**Autonomous and Scalable Computation of Spatial Correlation Function from Spatiotemporal Materials Data**

Tony Fast

Surya R. Kalidindi

**Introduction**

Material science is experiencing pressure to deliver new and improvement materials to the marketplace in half the time and at half the cost.[ref] To reach this goal, a hurdle that will need to be overcome is the increasing dissonance between the strength of material science in generating massive stores of empirical and simulated physical information and its weakness to utilize the data at scale. Materials science is challenged by the 3 V’s of Big Data as the Volumes of the unstructured datasets are reaching terabyte sizes, as high performance computing is increasing the Velocity which information is generated, and with the Variety of information being generated by conduits in the field from both experiment and simulation. The current subjective, undocumented, and ad-hoc data analysis tools cannot be scaled to suit these needs. New data science techniques will need to be explored to offer objective statistical representations of materials science information.[ref]

Materials structure

Microstructure Informatics is an emerging framework that provides a growing suite of data science tools to extract bi-directional structure-property/processing relationships for most classes of materials science information. Microstructure Informatics uses digital signal processing and advanced statistics to properly encode material information into useful forms for machine learning, data structures, and algorithms that extract added value from the information. In microstructure informatics, the material structure, or microstructure, is considered to be the independent of the dependent property or processing response. Spatial statistics (i.e. Pair Correlations, N-Point statistics) are commonly used statistical utilities as it provides an objective statistical description of the material structure. The following successful case studies have been illustrated to shown the effective of spatial statistics in (1) determining objective microstructure comparisons in heat treated α-β Titanium [ref], (2) building regression models for the homogenized structure-property connections between the internal structure of fuel cell materials and their diffusivity [ref], and (3) determining the variance of properties associated with individual microstructures [ref].

This paper discusses fast algorithms to encode material structure information using parameterized basis function and spatial correlation functions. The concept of the microstructure function is employed to parameterize the material information and produce a digital microstructure signal; encoding can be performed on most classes of experimental and simulated material structure information.[ref] The digital signals are convolved using embarrassingly parallel Fast Fourier methods to compute the spatial statistics of the digitized microstructure. The following properties of the spatial correlation functions make them a worthy candidate for an objective material structure descriptor: several widely used statistical metrics are embedded in the correlations such as volume fraction [ref] and specific surface area [ref]; for binary images, they contain information about the original material structure information within a translation [ref]; and they can describe most types of materials science information in raw signal is processed appropriately. N case studies will be presented to illustrate the generality of the technique when applied to things.

This paper will begin with a discussion on classifications of materials science information and their proposed conversions to a digital signal using the microstructure function. Once the raw material structure information is digitized, they will form the foundations that allow spatial statistics to be computed using fast, scalable FFT algorithms. N case studies will be shown to express the diversity and general applicability of the spatial correlation functions in structure-structure comparison.

**1. A Generic Framework to Quantify Materials Science Information**

*1.1 The Microstructure Function*

The microstructure function expresses spatially resolved material structure information from physics-based models as digital signals. A physics-based model extracts material structure-response behaviors with either simulation or experimental techniques; materials information generated by either technique, or source, are uniquely identified by a set of boundary conditions and set of modeling conditions (i.e. control parameters). A model includes spatially resolved information about the material structure, , as either an input or output with local material states, , at disparate positions, , in the sample volume, . The sample volume has dimensions where is the dimensionality of the dataset; the dimensionality corresponds to the number of independent axes in the spatial domain and will have the property, . The local material state is an ordered set of salient material features such as phase, classification, grain orientation, volume fraction, spin, curvature, etc. The generality of the local material state definition enables a framework to describe most material information from most sources. The microstructure function digitizes the raw model information by the following equation

where is a digitized coefficients of the raw material information corresponding to any normalized basis functions and applied to represent the spatial domain and local state of the material, respectively. The spatial domain and local state space are orthogonal to one another and will require different basis functions. Each basis function is normalized requiring that is bounded between zero and one. The selection of the basis functions is an extremely important matter and will be discussed extensively throughout this paper. From this point on, will be referred to as the microstructure function where is a local state index corresponding to the thfunction in the local material state basis; similarly, is an index pertaining to a volume contained within the sample volume prescribed by the spatial basis function.

