PNNL-21448

U.S. DEPARTMENT OF ENERGY
Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Annotated Bibliography for the MATADOR Project

T Janik
AM White
NH Pawley
K Myers
CS Oehmen

June 2012

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

**Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov**

**Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov**

This document was printed on recycled paper.
(8/2010)

PNNL-21448

Annotated Bibliography for the MATADOR Project

T Janik
AM White
NH Pawley[1]
K Myers[1]
CS Oehmen

June 2012

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830 (PNNL) and
DE-AC52-06NA25396 (LANL)

Pacific Northwest National Laboratory
Richland, Washington 99352

[1] Los Alamos National Laboratory

Acronyms and Abbreviations

CDMs	committee decision mechanisms
FHMM	factorial hidden Markov model
HMM	hidden Markov models
HHMM	hierarchical hidden Markov model
LHMM	layered hidden Markov Models
NIALM	nonintrusive appliance load monitoring
NIALMS	nonintrusive appliance load monitoring system
NILM	non-intrusive load monitoring
RVM	relevance vector machine
STFT	short-time Fourier transform
SVM	support vector machine
WT	wavelet transform

Contents

MATADOR Project

The MATADOR project is focused on developing methods to infer the operational mode of facilities that have the potential to be used in weapons development programs. Our central hypothesis is that by persistent, non-intrusive monitoring of such facilities, differences between various use scenarios can be reliably discovered. The impact of success in this area is that new tools and techniques for monitoring and treaty verification would make it easier to reliably discover and document weapons development activities.

This document captures the literature that will serve as a basis to approach this task. The relevant literature is divided into topical areas that relate to the various aspects of expected MATADOR project development. We have found that very little work that is directly applicable for our purposes has been published, which has motivated the development of novel methods under the project. Therefore, the manuscripts referenced in this document were selected based on their potential use as foundational blocks for the methods we anticipate developing, or so that we can understand the limitations of existing methods.

Statistical Learning Methods

Aharon M, M Elad, and A Bruckstein. 2006. "K-SVD: An algorithm for designing over complete dictionaries for sparse representation." *IEEE Transactions on Signal Processing*, 54(11): 4311-4322. Aharon et al. present a new method for dictionary learning that searches for the dictionary that gives the best representation of each data object, subject to sparsity constraints. The algorithm is a generalization of the k-means algorithm, and alternates between finding the best sparse representation of the data items according to the current dictionary and updating the dictionary based on its fit with the data. This method could be used on the MATADOR project to identify unknown features in the data related to the events or as the basis for a letterization scheme for input into BLAST.

Finn J, J Goettee, Z Toroczkai, M Anghel, and B Wood. 2003. "Estimation of entropies and dimensions by nonlinear symbolic time series analysis." *Chaos*, 13 (2): 444-456. The authors apply symbolic nonlinear time series analysis methods to analyze nonlinear data efficiently with low sensitivity to noise. In symbolic nonlinear time series analysis, a time series for a fixed delay is partitioned into a small number (called the alphabet size) of cells labeled by symbols, creating a symbolic time series. Symbolic methods involve computing the statistics of words made from the symbolic time series. Specifically, the Shannon entropy of the distribution of possible words for a range of word lengths can be computed. The approach is interesting, and we are planning to include this analysis in the MATADOR work.

Hamerly G and C Elkan. 2004. "Learning the k in k-means." In *Advances in Neural Information. Processing Systems 16*, eds. S Thrun, L Saul, and B Scholköpf, p.281. This paper presents an extension to the classic k-means clustering algorithm that uses a Gaussian goodness-of-fit test to find the optimal number of clusters. The algorithm repeatedly runs the k-means algorithm for k = 1, 2, 3, …n, and at each step tests each cluster for normality. Clusters that do not meet the normality test are split into two (via k-means with k=2). The drawback of this method is that it assumes that each data cluster is distributed according to a multivariate normal distribution, which may not hold true for MATADOR data. Regardless, this is a clustering option that could be used as the basis of a letterization scheme.

Hastie T, R Tibshirani, and J Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition. Springer, New York. (http://www.stanford.edu/~hastie/local.ftp/Springer/ESLII_print5.pdf). The book provides an introduction to various methods of statistical learning, including supervised and unsupervised methods. The supervised learning algorithms may be used to develop a classifier that could distinguish between known events in the power data. Alternatively, the unsupervised methods could be used to develop a letterization scheme to create inputs for the BLAST algorithm.

