

# 大規模エンジニアリングデータ解析と MATLABを活用したエンジンおよび 車両の設計改善

## はじめに

自動車業界のエンジニアは長年にわたり、MATLAB®でデータを分析し、性能や効率の向上を実現してきました。自動車開発に携わるエンジニアは、エンジン、トランスミッション、シャーシなど、実質的に車両の全ての部分から収集したデータを分析するためにMATLABを活用しています。しかし、収集したデータがテラバイト単位、あるいはそれ以上になった場合はどうしたらよいのでしょうか。多数の車両からなるフリートレベルになると、車両データの規模が数テラバイトを超えることも珍しくありません。

ますますハードルが上がっているフリートデータ解析の問題に対し、MATLABをスケールアップすることで対応できるのでしょうか。また、大規模エンジニアリングデータの解析アルゴリズムを構築し、実際にデータサイエンティストや製品開発エンジニア、ビジネスアナリストが利用できるように実稼働環境に実装することは可能なのでしょうか。これら二つの質問に対する答えは、端的に言うと「YES」です。

このホワイトペーパーでは、MATLABを用いて大規模エンジニアリングデータから重要な知見を導き出し、エンジンおよび車両の設計改善に役立てる方法を説明します。具体的には、MATLABを使用して大規模エンジニアリングデータを解析するために構築した実際のシステムについて紹介します。私たちは独自データを収集するこのシステムを構築することで、MATLABによるデータアナリティクスをビッグデータに適用する方法を効率化し、改善できるようになりました。このシステムには、実際のフリートデータ収集のためのインフラ、ならびにMATLABを用いて作成・実装したウェブベースのフリートデータ解析アプリケーションが含まれています。ホワイトペーパーの最後では、このセットアップを用いて得ることができる有益な知見の例を紹介합니다。

## 大規模エンジニアリングデータとは何か？なぜ大規模エンジニアデータから知見を導き出すことが難しいのか？

各業界において、ビッグデータは一般に「3つのV」で説明されます。すなわちVolume（量）、Variety（多様性）、Velocity（速度）です。（4番目のVとして、しばしばVeracity（正確さ）が挙げられますが、Veracityについては後述します。）エンジニアリングアプリケーションでは、上記のような性質を持ったデータの特徴付けるために「大規模エンジニアリングデータ」という用語を用いますが、エンジニアリングデータの場合は、一般的なビッグデータとは微妙に異なる問題を伴います。一般に、工学技術分野のビッグデータは、他の業種のビッグデータに比べてしっかり構造化されており、種類も多く、はるかに速いスピードで増加する可能性があります。例えば、フリートデータの場合は次の通りです。

- Volume（量）とは、データの大きさを指します。自動車メーカー（OEM）の場合、一度に20TBものデータを扱うことも珍しくなく、1日1台当たり30 GBを上回るデータを収集することもあります。この速さでデータを収集していると、データサイズが簡単にペタバイト規模にまで増加してしまい、一度ではメモリ内で分析できなくなります。私たちが構築した比較的小規模な実証システムでさえ、1日1台当たり25 MBのセンサーデータが収集可能で、ほんの数人のドライバーだけで、数カ月間のうちに約2 GBのデータが生成されました。この数値には、先進運転支援システム（ADAS）でよく使用される映像データは含まれていません。映像データを含んだ場合、データの蓄積速度は1時間当たり数ギガバイトに及ぶこともあります。
- Variety（多様性）は、様々なソースからのデータを活用することで、意外な知見やより価値の高い知見が得られる可能性があるというアナリストの認識を反映しています。今日の車両は何十

ものセンサーを搭載しています。ましてや複数の計測車両を用いる場合、さらに多くのセンサーが、運転速度、燃費、温度など様々な信号を生成することになります。車両の種類によって、生成されるデータの種類も異なります。例えば、ハイブリッド車や電気自動車の場合は、燃料流量ではなく、充電情報を報告するでしょう。私たちのシステムでは、車両からのデータと、GPS機器など他のセンサーからのデータとを組み合わせ、時間的整合性を調整しました。そうしてようやく、これらの時系列データを、シミュレーションデータ、音声、映像、CANのログデータなど他の種類のデータによって補完できるようになります。

- Velocity (速度) は、ビッグデータ用語では、データが蓄積される速度を指しています。MathWorksのデモ用システムでは、1秒ごとに複数のセンサーからデータを収集しました。実際には、サンプリング周期が10ミリ秒の車載センサーであれば、1秒あたり100のデータポイントが生成されます。多数の車両の複数のセンサーからデータが流れ込めば、あっという間にギガバイト、テラバイト規模に膨れ上がります。