*1.2 Model Sampling Patterns in the Spatial Domain*

A physics-based model will sample material information at a position ; sampling does not occur at infinitesimal points in space; rather, the position corresponds to a sampling within the precision of the simulation or the probe volume of an experiment. Material information is sampled from a finite volume corresponding the resolution of the model and the resulting raw information is a probability density function of the local material states in that volume; concurrently, uncertainty is propagated by the precision of the local material states output by the model.[ref]

Two classifications of sampling patterns exist in the spatial domain: gridded and non-gridded.(Figure qq) Gridded sampling corresponds to information that is evenly spaced within the sample volume. Voxel, or 3-D pixel, based information is a gridded dataset because of this criteria; it is assumed that the voxel is the probe volume of the model. Non-gridded data is extracted from a sample volume of finite dimensions wherein can take any value within the range. Models such as Atom Probe Microscopy [ref] and Molecular Dynamics [ref] will often exhibit non-gridded, or point cloud, sampling characteristics. The probe volume associated with point cloud datasets will relate to the precision of the model from which the information is generated.

In prior work and this paper, indicator functions provide the basis to partition the spatial domain into non-overlapping, evenly spaced, cuboidal volumes [ref]; an investigation of other basis functions are currently underway (e.g. wavelets). Equation xx defines the basis for the spatial domain as . is an index to a unique cuboidal volume in the spatial domain with the properties

and

Each underlying volume is identified spatially by its centroid and has a volume of where is the number of spatial dimensions of the sample volume. [ref]

The work presented in this paper is limited to microstructure functions that are expressed with an evenly gridded spatial basis function. This transformation is trivial for material information that is generated on an even grid (e.g. microscopy images). Point cloud information can be transformed to uniform grid using the microstructure function by the appropriate basis function; it is acknowledged that some uncertainty will be propagated in this transformation. Alternate techniques are being developed to treat point cloud data on a non-uniform grid. A note: tree data structures may be most efficient to partition the spatial domain for point cloud datasets especially when the dataset is of a high-dimension.[ref] In applied problems, the microstructure function is a means to coarsen oversampled data or very large datasets; this will reduce computational demands incurred later.

*1.3 Local States and Local States Spaces of Material Dependent Information*

Independent of the spatial domain is the *n*-tuple of material features at each position ; a tuple is an ordered sequence of values. Each tuple in the sequence describes a local material state with values constrained are by a bounded space called the local state space; the local state space is the complete set of possible local states for a material feature. It is possible for one or many local material states to be extracted from the same model; several models can be combined to form the ordered set of local material states for the same sample volume. The microstructure function intrinsically accommodates multiple local state spaces. This can be demonstrated by expanding the Eq. xx as follows

where is that basis representation of the th tuple of the local material state . The local state index is uniquely mapped to . It is expected that each local state space expressed by the microstructure function will require a different basis function. For example, if the local material state is expressed as a 3-tuple,, that corresponds to the discrete phase indicator of the material , the volume fraction of an element within a phase , and the grain orientation , then a different basis function will be needed for each element.

Literature reports an extensive variety of basis functions.[ref] Fortunately, characteristics of the local state space inform the choice of basis functions for the th mode of the local state, . The following list describes all of the possible characteristics of the local state spaces:

* Discrete LSS – The local state is identified by a discrete index that demarcates a particular class of material features. For example in steel, a discrete basis will uniquely index different phases of steel (e.g. martensite, austenite, pearlite). Indicator basis functions can suite this application.
* Bounded, Periodic LSS – The local states on opposite boundaries exhibit similar material behavior(this sentence stinks). A local state space that identifies the angle of a material feature over is periodic. The cosine transform provides a basis for this type of local state space. Some new work will show an application of generalized spherical harmonics to grain orientation.
* Bounded, Non-periodic LSS – I need help with this description. Legendre polynomials are an example of a basis function for this application. Volume fraction is the example here.
* Semi-infinite LSS - I need help with this description.
* Infinite LSS - I need help with this description.