Kohonen T. 1990. "The Self-Organizing Map." In *Proceedings of the IEEE*, 78(9):1464–1479. Kohonen describes an unsupervised learning technique called a self-organizing map, in which high-dimensional observations are mapped into a one- or two-dimensional grid of points in a way that preserves the relationships between the original observations at much as possible. This method could be used as the basis of a letterization scheme, which would be an input to BLAST.

Mairal J, F Bach, and J Ponce. 2012. "Task-Driven Dictionary Learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):791-804. Mairal et al. present an efficient method for dictionary learning from a supervised formulation. This paper describes an algorithm for solving the optimization problem that will provide the dictionary, and demonstrates the approach with a few examples. This approach may be useful to this project for identifying new features in the data that are related to known events and for developing a classifier to identify events in new data.

Olshausen BA and DJ Field. 1997. "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" *Vision Research*, 37(23):3311–3325. Olshausen and Field introduce the sparse coding method of dictionary learning which uses a probabilistic framework for choosing basis vectors. Their application area is in image reconstruction with implications to human vision and recognition. This is an unsupervised dictionary learning technique that could be used for feature identification in the MATADOR data.

Pelleg D and A Moore. 2000. "X-means: Extending K-means with Efficient Estimation of the Number of Clusters." In *Proceedings of the 17th International Conference on Machine Learning*, pp. 727-734. Morgan Kaufmann, San Francisco, California. This paper presents an extension to the k-means clustering algorithm that uses the Bayes Information Criterion or the Akaike Information Criterion to determine the best value for k. Like the Hamerlay and Elkan (2003) paper, k-means clustering is performed repeatedly for increasing values of k, and at each step, each cluster is tested to see if it should instead be two clusters. This is another clustering option that could be used as the basis of a letterization scheme.

Tibshirani R, G Walther, and T Hastie. 2001. "Estimating the number of clusters in a data set via the gap statistic." *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 63(2):411–423. doi: 10.1111/1467-9868.00293. The authors propose a method (the "gap statistic") for finding the appropriate number of clusters in a data set. The method uses measured changes in cluster dispersion for k = 1, 2, 3,…, n clusters (i.e., the "gap" statistic) and lays out a statistical framework to justify how to choose the appropriate cluster size k. This can be used with many clustering algorithms, and is demonstrated with k-means and hierarchical clustering in the paper. This could be used on the

MATADOR project to identify data clusters for letterization (e.g., for input into BLAST) or for differentiating between different types of events in the data.

Tipping ME. 2001. "Sparse Bayesian Learning and the Relevance Vector Machine." 2001. *Journal of Machine Learning Research*, 1: 211-244. This paper introduces a general Bayesian framework for obtaining sparse solutions to regression and classification tasks utilizing models linear in the parameters. Although this framework is fully general, the author illustrates his approach with a particular specialization that he denotes the relevance vector machine (RVM), a model of identical functional form to the popular and state-of-the-art support vector machine (SVM).

Tosic I and P Frossard. 2011. "Dictionary Learning." *IEEE Signal Processing Magazine*, 28(2):27-38. This is an overview of the dictionary learning problem, solution methods, and applications. Dictionary learning is the process of finding a set of vectors for which any data item may be represented by a linear combination of a small (sparse) subset of this set. Dictionary learning is often used for image de-noising, dimension reduction and classification, and could be used for this project to identify unknown features in the data (unsupervised learning) or additional unknown features associated with the known events (supervised learning).

Vapnik VN. 1995. *The Nature of Statistical Learning Theory*, Springer, New York. Vapnik introduces the concept of SVMs, a supervised learning algorithm for finding an optimal linear discriminant in a transformed feature space. On the MATADOR project we have truth data in the form of event dates, times and types, so SVMs could be used to develop an event classifier which would take advantage of the full frequency spectrum. However, SVMs have the disadvantage that they do not identify previously unknown events in the data (i.e., other types of events than the ones currently under consideration.).