大規模エンジニアリングデータから価値を引き出そうとするなら、まずはそのデータの量、多様性、蓄積速度に対応できるシステムを用意する必要があります。そうしたシステムには、柔軟でスケールラブルなデータ解析アルゴリズムが不可欠であり、それが実稼働環境内に展開されて初めて、データサイエンティストや開発エンジニア、ビジネスアナリストがデータから知見を導き出すことができるのです。

## MATLABによる大規模エンジニアリングデータ解析システムの構築

MathWorksコンサルティングサービスチームは、これまでたくさんの顧客と共同でデータアナリティクス・アプリケーションを構築してきましたが、このチームが望んだのは、実験、調査、研究、共有できる大規模エンジニアリングデータの収集と分析のためのフレームワークでした。

MATLABデータアナリティクス・アプリケーションには、多くのアーキテクチャーオプションが用意されています。私たちが構築したシステム(図1)では、まずデータ収集用のアダプターを車両のODB2ポートに接続し、車両のデータをBluetoothでスマートフォンに無線送信します。スマートフォンは、受け取ったデータをAmazon EC2のLinux-Apache-MongoDB-Rails (LAMR) スタック上で動作するクラウドアプリケーションに転送します。このシステムでは、車両はIoTソリューションのエッジノードとしての役割を果たし、車両が生成したデータは収集されてデータアグリゲーターへと送信されます。システムには、JavaベースのHadoopソフトウェアフレームワークが用いられており、分散型ストレージやコンピュータクラスター間での大規模エンジニアリングデータの処理を簡素化しています。

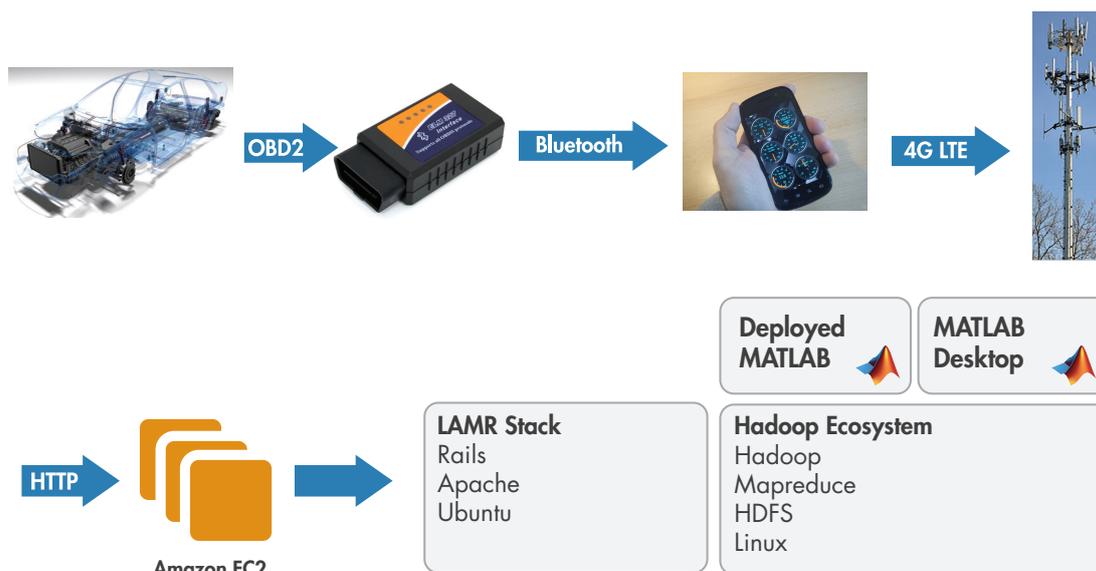


図1. フリートデータ収集、ならびにMATLABとHadoopによるデータアナリティクスのためのハードウェア・ソフトウェアインフラ

上記のようなデータ収集インフラを構築した上で、MATLABベースのワークフローを用いて、単一コンピュータ上での標準的なデータ解析手法から始め、最終的にはプロダクション版ウェブアプリケーションの実装まで実現しました。私たちはMATLABを用いてデスクトップ上で収集したデータを調べ、様々な処理方法を考案、テスト、視覚化しました。デスクトップでの検討に基づき、フリートデータ解析ウェブアプリケーションの開発・実装を行いました。このウェブアプリケーションは、Hadoopを用いたMATLABアナリティクスパッケージとウェブ技術とが組み合わされており、車両群（フリート）の性能特性の可視化、最適化、分析機能を備えています（図2）。