Generally speaking, some local state spaces may intrinsically be unbounded spaces, however prior information about of the material system being interrogated may place specific bounds on the local state space. The choice of the basis function is reflected in the fidelity and compaction of the microstructure function and in the computational demands to extract the statistical quantities explored in this paper.

*1.4 Partial and Uncertain Information*

In many experiments and some simulations, the material information extracted by a model may be uncertain or partial. A partial dataset will contain empty data points. An uncertain dataset will contain statistical information corresponding to the uncertainty of each data point. A dataset can be both partial and uncertain. These types of datasets may arise from poor boundary conditions or numerical instabilities in the simulation. In experiments, the resolution of the detector may impose epistemic uncertainty in the local material information. For example, in Electron Back Scattered Detection (EBSD) the confidence index at each position that sampled is provided as an output from the model. The grain boundaries in the material are often recorded as low confidence parameters. Figure yy(a) shows an EBSD scan of 7xxx series Aluminum and Figure yy(b) is a map of the confidence indexes associated with each position. It can be seen in this figure that there are regions of both high and low certainty.

A model may provide the confidence of each data point extracted within the volume. Using the same basis function that partitions the spatial domain, the digitized weight of uncertainty for each partition is derived from

where is a weight associated with the confidence of the th mode of the local material state sampled described by the weight basis at provided by the model. is the spatially resolved weighting signal. The confidence of the sampled information is incorporated into the microstructure function as

The weights are bounded between zero and one. Partial data points will assume a weight of zero and will be completely ignored from the analysis of the sample volume. Meanwhile, a weight of one indicates complete information and assumes total confidence in the local state information that is sampled. A dataset is partial if such that . Any weights between zero and one indicate uncertain data points.(Figure yy(b))

**3. Spatial Statistics**

Spatial statistics employ the microstructure function to rapidly compute an objective description of the material information provided by model(s) within a similar sample volume. The spatial statistics are computed by the following relationship

where is the probability of finding local states and separated by a vector ; is a local state derived from signal at the tail of and is local state derived from signal is at the head. The complete set of statistics includes the discrete set of statistics for all possible vectors for the sampling pattern of the model. To better understand the definition above, it is useful to consider the numerator and denominator individually. The numerator is a cumulative sum of the positive outcomes where and were observed to be separated by . The denominator provides the total number of trials conducted with a vector from the signal sources corresponding to the local state indices and . (Figure to illustrate statistics)

The spatial statistics are computed for all vectors in the sample volume, , that satisfy the Nyquist criteria, for .[ref] The correlation function of all vectors for states and is defined as . There are two types of correlations that are computed

1. Auto-correlation – occurs when and is represented as .
2. Cross-correlation – occurs when .

Both correlations functions are smooth and differentiable; they are readily amenable to interpolation methods to extract correlations of arbitrary vectors . Auto-correlation functions maintain , 2-fold symmetry, about the origin of the statistics, or the vector; slightly more than half of the vectors are unique. The cross-correlation functions and are respectively anti-symmetric. (IS THIS TRUE FOR COMPLEX BASIS?) The complete set of spatial statistics is defined for all combinations of local state indices of the microstructure function in the following anti-symmetric block matrix

with the auto-correlation functions on the diagonal and the cross-correlation functions on the upper and lower triangle. Previous literature reports the independent and dependent features of the complete statistics for an isolated, but common set of raw material information.[STEVEref]

2.1 The Numerator – An FFT Algorithm to Convolve Digitized Microstructure Signals

Both the numerator and denominator of the spatial statistics decompose into discrete convolutions of digital signals. The convolution has unique properties in the Fourier space that are often leveraged in algorithms to expedite computations. Fast Fourier Transforms (FFT) reduce the computational demands required to evaluate the expression; explicitly computing the convolution has a complexity of whereas the FFT approach has complexity, is the number of spatial points in the microstructure function.[ref] FFT methods extract the correlation function in a single computation; all vectors are accounted for at once. In materials science applications to date, spatial statistics have only been computed using uniform FFT algorithms.[ref] This class of algorithms requires that the spatial domain of the microstructure function is on an evenly spaced grid in three dimensions. Future work will include an application of non-uniform FFT methods to point cloud datasets.[ref]