State-space Models

Bui H, D Phung, and S Venkatesh. 2004. "Hierarchical hidden Markov models with general state hierarchy," In *Proceedings of the Nineteenth National Conference on Artificial Intelligence*, pp. 324-329. AAAI Press. The authors generalize the form of the hierarchical hidden Markov model (HHMM) of Fine et al. (1998), by allowing a more general lattice structure instead of restricting to a tree topology.

Fine S, Y Singer, and N Tishby. 1998. "The Hierarchical Hidden Markov model: Analysis and Applications." *Machine Learning*, 32(1): 41-62. Introduces a tree structure for hidden Markov models (HMMs) called a hierarchical hidden Markov model (HHMM) to capture hierarchies among processes. This is a generalization of a segmental HMM (e.g., Russell 1993) which allows each segment to contain subsegments, each modeled by an HMM. In this framework, a tree HHMM can capture a complex process, while subtrees can capture more primitive component processes. The MATADOR effort proposes to capture hierarchical structure as well, inspired by this methodology.

Murphy KP and MA Paskin. 2001. "Linear time inference in Hierarchical HMMs." In *Advances in Neural Information Processing Systems 14*, eds. T. G. Dietterich, S. Becker, and Z. Ghahramani. MIT Press, Cambridge, Massachusetts. Demonstrates how to make HHMMs practical computationally by reducing the cost from $O(T^3)$, where T is the length of the sequence, to $O(T)$.

Myers, K. 2011 “Malt balls or malt beer? Detecting the prohibited operation of dual-use facilities.” Poster session presented at Experiments for Processes with Time or Space Dynamics, Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, July 18-22, 2011. This poster presents our template approach for using hidden Markov models to distinguish known behavior patterns in the presence of severe noise and mismatches between the measurements and the underlying operations. This template approach is related to the LHMMs presented in Oliver et al. (2004) and to the segmental HMMs presented in Russell (1993).

Oliver N, A Garg, and E Horvitz. 2004. “Layered representations for learning and inferring office activity from multiple sensory channels.” *Computer Vision and Image Understanding*, 96(2):163-180. Presents the use of a cascade of hidden Markov models, which the authors call layered hidden Markov models (LHMMs), to combine multiple streams of information and identify activities. Their LHMMs can be considered a generalization of our template-based HMMs, and we see an opportunity to use their methods to make our approach more flexible. In addition, they provide an excellent explanation of why standard HMMs are too brittle for the problems we are considering.

Rabiner LR. 1990. “A Tutorial on Hidden Markov-Models and Selected Applications in Speech Recognition.” In *Proceedings of the IEEE*, 77(2):257-286. A classic paper on hidden Markov models (HMMs). Our approach to decoding sequences of activity templates relies on HMMs.

Russell M. 1993. “A segmental HMM for speech pattern modeling.” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2(2):499-502. Introduces a simple segmental HMM that allows the underlying process to be semi-Markovian rather than Markovian. Duration is modeled explicitly. The methods used in MATADOR also require the flexibility of a semi-Markov model and are a simplification of this framework.

Biology-inspired Engineering Techniques

Altschul SF, W Gish, W Miller, EW Myers, and DJ Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology*, 215(3):403-10. This paper describes a text alignment method called BLAST that is the foundation of most of modern bioinformatics. The method is used to locate statistically significant subsequences in a pair of text strings given a scoring matrix that penalizes or rewards alignment of text based on the characters being aligned. The method allows for insertions and deletions of text, and masks regions of a query string having low complexity. This is a foundational method that we intend to apply in the MATADOR project to discover when sequences are derived from a common ancestor.

Eddy SR, G Mitchison, and R Durbin. 1995. “Maximum discrimination hidden Markov models of sequence consensus.” *Journal of Computational Biology*, 2(1):9–23. This paper describes a computational method to construct a model of the consensus region in a collection of text strings representing sequences that are in a single family. The Markov model is at the heart of this method. The consensus model captures the relationships in subsequences that are present in the members of the sequence family. We plan to use this method in the MATADOR project to express complex sequence families where a simple linear sequence is inadequate.