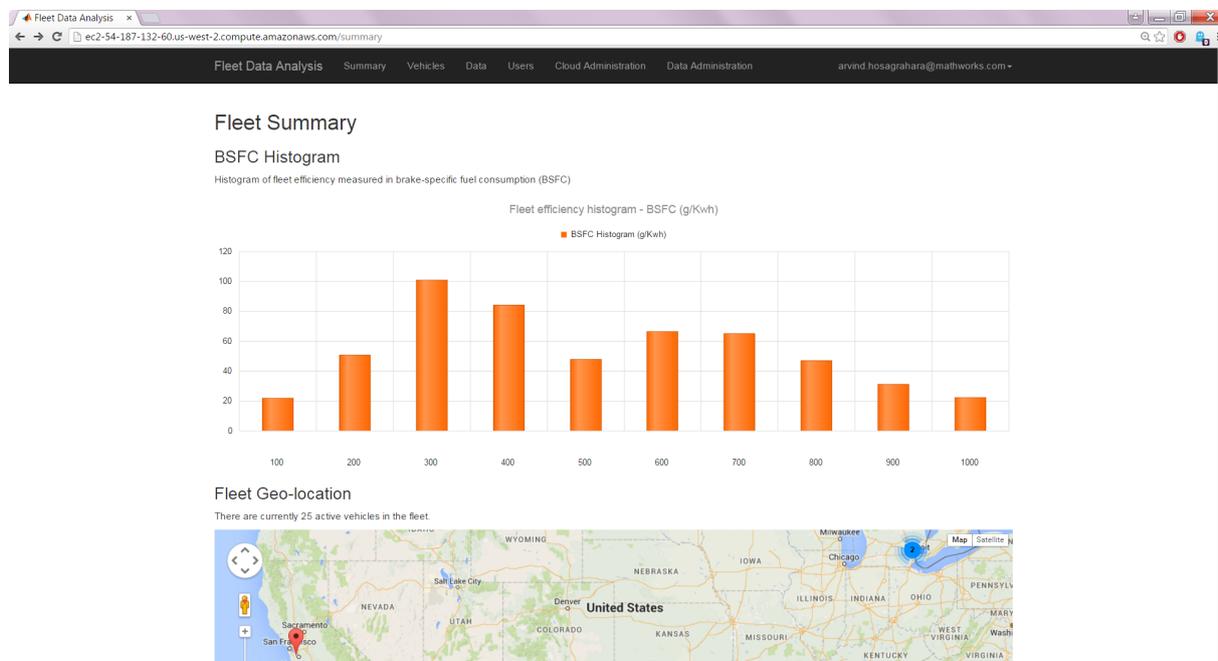


図2. MATLABによるフリートデータ解析ウェブアプリケーション

### MATLABとHadoopを使ったMapReduce

大規模エンジニアリングデータを計算する際の課題の一つとして、データが大きすぎて利用可能なメモリ内に収まらないことが多く、単一のプロセッサでは分析に時間がかかり過ぎることが挙げられます。MapReduceプログラミングは、このような課題に対処するための手法です。データをメモリに収まるサイズの塊（チャンク）に分割し、それぞれを処理して中間結果を得た後、その中間結果を集約して最終的な処理結果を導き出します。Hadoop MapReduceはよく利用されている方法で、Hadoop分散ファイルシステム（HDFS）と共に動作します。また、MATLABにはMapReduce手法を実現する独自の方法があります。

デスクトップでMATLABを動作させ、MapReduceを用いてインタラクティブにアルゴリズムの開発やデバッグを行うことができます。その後コードの修正なしにHadoopを使用してアルゴリズムを実行するには、二つのオプションがあります。一つは、MATLAB Distributed Computing Server™を使用してHadoop MapReduce内でMATLAB MapReduceに基づくアルゴリズムを実行し、Hadoopで蓄積・管理されているデータの処理を行う方法です。もう一つは、MATLAB Compiler™を使用し、MATLAB MapReduceを基にしたアプリケーションを作成し、Hadoopのプロダクションインスタンス内に実装する方法です。

従来型のデータ分析手法では、一般に、分析を行うコンピュータ環境上にデータを移動させる必要があります。しかし大規模エンジニアリングデータの場合、データを分析環境に移す方法は取れないことがあります。MATLABがHadoopをサポートすることにより、チームはデータのある場所で分析を行うことができるようになり、保存されている場所からビッグデータを動かす必要がなくなります。