Another caveat of the FFT methods is that they naturally impose periodicity on the volume being evaluated. This assumption is incorrect for a large variety of physics-based models. In fact, this assumption is erroneous for all experimental models. Figure rr explores the intrinsic problem of using a periodic assumption on an artificial non-periodic dataset using a single local state index, , from the microstructure function. Figure rr(a) shows that if periodicity is assumed, a vector that goes beyond the boundary of the sample volume will enter in on the opposing side. When vectors transverse a boundary in a non-periodic sample volume then dubious counts are recorded into the numerator of the statistics; vectors with a dotted line identity these counts. Non-periodic statistics are computed using Fourier methods by padding the microstructure function with zeros to twice its length in each spatial dimension. Padding ignores vectors that transverse the boundaries because either or is zero thereby excluding the count from the sum of probabilities in the numerator. Pseudocode to compute the numerator is shown in Table YY. Periodic boundary conditions are discussed as a special case later in this paper; mixed periodic/non-periodic boundary conditions (i.e. slab boundary conditions) can exist when modeling surfaces/interfaces.[ref]

|  |  |
| --- | --- |
| Pseudocode to compute the numerator in spatial statistics | |
| 1.  2.  3.  4.  5.  6.  7.  8.  9.  10.  11.  12.  13.  14. | Initialize raw spatial information,  Initialize the spatial Basis Function,  For number of local material states,  Initialize Local State Space basis,  Generate Microstructure Function,  For local state index  For local state index  Pad Microstructure Function signal in each non-periodic spatial dimension with zeros  Pad Microstructure Function signal in each non-periodic spatial dimension with zeros  Fast Fourier Transform padded signal  Fast Fourier Transform padded signal  Element-wise multiplication of the Fast Fourier Transforms of and  Inverse Fourier Transform element-wise multiplication  Return Cumulative Sum of True Events in the Numerator |

Table YY

2.2 The Denominator – A Generalized Approach to Normalize Spatial Statistics

To cast the correlations as a probability, the cumulative sum of probabilities in the numerator must be normalized. is the number of samples taken by a vector indexed by for local states in the numerator and derived from signals and , respectively. The simplest normalization criteria occurs when complete material information is provided, or a weight, , such that . Traditionally, this normalization (for complete non-periodic information) is computed using the following relationship

where and are the size of the sample volume and length of the vector in the th dimension, respectively. This relationship is only useful for complete material information. This relationship provides a count of vectors of length that are sampled when computing non-periodic statistics. For incomplete and partial datasets, a more generalized approach is necessary for the normalization.

In this paper, we restrict the normalization to partial datasets; a more complete definition for uncertain datasets will be presented in future work. Partial datasets are defined by a dataset with weights, ; a complete dataset is a partial dataset. The normalization for each vector is expressed as the convolution

where corresponds to the completeness of the th tuple in the raw material information. FFT methods, again, enable a rapid computation of the convolution in the normalization of any partial dataset in a data-driven manner. This approach to computing the denominator will support arbitrary sample geometries and sampling patterns. For example, the reconstruction of micro-CT information often has a cylindrical shape as due to the sampling pattern in the experimental. The prescribed normalization allows all of the available information to in the reconstructed image to be sampled for the spatial statistics without having to ignore information to suit the requirement of a complete cuboidal dataset previously required. Eq. uu is a generalized expression for the denominator in the statistics; it will exactly reproduce the results the normalization in Eq. ww. The normalization terms only need to be computed once for each local material state, , while the numerator needs to be computed for every local state index. A pseudocode to compute the denominator of the spatial statistics is provided in Table zz.

|  |  |
| --- | --- |
| Pseudocode to compute the denominator in spatial statistics | |
| 1.  2.  3.  4.  5.  6.  7.  8.  9.  10.  11.  12. | Initialize spatial basis function from the numerator computation,  For number of local material states,  Initialize the weighting Basis Function,  For weighting signal corresponding to local state ,  For weighting signal corresponding to local state ,  Pad weighting signal for in each non-periodic spatial dimension with zeros  Pad weighting signal for in each non-periodic spatial dimension with zeros  Fast Fourier Transform padded weighting signal  Fast Fourier Transform padded weighting signal  Element-wise multiplication of the Fast Fourier Transforms of and signals  Inverse Fourier Transform element-wise multiplication  Return Number of Vectors sampled in the denominator |