Eddy SR. 2004. “Where did the BLOSUM62 alignment score matrix come from?” *Nature Biotechnology*, 22(8):1035-6. This paper describes the generalization of the scoring matrix, which is an essential input for performing text alignment. An essential task of the MATADOR project will be to discover the space of characters that cover the sequence primitives in the data. This generalized scoring matrix takes into account the underlying distribution of characters in the dataset and a presumed background rate of mismatch. We plan to use this technique to construct a domain-specific scoring matrix that will enable BLAST calculations on sequences of events from facility monitoring data.

Oehmen CS and J Nieplocha. 2006. "ScalaBLAST: "A scalable implementation of BLAST for high-performance data-intensive bioinformatics analysis." *IEEE Transactions on Parallel and Distributed Systems*, 17(8):740-749. This paper describes the parallelization of the BLAST method for use in large-scale datasets using multiprocessor architectures. If the sequence datasets in the MATADOR project exceed the size where serial BLAST processing is reasonable, we will use this multiprocessor version of BLAST to rapidly complete calculations.

Oehmen CS, ES Peterson, and ST Dowson. 2010. "An Organic Model for Detecting Cyber Events." In *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research*, Article No. 66. Association for Computing Machinery, New York. doi:10.1145/1852666.1852740. This paper describes a method for generalizing BLAST-based sequence alignment for use in non-biological applications. The method includes a discovery phase in which the space of character primitives is identified and mapped to characters of text. Entities of interest are then expressed as sequences of text based on the sequence of behaviors they exhibit and the mapping of these behaviors to text characters. BLAST calculations, multiple alignments, clustering, family tree analysis, and other calculations can be applied to this text-based representation. In the MATADOR project, we plan to use a similar technique to discover the space of primitives that capture facility operations and express facility modes in terms of sequences, performing similar calculations to discover similar modes and signatures of these modes.

Nonintrusive Monitoring of Electric Loads

Berges M, E Goldman, H Matthews, and L Soibelman. 2010. "Enhancing electricity audits in residential buildings with nonintrusive load monitoring." *Journal of Industrial Ecology, 14(5): 844-858.* This article provides an overview of the practicalities of nonintrusive appliance load monitoring system (NIALMS). Since many approaches require a specific monitoring device, it is useful to understand the range of hardware available and their associated costs. The authors relate the capabilities of such hardware to the potential for nonintrusive appliance load monitoring (NIALM), including an overview of general appliance behavior and their power demands.

Cox R, R Leeb, S Shaw, and L Norford. 2006. "Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion." In *Proceedings of 21st Annual IEEE Applied Power Electronics Conference and Exposition (APEC 2006), pp. 1751-1757.* This paper demonstrates a new monitoring system for tracking individual load operation on an aggregate power service using only voltage measurements, i.e., with no current sensor.

Drenker S and A Kader. 1999. "Nonintrusive monitoring of electric loads." *IEEE Computer Applications in Power,* 12(4):47-51. Authors present an implementation and commercialization of a NIALMS for residential appliances. The algorithm consists of five phases: edge detection, cluster analysis, cluster matching, anomaly resolution, and appliance identification. The classification/cluster feature space is a simple 2D plane with axes being real and reactive power.

Hart G. 1992. "Nonintrusive appliance load monitoring." *Proceedings of the IEEE, 80(12):870-1891.* This is a pioneering paper where the author defines the field of NIALM. The paper provides insight into the aims and applications of such technology, in addition to summarizing both his own and other related work in the late 1980s and early 1990s. Hart explains the intuition behind many approaches to NIALM, including his total load model (also known as integer programming or combinatorial optimization) and appliance model (which has since been extended using HMMs).

Jiang L, J Li, S Luo, J Jin, and S West. 2011. "Literature Review on Power Disaggregation." In *Proceedings of 2011 International Conference on Modeling, Identification and Control,* pp. 38-42. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey. This paper presents the literature review on the non-intrusive load monitoring (NILM) approaches to the power disaggregation problem.

Jin Y, E Tebekaemi, M Berges, and L Soibelman. 2011. "Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities." In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2011.* The paper presents an adaptive goodness-of-fit χ^2 test event detector. Signal under consideration is the average power in time collected from three residential buildings.

Kolter JZ and MJ Johnson. 2011. "REDD : A Public Data Set for Energy Disaggregation Research." In *Workshop on Data Mining Applications in Sustainability (SIGKDD): 1-6.* This paper describes a recently published public data set for evaluating NIALM methods. The issue is that, generally, energy disaggregation approaches are evaluated on different data sets, making direct performance comparison impossible. The data collected includes energy monitors at premises-level, circuit-level, and appliance-level in six homes in the United States.