## 大規模エンジニアリングデータ解析から知見を得る

ビッグデータを使用して実際にエンジンおよび車両の設計を改善しようとするのであれば、組織内の誰もが収集・蓄積されたデータから知見を導き出すことができるようにする必要があります。本セクションでは、MATLABを用いたデータアナリティクスにより得た知見の例を3つ紹介します。

### 実際の条件下での正味燃料消費率の把握

自動車開発に携わるエンジニアは、どんなに巧みに設計されたエンジンでも、トランスミッションがそのエンジンに合わせて動作するよう調整されていなければ、最大効率で動作しないことを知っています。正味燃料消費率（BSFC）は、燃費に加えてエンジンの調整度合も反映した指標で、ギアボックスやシフトスケジュール、その他のトランスミッションのパラメーター最適化に用います。BSFCはシミュレーション最適化の問題として取り上げられることが多いのですが、データ駆動型アプローチによって、異なる視点から実世界の様々な運転パターンに基づく燃費に関する理解を深めることができます。

私たちは、MATLABのデータアナリティクスを利用して、約25台のサンプルを用いたBSFCの分析を行いました。実際にドライバー達が日々の通勤や移動に使った際の走行データです。MATLABを用いてエンジン速度に応じたトルクをプロットするとともに、一定出力の双曲線を追加してBSFC図を作成しました（図3参照）。

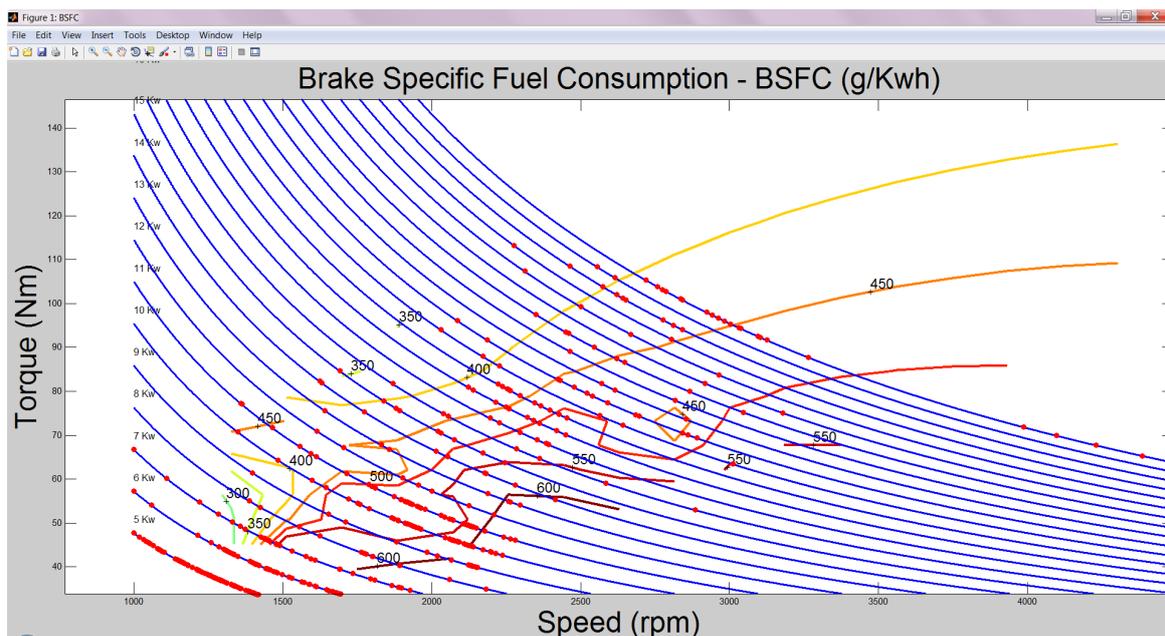


図3. 実際の車両走行データからMATLABで作成したBSFC図

BSFC図の作成に用いた走行データからは、エンジン速度とトルクだけでなく、さらに多くの情報を得ることができます。緯度、経度、高度といったデータを利用すれば、エンジンが効率よく動作した時の車両位置を特定できます。また、騒音や振動のデータを加えれば、燃費改善策が騒音・振動・ハーシュネス (NVH) にどのような影響をもたらすか知ることができます。