*2.3 Periodic Boundary Conditions as a Special Case*

Periodic boundary conditions (PBC) are often imposed in materials simulations using methods like the minimum image convention.[ref] PBC explicitly require that information passing out of a boundary during the simulation is reflected into the opposing boundary (Figure rr(a)). Under PBC, non-periodic statistics will yield valuable insight into the material structure statistics. However, since the PBC models are translation-invariant; their non-periodic spatial statistics are non-unique and under-sampled since sampling across the sample boundaries is deliberately suppressed (Figure rr(b)). Non-periodic statistics will provide one of many possible statistical descriptors. It is possible to allow the spatial statistics to sample information across the boundaries to extract a complete set of unique spatial statistics for the periodic simulation.

Extracting periodic statistics is a less computationally intensive task than the non-periodic boundary conditions. In lines 8-9 in the numerator psuedocode (Table yy) and lines 6-7 (Table zz) in the denominator, a decision is made to pad the microstructure function with zeros. This decision is based on the periodicity; if no periodicity is enforced by the physics based model then each spatial dimension is padded with zeros to avoid any dubious samples in the spatial statistics. If a simulation is perfectly periodic then the microstructure signal is not padded. In a perfectly periodic simulation with complete material information, the normalization in the denominator simply becomes and need not be rigorously computed as suggestion in the pseudocode for the denominator; if the denominator is computed then . If the physics based model has any non-periodic dimensions or is partial, then Eq. ll will need to be used to compute the denominator. During this computation, it is important to make sure that the non-periodic dimensions are the only dimensions that are padded in lines 6-7.

STOP HERE

PSUEDOCODE FOR THE STATISTICS

1. Initialize spatially resolved raw model information, , from physics-based models

*Digitize Raw Model Information using the microstructure function*

2. Initialize the Spatial Basis function

3. For the number of local material states,

Initialize the local state space for the th state

Output the microstructure function of using

4.

weights

For all unique pairs of and

FFT

Numerator Counts Complete Spatial Statistics

Denominator normalize, Convolve the signals

Numerator

For all combinations of and

Augment the microstructure with zero padding the non-periodic dimensions with zeros

FFT the augment microstructure function for states and

Denominator

Normalize

Initialize raw spatial information,

Initialize the spatial Basis Function,

For number of local material states,

Initialize Local State Space basis,

Generate Microstructure Function,

For local state index

For local state index

Pad Microstructure Function signal in each spatial dimension\* with zeros

Pad Microstructure Function signal in each spatial dimension\* with zeros

Fast Fourier Transform padded signal

Fast Fourier Transform padded signal

Element-wise multiplication of the Fast Fourier Transforms of and

Inverse Fourier Transform element-wise multiplication

PERIODIC SIMULATION BOUNDARY CONDITIONS

Periodic boundary conditions can be treated as a special cause because the material is spatially invariant in the sample volume. The prescribed approach for non-periodic material information can be applied to periodic material information. Computing the statistics in this manner will correspond to a sub-sampling of the complete statistics because vectors are not allowed to transverse the boundaries of the sample space. Since FFT methods are periodic by definition, the padding step presented earlier is ignored for both the numerator and denominator of the statistics. If all dimensions of the dataset are periodic. For interfacial simulations, one may incur partial periodic boundary conditions wherein parallel to the interface the conditions are periodic, but perpendicular they aren’t.

For a periodic dataset, the padding step is removed from the computation of both the numerator and denominator. T

For voxel based raw material information, the transition from non-periodic statistics is trivial. For point cloud data it is necessary to know the exact dimensions of the boundaries in the simulation others the vector lengths will be perturbed across boundaries.

For point cloud, it is required that the bounds of the simulation are known.