Lam HY, GSK Fung, and WK Lee. 2007. "A novel method to construct taxonomy of electrical appliances based on load signatures." *IEEE Transactions on Consumer Electronics,* 53(2):653-660. The authors use voltage-current (V-I) trajectories to characterize household appliances. They explore the use of dendrograms as a hierarchical clustering tool to organize the trajectories into families of appliances.

Norford L and S Leeb. 1996. "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms." *Energy and Buildings,* 24: 51-64. The authors focus on commercial buildings, which is not so common in the published literature. They have shown that electrical loads from space-conditioning equipment in commercial buildings can be detected centrally on the basis of appropriate filtering and changes in steady-state power. The paper can be useful since the presented approach will be compared to the MATADOR procedures.

Zeifman M and K Roth. 2011. "Nonintrusive appliance load monitoring: Review and outlook." *IEEE Transactions on Consumer Electronics,* 57 (1): 76-84. This paper presents a description and critique of all techniques applied within the NIALM field over the past 25 years. The range of approaches identified by the authors make this paper very useful when viewing the field as a whole. The authors discuss consumer systems (residential).

Classification Techniques for Load Identification in Nonintrusive Monitoring

Lin Y and M Tsai. 2010. "A novel feature extraction method for the development of nonintrusive load monitoring system based on BP-ANN." In *Proceedings of the 2010 International Symposium on Computer, Communication, Control and Automation*, pp. 215-218. The authors propose an approach that integrates an artificial intelligent recognition technique (back-propagation neural networks) and load current acquisition method. The feature computation is based on time domain information.

Lin Y and M Tsai. 2011. "Applications of Hierarchical Support Vector Machines for identifying load operation in nonintrusive load monitoring systems." In: *Proceedings of the 8th World Congress on Intelligent Control and Automation*, pp. 688-693. In this paper, an NILM system that integrates transient load current feature extraction methods and an artificial intelligent recognition technique is proposed. The proposed system that uses the hierarchical SVM as the load identifier is able to identify load operation status with single load and multiple load operation scenarios in different experimental environments. This approach is in line with the MATADOR project classification techniques.

Su Y, K Lian, and H Chang. 2011. "Feature selection of non-intrusive load monitoring system using STFT and wavelet transform." In *Proceedings of the 2011 Eighth IEEE International Conference on e-Business Engineering*, pp. 293-298. This paper employs a scheme for a non-intrusive load monitoring system by extracting the significant and representative power signatures of voltage and current at a utility service entry in identifying loads and analyzing the characteristics of loads, and then finds out the physical behavior of the operation of loads to establish the model of loads. This paper uses short-time Fourier transform (STFT) and wavelet transform (WT) of time-frequency domain data to analyze and compare different loads in the experiments. In the experiments, the results reveal that the wavelet transform is better than STFT on transient analysis of loads. We are planning to compare the MATADOR feature extraction with those presented in this paper.

Wang Z and G Zheng. 2011. "New method for non-intrusive data extraction and classification of residential appliances." In *Proceedings of 2011 Chinese Control and Decision Conference (CCDC)*, pp. 2196-2201. The paper is focused on the NILM system, which allows the identification of residential appliances in a non-intrusive way. It contains some interesting mathematical models based on probability theory but the discussed time domain signals (power in time) are very simple and belong to 3 categories only.

Yu Y, P Li, and C Zhao. 2008. "Non-intrusive method for on-line power load decomposition." In *Proceedings of 2008 China International Conference on Electricity Distribution (CICED).* A non-intrusive method for on-line power load decomposition is introduced. This method is based on the following fact, that in normal operational states of electrical appliances, steady-state currents (including fundamental current and harmonic currents) of electrical appliances normally have certain statistical regularity. The authors use a simple combination of harmonic components for each appliance. They estimate a power load current by the linear superposition of the current components of n main types of electrical appliances.

