Journal of Power Sources 307 (2016) 368–378

Contents lists available at ScienceDirect

Journal of Power Sources

journal homepage: www.elsevier.com/locate/jpowsour

# Critical parameters governing energy density of Li-storage cathode materials unraveled by confirmatory factor analysis

Kee-Sun Sohn [a], Su Cheol Han [b], Woon Bae Park [a], Myoungho Pyo [b, *]

[a] Faculty of Nanotechnology and Advanced Materials Engineering, Sejong University, Seoul 143-747, South Korea
[b] Department of Printed Electronics Engineering, Sunchon National University, Chonnam 540-742, South Korea

## HIGHLIGHTS

- Layer and spinel cathode materials of LIBs are data-mined by CFA through PCA.
- Critical descriptors governing energy densities (ED) and stability are extracted.
- SEM reveals that Descriptors governing ED are different with cathode structures.
- Stability is irrelevant to the descriptors chosen in both materials.
- This work provides insight which is impossible to obtain only by human cognition.

## ARTICLE INFO

*Article history:*
Received 17 July 2015
Received in revised form
30 November 2015
Accepted 1 January 2016
Available online xxx

*Keywords:*
Confirmatory factor analysis
Data mining
Structural equation model
Cathode
Lithium ion batteries

## ABSTRACT

Despite extensive effort during the past few decades, a comprehensive understanding of the key variables governing the electrochemical properties of cathode materials in Li-ion batteries is still far from complete. To elucidate the critical parameters affecting energy density (ED) and capacity (Q) retention in layer and spinel cathodes, we data-mine the existing experimental data via confirmatory factor analysis (CFA) based on a structural equation model (SEM), which is a proven, versatile tool in understanding complex problems in the social science. The data sets are composed of 18 and 15 parameters extracted from 38 layer and 33 spinel compounds, respectively. CFA reveals the irrelevance of Q retention to all the parameters we adopt, but it also reveals the sensitive variations of ED with specific parameters. We validate the usefulness of CFA in material science and pinpointed critical parameters for high-ED cathodes, hoping to suggest a new insight in materials design.



## 1. Introduction

Since the market launch of Li-ion batteries (LIBs) by Sony Inc. in 1997, enormous effort has been devoted to the discovery of new cathode materials that can deliver high energy density (ED) without serious decay during repeated charge-discharge (C-D) cycles [1–57]. Despite the noticeable accomplishment of new Li-storage cathodes as a result of these long-term endeavors, however, the current state of LIBs remains far short of the high standards required by power-generating systems for long-distance electric vehicles, and cannot satisfy the demands of the energy storage systems for a 'so-called' smart-grid society.

The current devotion to high-ED LIBs has been carried out via individual confirmations of one-to-one relationships between material characteristics and electrochemical properties rather than by reliable and comprehensive model systems. Although such a scrutinized investigation into specific cathode materials has elicited significant scientific findings, the sensitive dependence of electrochemical properties on crystallographic arrangements and physicochemical properties of constituents has made it difficult to establish general rules for such findings. Density functional theory (DFT) calculation can compensate for these shortcomings to some extent [58,59]. However, DFT is also far from complete in explaining the overall trend of reported experimental data, which indicates the necessity of a more systematic approach that can be used to prospect the extent of experimental data for a quantum-leap innovation in materials design.

Statistics-based data-mining could be used to integrate fragmented information surrounding LIBs research society to extract a

* Corresponding author.
*E-mail address:* mho@sunchon.ac.kr (M. Pyo).

http://dx.doi.org/10.1016/j.jpowsour.2016.01.007
0378-7753/

tangible understanding of the overall propensity for a variety of cathode materials. There have been several trials using meta-analysis for functional material systems such as inhomogeneous catalysts [60], perovskite-based ferroelectric materials [61], inorganic high-temperature superconductors [62], and luminescent materials [63]. In the data-mining of the above-mentioned materials, partial least square regression (PLSR) as well as typical machine learning techniques such as artificial neural network (ANN) and support vector machine (SVM) have been utilized along with principal component analysis (PCA) for data reduction. Confirmatory factor analysis (CFA) can be contrasted with these conventional data-mining processes, in that latent variables (so-called factors) play a central role in disclosing a generalized relationship (causal effect) between variables. These variables are not experimentally measurable, but they can be indirectly measured by other experimentally measurable variables. The existence of latent variables is a 'rule-of-thumb'. CFA is often used to analyze complex, large-scale data in the social science fields [64,65]. However, it has never been applied to functional materials informatics, either in the natural science or in engineering, despite its proven usability and reliable robustness in the field of social science and bioinformatics.

