Statistical Analysis of the Efficiency of an Integrated Voltage Regulator by means of a Machine Learning Model Coupled with Kriging Regression

R. Trinchero¹, M. Larbi², M. Swaminathan² and F.G. Canavero¹

¹Department of Electronics and Telecommunication, Politecnico di Torino, Torino, Italy
Email: {riccardo.trinchero,flavio.canavero}@polito.it

²School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA
Email: mourad.larbi@ece.gatech.com, madhavan.swaminathan@ece.gatech.edu

Intro: design process and needs

- Robust designs require assessment against uncertainty
- Optimize design to achieve better product performance

Need for a computational model, i.e. a procedure (e.g., analytical formula, algorithm, …) computing quantities of interest from input parameters
The crux of the matter

- **Computational models**
  - are demanding in terms of memory and CPU time

- **Statistical simulations**
  - are necessary for both
    - the assessment of the design robustness w.r.t. uncertainty and variability
    - the performance optimization

---

Surrogate Modeling

- A surrogate model $\hat{M}$ is an approximation of the full-computational model $M$

- Determined from a limited number of runs of the full model
- Simpler **closed-form** relationship
- Faster than the full-model $M$

Several fitting techniques are available
Least Squares Support Vector Machines (LS-SVM)

- Machine learning technique for both classification and regression
- Kernel method: mapping data from the vector space to the feature space
- The number of unknowns is independent from problem dimensionality (non-parametric regression)

**Generic formulation**

\[ M(x) = \sum_{i=1}^{L} \alpha_i K(x_i, x) + b \]

- Linear kernel: \( K(x, x_i) = x^T x_i \)
- Polynomial kernel of degree \( d \): \( K(x, x_i) = (1 + x^T x_i)^d \)
- Radial basis function RBF kernel: \( K(x, x_i) = \exp(-\|x - x_i\|^2/\sigma^2) \)
- MLP kernel: \( K(x, x_i) = \tanh(k x^T x + \theta) \)

The response \( y = M_{LS-SVM}(x) \) is given as the linear combination of the kernel evaluated at the training samples \( x_i \).

**LS-SVM (cont’d)**

\[ M_{LS-SVM}(x) = \sum_{i=1}^{L} \alpha_i K(x_i, x) + b \]

Least squares solution

\[
\begin{bmatrix}
0 & 1 & \cdots & 1 \\
1 & K(x_1, x_1) + \frac{1}{y_1} & \cdots & \frac{1}{y_1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & K(x_L, x_1) & \cdots & K(x_L, x_L) + \frac{1}{y_L}
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_L \\
b
\end{bmatrix}
= \begin{bmatrix}
0 \\
y_1 \\
\vdots \\
y_L
\end{bmatrix}
\]

The coefficients \( \alpha \) and \( b \) are estimated by solving a simple linear system

**How accurate is the model prediction?**
**Deterministic Regression & Error Function**

- Training samples \( \{(x_1, y_1), \ldots, (x_4, y_4)\} \)

- **Build a surrogate mode via a deterministic regression**

\[
y_i = M_{LS-SVM}(x_i) + e(x_i)
\]

- Surrogate Model Error

**How accurate is the prediction?**
Confidence bounds are needed!!!
Probabilistic Interpretation

We can think of the error function $e(x)$ in a probabilistic sense.

In practice, error is a smooth function.

Assumption:
$e(x)$ is smooth $\rightarrow$ The values $\eta(x_1), \ldots, \eta(x_n)$ are correlated.

Correlation via Covariance function, e.g.,
\[
k(x_j, x_j) = \sigma^2 \exp\left(-\frac{(x_j - x_i)^2}{2\sigma^2}\right)
\]
Probabilistic Interpretation

We can think of the error function $e(x)$ in a probabilistic sense.

