Subject: "Correct" Data Philosophy
Posted by David Fanning on Thu, 17 Dec 2009 16:43:48 GMT
View Forum Message <> Reply to Message

Folks,

Every couple of weeks I get an e-mail from someone whose data is "missing" and they want to replace it with the "correct" value. These e-mails bug me because if the data is "missing" how the hell would I know what the "correct" value is suppose to be?

But, generally speaking, they want some method to guess at the "correct" values by looking around the neighborhood, shuffling their feet, etc. I guess we have all been tempted to fudge data, if only for aesthetic reasons, so maybe it is a legitimate request.

What would you tell them to do?

If I get some good suggestions I'll write an article so I can get rid of these requests in the future. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by Laura on Thu, 17 Dec 2009 20:33:19 GMT
View Forum Message <> Reply to Message

On Dec 17, 11:43 am, David Fanning <n...@dfanning.com> wrote:
> Folks,
>
> Every couple of weeks I get an e-mail from someone whose
> data is "missing" and they want to replace it with the
> "correct" value. These e-mails bug me because if the
> data is "missing" how the hell would I know what the
> "correct" value is suppose to be?
>

> But, generally speaking, they want some method to
> guess at the "correct" values by looking around the
> neighborhood, shuffling their feet, etc. I guess we
> have all been tempted to fudge data, if only for
> aesthetic reasons, so maybe it is a legitimate request.
>
> What would you tell them to do?
>

Is it similar to "interpolation" or "approximation" or "estimation"?

How about linear/bilinear/trilinear interpolation? Or minimum curvature surface or thin-plate-spline? It also depends on how many values are available and/or missing. There are other fitting/ interpolation functions too.

Subject: Re: "Correct" Data Philosophy
Posted by Kenneth P. Bowman on Thu, 17 Dec 2009 21:23:16 GMT
View Forum Message <> Reply to Message

In article <MPG.25940db1221ff3269896aa@news.giganews.com>,
David Fanning <news@dfanning.com> wrote:

> Folks,
>
> Every couple of weeks I get an e-mail from someone whose
> data is "missing" and they want to replace it with the
> "correct" value. These e-mails bug me because if the
> data is "missing" how the hell would I know what the
> "correct" value is suppose to be?
>
> But, generally speaking, they want some method to
> guess at the "correct" values by looking around the
> neighborhood, shuffling their feet, etc. I guess we
> have all been tempted to fudge data, if only for
> aesthetic reasons, so maybe it is a legitimate request.
>
> What would you tell them to do?
>
> If I get some good suggestions I'll write an article
> so I can get rid of these requests in the future. :-)
>
> Cheers,
>
> David

The problem of estimating values where you have no data is

very common and often very difficult. The best approach depends
on the character of the data, the size of the gaps, the methods used,
and the purpose of the analysis.

It is very important to not mislead yourself or your readers.
My first recommendation is *not* to fill gaps whenever possible --
instead, adapt your analysis and display methods to the data.
If you are displaying an image or contour, for example, show
the viewer where the data is missing with a special color
and don't display contours where there is no data.

If I am plotting global maps of 5 deg x 5 deg data, it should
look chunky (pixelated), not smooth. That reminds the viewer
what the actual resolution of the data is.

If you need to do a Fourier transform, consider using
least-squares estimation rather than interpolating
and using an FFT.

If the data is smooth and the gaps are small, interpolation
will probably work well. If the data is noisy and the gaps are
large, it is possible that nothing will work well.

If you do fill gaps, always test the impact on your results.
Does it matter whether you use linear or cubic interpolation,
for example?

In the end, you need to be confident that your results do not
depend significantly on how you chose to estimate the missing
data.

Cheers, Ken

Subject: Re: "Correct" Data Philosophy
Posted by rogass on Thu, 17 Dec 2009 21:41:14 GMT
View Forum Message <> Reply to Message

On 17 Dez., 21:33, Laura <haixia...@gmail.com> wrote:
> On Dec 17, 11:43 am, David Fanning <n...@dfanning.com> wrote:
>
>
>
>> Folks,
>
>> Every couple of weeks I get an e-mail from someone whose
>> data is "missing" and they want to replace it with the
>> "correct" value. These e-mails bug me because if the

>> data is "missing" how the hell would I know what the
>> "correct" value is suppose to be?
>
>> But, generally speaking, they want some method to
>> guess at the "correct" values by looking around the
>> neighborhood, shuffling their feet, etc. I guess we
>> have all been tempted to fudge data, if only for
>> aesthetic reasons, so maybe it is a legitimate request.
>
>> What would you tell them to do?
>
> Is it similar to "interpolation" or "approximation" or "estimation"?
>
> How about linear/bilinear/trilinear interpolation? Or minimum
> curvature surface or thin-plate-spline? It also depends on how many
> values are available and/or missing. There are other fitting/
> interpolation functions too.

As Laura said you can't give general recommendations - it always
depends on that special case. Maybe you can suggest to leave some data
out and to test then the goodness of fit due to the missing but
existent data.