Although there exist many reports on experimental or calculated cathode performance, there has been no attempt to collect them and to understand the overall trend in a systematic manner. In fact, the enormous amount of data and the erroneous nature of the data have prevented investigators from pursuing any sort of analytical process, but it motivated us to employ a statistical approach. Although the systematic raw data acquisition and the relevant descriptor extraction can be completed via tremendous effort, an ensuing data-mining process to obtain a deeper understanding has been an additional hurdle in the statistical approach. Recent advancements in data-mining techniques, however, enabled us to handle such a task with ease. CFA is a typical example of such an improvement [64,65]. Therefore, this is the right time to perform the integrated management of all data collected from many different sources in order to promote in-depth understanding of cathode performance and, thereby, to provide guidance to the discovery of novel high-ED cathode materials via reliable prediction (See Fig. 1 for a typical CFA process.).

In this work, we data-mine the cathode materials of LIBs through CFA. The layer and spinel structures are separately subjected to the statistical process. By collecting accurate structural and electrochemical data of both types, we prepare comprehensive data sets and utilize them in the CFA through PCA. The data sets we adopt for CFA are relatively large in size: 18 descriptors X 38 cathodes for layer and 15 X 33 for spinel. The CFA for layer cathodes reveals that the ED during 1st discharge at a given current density results in 4 critical descriptors with 2 latent factors, showing a causal effect between the 2 latent factors. In spinel cathodes, 3 descriptors are evaluated critical to understand the ED. The capacity (Q) retention, however, is in no way to be related to our selected descriptors, albeit the introduction of latent factors. We interpret these findings based on the pre-knowledge accepted by the LIBs society, which is virtually impossible only by human cognition capability, without a smart statistical approach such as CFA.

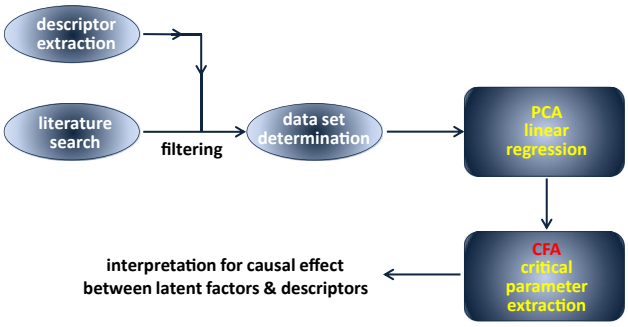


**Fig. 1.** Data mining process for the extraction of critical parameters. Prior to PCA and CFA, a reliable data set should be deliberately determined through the selection of flawless examples of literature and by most-probable descriptors.

## 2. Methods

### 2.1. Literature data filtering

Although CFA can account for the data scattering, the collection of reliable data was the most important part of the present study. From the vast quantity of literature, we collected structural and electrochemical data for layer and spinel cathode materials. The total number of literature references examined included ca. 2200 for layer cathodes and ca. 2500 for spinel cathodes. We down-sized the scope of the material by restricting it to the general formulas of $[Li_\alpha M_\beta]_{3b}[Li_\gamma M_\delta]_{3a}[O_2]_{6c}$ (layer) and $[Li_\alpha M_\beta]_{8a}[Li_\gamma M_\delta]_{16d}[O_4]_{32e}$ (spinel), where M could be multiple species, but was mostly transition metals (TM). The space groups were limited to *R-3m* and *Fd-3m* for layer and spinel cathodes, respectively. No cation-ordered structures, which, in general, have space groups of *C*2/*m* (layer) and $P4_332$ (spinel), were included. Oxygen-deficient and anion-doped compounds were also excluded. The compounds synthesized via a conventional solid-state method were the first choice to reduce the complexity incurred by morphology-dependence. If possible, papers that described all the descriptor values were preferred. When there were several literature for the similar compounds, the main criterion was the credibility of Rietveld refinements. Compounds with impurity phases and refinement results showing a greater $R_{wp}$ were discarded. For a series of compounds composed of the same elements, but with different compositions, the selection was limited to 3 compounds to avoid overestimation. A compound, which was well known to decay fast during C-D cycles, was also excluded since it could severely distort a statistical approach. For example, $LiCrO_2$ was excluded from the data set, because the inhibition of $Li^+$ transport by $Cr^{6+}$ produced via a disproportionation reaction during charge resulted in abnormal discharge behaviors [66]. In the same context, 3 V-class manganese spinels were not included. The compounds selected numbered 38 and 33 for layer and spinel cathodes, respectively. These are listed in Tables 1 and 2.

### 2.2. Descriptor extraction

We defined a number of structural, elemental, and electrochemical parameters as 'descriptors'. They were also used to account for 2 target descriptors (the ED and the Q retention of cathode materials). To express the descriptors excluding target descriptors, we used the term 'indicators'. Initially, we selected more than 32 descriptors and executed PCA with all. The preliminary outcome revealed that some subsets of these showed a perfect colinearity, leading to singular matrix problems. For example, a void volume in a unit cell was synchronously changed by the unit cell volume in the layer and therefore was removed from the data set. A solvent dielectric constant was also discarded because of negligible variations. Finally, we extracted 16 indicators (9 structural, 6 elemental, and 1 electrochemical) for the layer structures. The structural descriptors consisted of crystallographic parameters (e.g., lattice parameters and interatomic distances), and the elemental parameters included the intrinsic properties of constituent elements (e.g., atomic numbers and ionization