2003, Las Vegas, NV, pp 131-135.

[7] Alexandria Digital Library Feature Type Thesaurus. University of California, Santa Barbara. Version of July 3, 2002. http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/index.htm

[8] Introduction to ArcView 3.x. ESRI Virtual Campus, GIS Education and Training on the Web. http://campus.esri.com/

[9] Takeyama, M., and Couclelis, H., 1997, Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of geographical Information Science* 11: 73-91.

[10] Jensen, C.S., and R. Snodgrass. 1994. Temporal Specialization and Generalization. *IEEE Transactions on Knowledge and Data Engineering* 6(6): 954-974.

[11] Onsrud, H.J., and G. Rushton. 1995. Sharing Geographic Information. Center For Urban Policy Research, New Brunswick, N.J. 510pp.

[12] Guimaraes, G., V.S. Lobo, and F. Moura-Pires. 2003.A Taxonomy of Self-Organizing Maps for Temporal Sequence Processing. *Intelligent data Analysis* 4:269-290.

[13] Heaton, Jill. Class lecture. University of Nevada, Reno. 08/23/2004.

[14] Calkins, H. W.; Obermeyer, N. J.; Taxonomy for Surveying the Use and Value of Geographical Information. *International Journal of Geographic Information Systems* V. 5, N. 3, July-September 1991, pp. 341-351.

[15] Phinn, Stuart R., Menges C., Hill, G. J. E., Stanford, M. 2000. Optimizing Remotely Sensed Solutions for Monitoring, Modeling, and Managing Coastal Environments. *Remote Sensing of Environment* 73: 117-132.