### インフラ変更の評価

交通関連のエンジニアであれば、同じ車両走行データを用いて問題のある交通パターンを特定すると共に、その解決方法の評価を行うことができます。一つの例として、車両群が最も燃料を多く消費する地域を特定するアナリティクスを開発しました。MATLABアナリティクスとサードパーティのBI (ビジネス・インテリジェンス) 解析ソフトウェアとを組み合わせることによって、インタラクティブなウェブダッシュボードを短期間で構築しました (図4)。ユーザーがダッシュボードのトリガーを操作すると、MATLABコードが呼び出されて再計算が行われ、その結果に基づいてダッシュボードのグラフィック表示が更新されます。

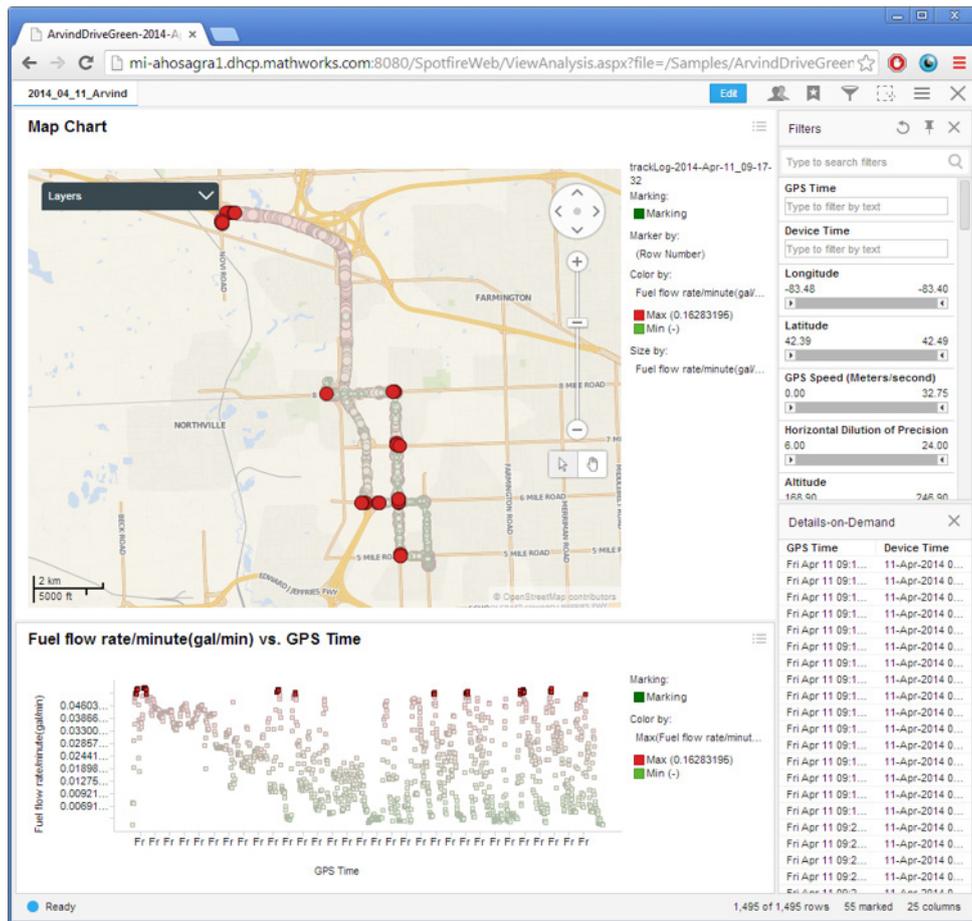


図4. MATLABとBIソフトウェアを組み合わせて作成したインタラクティブなウェブダッシュボード

このダッシュボードでは、待ち時間が長く、燃料の消費がとりわけ非効率になる交差点を強調表示しました。さらに、MATLABを用いてデータを詳しく調べ、最も問題のある交差点の一つで、車両が約90秒間停止する間に消費する燃料を計算しました。この結果（1台当たり0.035ガロン（0.132リットル）をもとに試算したところ、交差点をラウンドアバウト方式に変更することにより、その交差点の通行車全体で1日当たり120ガロン（452リットル）を上回る燃料が節約され、年間約450万ポンド（約2千トン）二酸化炭素排出量が削減されると算出されました。自動車メーカー（OEM）のエンジニアであれば、実際の交通パターンを利用したこのような分析を通じて、アイドリングストップシステムなど車載システムの最適化を行うことができるでしょう。