Computing periodic statistics is less computationally expensive

COMPUTING SPATIAL STATISTICS

The spatial statistics of the microstructures is an emerging quantification tool for materials science datasets. They are an objective and rigorous quantification of the internal structure of the materials.

Previous examples

Pseudocode for the examples

Figure

THE MICROSTRUCTURE FUNCTION

In recent work [1-2], a framework has been presented for the stochastic description of microstructure. In this description, the microstructure function is denoted as , and represents the probability density associated with finding local state at spatial location . The local state is often described by a set of parameters that may be organized as a k-dimensional vector, . For example, in an annealed multi-phase metallic alloy, the complete description of the local state may include the thermodynamic phase identifier, the elemental composition of the phase, and the crystal lattice orientation, among other descriptors. In the notation presented above, the space of all possible local states would be a *k*-dimensional space called the local state space, and is denoted .

A discretized version of the microstructure function can be extracted by binning (typically a uniform tessellation using an invariant measure) the spatial domain as well as the local state space [1]. Let *s =* 1, 2, …, *S* and *h =* 1, 2, …, *H* enumerate the individual bins in spatial domain and the local state space, respectively. With these conventions, the discretized microstructure function is denoted as and represents the total volume fraction of all local states from bin *h* in the spatial bin *s*. As per this definition, the microstructure function is subject to the following constraints:

|  |  |
| --- | --- |
|  | () |

Note that the framework for the description of the microstructure presented above is fairly general and can accommodate any combination of material features by simply increasing the dimensionality of the local state space; it is also not limited to any specific length or time scales. It is noted that there is an inherent approximation in the discretization process. The fidelity of this process depends on the details of the discretization (e.g. higher values of *S* and *H* improve the accuracy). However, for large spatial domains and local state spaces, the storage demands on the discretized microstructure function can increase rapidly. Therefore, one has to be extremely judicious in selecting the list of the parameters of interest in identifying the local state, especially when seeking a viable, but sufficiently accurate, description of the microstructure function.

The discretized representation of microstructure offers many advantages in fast computation of microstructure measures/metrics [3-4], automated identification of salient microstructure features in large datasets [5], extraction of representative volume elements from an ensemble of datasets [6], reconstructions of microstructures from measured statistics [7-8], building of real-time searchable microstructure databases [9], and mining of high fidelity multi-scale structure-performance-structure evolution linkages from physics-based models [10-14]. In most of the prior work, the material system was treated as a multi-phase composite with a small number of distinct phases. In other words, the local state space was inherently discrete and one-dimensional. However, in recent work dealing with extraction of reliable structure-performance linkages in high-contrast composite systems, we have introduced the notion of a higher-order microstructure function that captured systematically the details of the local conformations (i.e. the local topology). For example, if one were to consider a two-phase composite material microstructure in a uniform cuboidal tessellation of the spatial domain and define the local conformation based on the local states present in each spatial cell as well as the six adjoining nearest neighboring cells, there would be a total of 27 (= 128) distinct local conformations [14]. If these local conformations are then used to define the local state, then the size of the local state space in this specific example would increase from *H* = 2 in the simple description of the microstructure function to *H* = 128 in the seventh-order description of the microstructure function. Such higher-order microstructure functions and have been found to be exceptionally useful in establishing high fidelity structure-performance linkages in high-contrast composite material systems [14].

The main difficulty with the higher-order microstructure functions described above are that the size of the local state space increases rapidly with either the addition of more local states (for example, if 3 local states are allowed in each cell, the value of *H* would increase from 3 to 37 (= 2187) when considering only the first six neighbors in a 3-D microstructure) or the addition of more neighbors (for example, addition of second nearest neighbors in a two-phase composite would increase the value of *H* to 215 (= 32768)). The second difficulty of this prior approach is that one is forced to take on large discrete jumps in the size of the local state description (i.e. from 2 to 27 to 215 in the examples described above).

BOUNDARY CONDITIONS: PERIODICITY VS WELL NOT

The computation of the spatial statistics requires prior information about the source and the veracity of the data.