Assumption:

- $e(x)$ is smooth → The values $\eta(x_1), \ldots, \eta(x_n)$ are correlated

Prior: before using the training samples

Correlation via Covariance function, e.g.,

$$k(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$

Gaussian Process (GP) Regression: from Prior to Posterior

- The value of the error $e(x_i)$ is known on the training samples $e(x_i)$
GP Regression: from Prior to Posterior

- The value of the error $e(x_i)$ is known on the training samples $e(x_i)$
- Discarding all the functions not compatible with the training samples

We consider only all functions fitting our data. Such functions can be described in terms of a probability law $\to$ probabilistic model.

GP Regression & Probabilistic Model

- The resulting model combines the deterministic regression $M_{LS-SVM}$ with a probabilistic model of the error function $e(x)$

The model provides for any value $x_*$ a distribution $[3]$

$$ M(x_*) \sim N(\mu(x_*), \sigma^2(x_*)) $$

most probable confidence value of the error bounds

GP allows converting any regression-based deterministic model into a probabilistic one.

---

System-in-package (SiP) solution with buck converter, low-dropout/load and an integrated inductor on an organic package [4]

The buck converter is used to drop the voltages of the power plane 1 to the level required by the power plane 2

The converter efficiency depends on the integrated inductor

The IVR efficiency has been investigated by considering 8 uniformly distributed parameters


IVR Results

- A subset of $L = 200$ training samples is selected via Latin Hypercube Sampling
- The surrogate model predictions are compared with the results of a MC simulation with 10000 samples

<table>
<thead>
<tr>
<th>Method</th>
<th>Kernel Regression</th>
<th>RMSE</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\ell_{\text{model}}$</th>
<th>$\ell_{\text{out}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>–</td>
<td>67.01</td>
<td>0.31</td>
<td>–</td>
<td>–</td>
<td>7 days</td>
</tr>
<tr>
<td>LS-SVM</td>
<td>Linear</td>
<td>0.158</td>
<td>67.03</td>
<td>0.28</td>
<td>&lt;1s</td>
<td>&lt;1s</td>
</tr>
<tr>
<td></td>
<td>Poly Order 2</td>
<td>0.166</td>
<td>67.04</td>
<td>0.29</td>
<td>&lt;1s</td>
<td>&lt;1s</td>
</tr>
<tr>
<td></td>
<td>Poly Order 3</td>
<td>0.443</td>
<td>67.00</td>
<td>0.54</td>
<td>&lt;1s</td>
<td>&lt;1s</td>
</tr>
<tr>
<td></td>
<td>RBF</td>
<td>0.152</td>
<td>67.02</td>
<td>0.28</td>
<td>1.4s</td>
<td>&lt;1s</td>
</tr>
<tr>
<td>GP</td>
<td>–</td>
<td>0.162</td>
<td>67.02</td>
<td>0.28</td>
<td>&lt;1s</td>
<td>2.4s</td>
</tr>
<tr>
<td>LS-SVM+GP</td>
<td>Linear</td>
<td>0.162</td>
<td>67.02</td>
<td>0.28</td>
<td>&lt;1s</td>
<td>&lt;1s</td>
</tr>
<tr>
<td></td>
<td>RBF</td>
<td>0.156</td>
<td>67.02</td>
<td>0.28</td>
<td>1.5s</td>
<td>&lt;1s</td>
</tr>
<tr>
<td>Sparse PC</td>
<td>Poly Order 9</td>
<td>0.170</td>
<td>67.02</td>
<td>0.28</td>
<td>&lt;1s</td>
<td>&lt;1s</td>
</tr>
</tbody>
</table>

LS-SVM and LS-SVM+GP regression with RBF kernel provides the most accurate metamodel.

IVR Results (cont’d)

- The prediction and 95% confidence intervals estimated by the GP+LS-SVM (RBF) are compared with the results of a MC simulation with 10000 samples.

Excellent agreement for both the model prediction and Confidence Intervals.
Conclusions

- **Surrogates** based on Machine Learning regressions represent effective solution for the UQ in nonlinear problems.

- **Surrogates** are built from a limited set of training samples provided by the full-model.

- **Gaussian Process regression (aka Kriging)** allows building accurate probabilistic models providing an estimation of the model output + confidence bounds.

- **FUTURE WORK:** What is the limit in term of number of parameters?