Zeifman M and K Roth. 2011. "Viterbi Algorithm with Sparse Transitions (VAST) for Nonintrusive Load Monitoring." In *Proceedings of 2011 IEEE Symposium on Computational Intelligence*

Applications in Smart Grid. This paper introduces a new disaggregation algorithm for NIALM based on a modified Viterbi algorithm. This modification takes advantage of the sparsity of transitions between appliances' states to decompose the main algorithm, thus making the algorithm complexity linearly proportional to the number of appliances.

Signal Processing Techniques for Nonintrusive Monitoring

Hanoch L and A Stankovic. 2003. "Hilbert space techniques for modeling and compensation of reactive power in energy processing systems." *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 50(4): 540-556. The paper provides a unified exposition of various definitions of reactive (or inactive) power and of various compensation methods. In particular, it presents a comparative Hilbert space analysis of definitions used in power system studies and in electric drives (the so called instantaneous reactive power). The main finding is that the notion of an orthogonal projection plays a central role that allows for a unified treatment of various compensation strategies, and also provides means for effective computations.

Kim H, M Marwah, M Arlitt, G Lyon, and J Han. 2011. "Unsupervised disaggregation of low frequency power measurements." In *The 11th SIAM International Conference on Data Mining*, pp. 747-758. This paper describes how Factorial Hidden Markov Models (FHMMs) from the machine learning domain can be used for NIALM. The authors cover the tractability of both training and inference over such models. Two extensions of the basic FHMM are described and evaluated; the use of additional data sources (e.g., time of day) and the dependencies between appliances (PC and monitor). We consider implementing this approach for unsupervised disaggregation of the power data into per-appliance usage information.

Laughman C, K Lee, R Cox, S Shaw, S Leeb, L Norford, and P Armstrong. 2003. "Power signature analysis." *IEEE Power and Energy Magazine*, 1(2):56-63. This article provides a high-level overview of the state of the art at the time of publishing. The approaches discussed include those applicable to high-granularity data, such as harmonic and transient analysis, in addition to steady-state disaggregation methods. Some intuition is given into how appliances with a continuously varying power draw can be disaggregated.

Liang J, SKK Ng, G Kendall, and JWM Cheng. 2010. "Load Signature Study—Part I: Basic Concept, Structure, and Methodology." *IEEE Transactions on Power Delivery*, 25(2):551-560. This paper depicts the basic concept, features of load signatures, structure and methodology of applying mathematical programming techniques, pattern recognition tools, and committee decision mechanism to perform load disaggregation. The authors provide a useful description of how multiple approaches can be combined to give a single, more accurate classification. An evaluation of many accuracy metrics is also provided.

Liang J, S Ng, G Kendall, and J Cheng. 2010. "Load signature Study—Part II: Disaggregation Framework, Simulation, and Applications." *IEEE Transactions on Power Delivery*, 25(2):561-569. Based on the proposed disaggregation framework, the authors use three advanced disaggregation algorithms, called committee decision mechanisms (CDMs), to perform load disaggregation at the metering level. Three random switching simulators are also developed to investigate the performance of different CDMs under a variety of scenarios. The proposed framework can be useful for the MATADOR work on disaggregation of the data.

Shaw S and C Laughman. 2007. "A Kalman-filter spectral envelope preprocessor." *IEEE Transactions on Instrumentation and Measurement*, 56 (5): 2010-2017. This paper presents a Kalman-filter approach for computing spectral envelopes of current waveforms for nonintrusive load monitoring on the electric utility. Spectral envelopes represent time-varying frequency content and phase of the current relative to the voltage. We are planning to implement some of these techniques in the MATADOR work.

Stoffer D, D Tyler, and D Wendt. 2000. "The spectral envelope and its applications." *Statistical Science* 15 (3): 224-253. The concept of the spectral envelope is discussed as a statistical basis for the frequency domain analysis and scaling of qualitative-valued time series. This technique is currently used in the nonintrusive load monitoring.

Wichakool W, A Avestruz, R Cox, and S Leeb. 2009. "Modeling and estimating current harmonics of variable electronic loads." *IEEE Transactions on Power Electronics*, 24(12):2803-2811. This paper demonstrates a method to extract the fundamental current harmonic of the uncontrolled, three-phase rectifier load using a linear combination of higher AC-side harmonics. The proposed algorithm enables NILM to work with a larger set of loads.

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)
www.pnnl.gov

U.S. DEPARTMENT OF
ENERGY