PERIODIC VS. NON-PERIODIC BOUNDARY CONDITIONS

Strictly speaking of the variety of the data, most materials science dataset will maintain non-periodic boundary conditions

INCOMPLETE OR PARTIAL DATASETS

Microstructure/Materials Informatics

Heterogeneous Data types in Materials Science Data

Fast Algorithms for Comprehensive/Hierarchical Spatial Statistics

# Versatile and Scalable Computation of Spatial Correlations from Spatiotemporal Materials Data (autonomous)

* Introduction
* Background (surya )
  + 2-pt statistics ( Make the case that this is an objective measure of microstructure)
    - Description/Illustration
    - Utility and Previous non MatSci uses of spatial statistics
      * Gokhale, Garmestani, Adams, Schuh
    - Successful matsci applications
      * Pseudocode
      * largely applied discrete local states
      * Reconstructions, Variance, Taxonomy
* Spatial Statistics
  + The microstructure Functions
    - Spatial Domain
    - Local States
    - Generalized representation of the concept of the microstructure function
  + The Spatial Domain and Experimental Conditions
    - Gridded vs. Nongridded.
    - Complete/Partial Datsets
      * Confidence Indices
  + Local states and spaces
    - Discrete vs. Continuous
      * Basis Functions to accommodate each
    - Multimodal States
  + A Fourier approach to spatial statistics
    - Counts
    - Normalization
    - Pseudocode
  + Periodicity the special case
* Case Studies: Applications to Structure-Structure Comparison
  + Principal Component Analysis
  + Karl Jacobs
  + ISU
    - Discrete vs. Legendre vs. Higher-Order
* Results & Discussion
  + Scalability to Higher-Order
  + Alternate Basis Representations
* Conclusions

DATA ON GROUP WEBSITE

What are statistics

Computing spatial statistics requires input from the metadata used in the physics-based model to generate the information that is analyzed. The key features are (1) the boundary conditions of the simulation, (2) the spatial sampling of the information, and (3) the local state information. The following section expands upon these datasets as follows.

s

DO YOU BREAK IT UP INTO THE NUMERATOR AND DENOMINATOR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Numerater | Periodicity | T | F |  |
|  | Datatype | Point | Gridded |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Denominator | Partial | T | F |  |
|  | Periodicity |  |  |  |
|  |  |  |  |  |

MASKING OF DATASETS

Spatial statistics provide a straight-forward way to manage uncertainty in the data. Such a technique allows value to be extracted from impartial datasets with confidence.

Much like the computation of the spatial statistics

sLOCAL STATES AND THE MICROSTRUCTURE FUNCTION

FAST FOURIER METHODS FOR COMPUTATION

The algorithm for computing spatial correlation functions below is only amenable to microstructure functions whose spatial domain is on an even grid. Point cloud datasets require intermediate pre-processing steps and one such case will be outlined in the example.

NORMALIZATION

PERIODICITY: THE SPECIAL CASE

All experiments and specific simulation tools will be generated with non-periodic boundary conditions. Non-periodic boundary conditions indicate that material moving out of one boundary will not return in the opposing boundary. (Figure xx) Regularly, materials simulation will impose periodic boundary conditions. In such cases it is possible to extract richer statistical information by a slight modification the pseudo-code.

The algorithm presented will provide one of many possible statistical representations of the microstructure. In Figure xx, a periodic microstructure is illustrated using the periodic boundary condition. The non-periodic statistics algorithm indicates that for periodic simulation there is a choice of the sample window to choose. Any window will provide one statistical representation of the microstructure. If windows A and B in Figure xx are used to compute the statistics then both statistics that are computed will be representative of the microstructure, but they will not be equivalent. This is because certain pair correlation are excluded from the computation because the vectors in the computation require that both the head and the tail of the vector lie within the region of interest. However, in a periodic simulation the vector can transverse the boundaries which implies that these vectors to contribute to the statistics.

There is an alternate quantification of the periodic statistics by making the following modifications to the pseudocode.

F2(A,B)

FFT(A)

If is cross-correlation

FIGURES

Gridded vs. Point cloud add the basis function to this

Periodic vs. non-periodic statistics padding

Partial datasets

Masking Datasets – Ali EBSD data.