### Veracity（正確さ）に関する理解

ビッグデータに関する議論で常に中心となるのはVolume（量）、Variety（多様性）、Velocity（速度）です。しかし、データアナリティクスから間違った結論を導き出さないためには、利用するデータのVeracity（正確さ）に対する理解が不可欠であることも認識されるようになってきました。一例として、自動車開発チームが予測モデルを作成する際に、フリートデータを用いてそのモデルのトレーニングを行う場合を考えてみましょう。トレーニングに使用するデータに誤りがあれば、そのモデル

に基づいてビジネス上の意思決定を行っても、根拠がないこととなります。機械学習アプリケーションのネットワークのトレーニングの場合も、同様の問題に直面します。誤りのあるデータを用いて分類モデルのトレーニングを行ったとしても、そのモデルは期待通りには動作しないからです。

今回のフリートデータ計測に参加したドライバーの一人が、アフターマーケットでアップグレードソフトウェアを購入し、車両に搭載しました。そのソフトウェアは、瞬間最大トルクを高め、過給圧を上げ、イグニッションのタイミングが最適化されると謳っていました。これを付けて実際に運転してみると、通常設定時よりもトルクが増加したことが容易に体感できました。ところがデータを検証してみたところ、各設定の違いはほとんどなく、購入したソフトウェアがデータを調整していたことが分かりました。言ってみれば、私たちが収集したデータの方が正確ではなかったのです。このような実地に収集したデータが真実でない可能性があるという発見自体が貴重な知見です。組織が間違ったデータに基づいて設計上の誤った意思決定を行わないようにする上で役に立つからです。

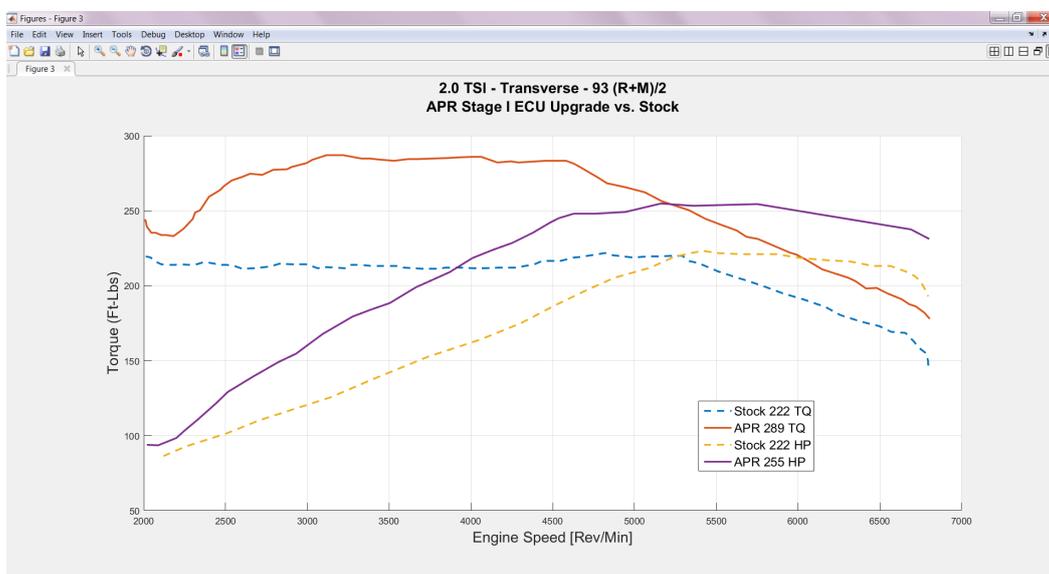


図5. 通常設定時および設定変更時のECUの推定性能の比較

## まとめ

自動車開発において収集された多くのデータに対して、MATLABはオープンで拡張性のあるデータアナリティクス・プラットフォームを提供することができます。このプラットフォームは、データアナリティクス・アプリケーションの迅速な開発、改良、実装のための実績あるワークフローをサポートしています。MathWorksはビッグデータ解析への投資を続けており、MATLABが既にサポートしているデータストア、MapReduce、Hadoop、その関連技術に基づいてさらなる開発を進めています。時間をかけずにシステムを導入したい方や専門家の助言を必要とする方には、MathWorksコンサルティングサービスが、大規模エンジニアリングデータ解析から貴重な知見を導き出す技術を身に着けるお手伝いをします。

MathWorksコンサルタントArvind Hosagraharaが、フリートテストデータ解析システムを実装するための1年に及ぶ研究プロジェクトの成果を紹介する[ビデオをご覧ください](#)。自動車開発におけるエンジニアリングデータの様々な利用事例に関する興味深い知見が盛り込まれています。