Just my 2 cents

CR

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Thu, 17 Dec 2009 21:56:56 GMT
View Forum Message <> Reply to Message

Kenneth P. Bowman writes:

> The problem of estimating values where you have no data is
> very common and often very difficult. The best approach depends
> on the character of the data, the size of the gaps, the methods used,
> and the purpose of the analysis.
>
> It is very important to not mislead yourself or your readers.
> My first recommendation is *not* to fill gaps whenever possible --
> instead, adapt your analysis and display methods to the data.
> If you are displaying an image or contour, for example, show
> the viewer where the data is missing with a special color
> and don't display contours where there is no data.
>
> If I am plotting global maps of 5 deg x 5 deg data, it should
> look chunky (pixelated), not smooth. That reminds the viewer

> what the actual resolution of the data is.
>
> If you need to do a Fourier transform, consider using
> least-squares estimation rather than interpolating
> and using an FFT.
>
> If the data is smooth and the gaps are small, interpolation
> will probably work well. If the data is noisy and the gaps are
> large, it is possible that nothing will work well.
>
> If you do fill gaps, always test the impact on your results.
> Does it matter whether you use linear or cubic interpolation,
> for example?
>
> In the end, you need to be confident that your results do not
> depend significantly on how you chose to estimate the missing
> data.

OK, here is my problem: I don't have any idea what you
people are talking about. And neither do the folks asking
me questions. :-(

This, in particular, is opaque to me:

> If you need to do a Fourier transform, consider using
> least-squares estimation rather than interpolating
> and using an FFT.

OK, I will, but *how*!?

> Is it similar to "interpolation" or "approximation" or "estimation"?

Yeah, it's similar to all of those, I guess. But, how
would you do it in IDL?

> How about linear/bilinear/trilinear interpolation? Or minimum
> curvature surface or thin-plate-spline? It also depends on how many
> values are available and/or missing. There are other fitting/
> interpolation functions too.

Does IDL even *do* these things!? Or do I have to go learn
Matlab?

I guess I was hoping for a couple of examples. I really don't
have the time or energy to open up a whole new research area
here, although I can see that it might occupy my time quite
fruitfully for a number of years. :-(

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by Giorgio on Thu, 17 Dec 2009 22:31:12 GMT
View Forum Message <> Reply to Message

On Dec 17, 1:56 pm, David Fanning <n...@dfanning.com> wrote:
> Kenneth P. Bowman writes:
>> The problem of estimating values where you have no data is
>> very common and often very difficult. The best approach depends
>> on the character of the data, the size of the gaps, the methods used,
>> and the purpose of the analysis.
>
>> It is very important to not mislead yourself or your readers.
>> My first recommendation is *not* to fill gaps whenever possible --
>> instead, adapt your analysis and display methods to the data.
>> If you are displaying an image or contour, for example, show
>> the viewer where the data is missing with a special color
>> and don't display contours where there is no data.
>
>> If I am plotting global maps of 5 deg x 5 deg data, it should
>> look chunky (pixelated), not smooth. That reminds the viewer
>> what the actual resolution of the data is.
>
>> If you need to do a Fourier transform, consider using
>> least-squares estimation rather than interpolating
>> and using an FFT.
>
>> If the data is smooth and the gaps are small, interpolation
>> will probably work well. If the data is noisy and the gaps are
>> large, it is possible that nothing will work well.
>
>> If you do fill gaps, always test the impact on your results.
>> Does it matter whether you use linear or cubic interpolation,
>> for example?
>
>> In the end, you need to be confident that your results do not
>> depend significantly on how you chose to estimate the missing
>> data.

>
> OK, here is my problem: I don't have any idea what you
> people are talking about. And neither do the folks asking
> me questions. :-(
>
> This, in particular, is opaque to me:
>
>   If you need to do a Fourier transform, consider using
>   least-squares estimation rather than interpolating
>   and using an FFT.
>
> OK, I will, but *how*!?
>
>> Is it similar to "interpolation" or "approximation" or "estimation"?
>
> Yeah, it's similar to all of those, I guess. But, how
> would you do it in IDL?
>
>> How about linear/bilinear/trilinear interpolation? Or minimum
>> curvature surface or thin-plate-spline? It also depends on how many
>> values are available and/or missing. There are other fitting/
>> interpolation functions too.
>
> Does IDL even *do* these things!? Or do I have to go learn
> Matlab?
>
> I guess I was hoping for a couple of examples. I really don't
> have the time or energy to open up a whole new research area
> here, although I can see that it might occupy my time quite
> fruitfully for a number of years. :-(
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming:http://www.dfanning.com/
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

My 2second thought:
I think it depends of the case. One example I can imagine is the removal of hot pixels from a CCD camera. Since you know that systematically your CCD camera is giving you a hot pixel at the same position, you can estimate its value from its nearest neighbours. However if you are not sure, its value can have a more profound meaning than your instrument has a different response function for

that point. You could be missing something then.
I agree with Kenneth, you must always present the raw data and then
the data treated so people can judge the difference. Or at least give
the option about it.

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Thu, 17 Dec 2009 22:58:55 GMT
View Forum Message <> Reply to Message

Giorgio writes:

> I think it depends of the case. One example I can imagine is the
> removal of hot pixels from a CCD camera. Since you know that
> systematically your CCD camera is giving you a hot pixel at the same
> position, you can estimate its value from its nearest neighbours.
> However if you are not sure, its value can have a more profound
> meaning than your instrument has a different response function for
> that point. You could be missing something then.
> I agree with Kenneth, you must always present the raw data and then
> the data treated so people can judge the difference. Or at least give
> the option about it.

I *completely* agree with this. In fact, "Data Provenance"
is my middle name! But, if I *did* want to correct the
data... Oh never mind. I'm just going to stick to my
indignant reaction to the very idea and leave it at
that. :-(

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by Kenneth P. Bowman on Thu, 17 Dec 2009 23:03:38 GMT
View Forum Message <> Reply to Message

In article <MPG.2594571640f8a8219896ab@news.giganews.com>,

David Fanning <news@dfanning.com> wrote:

> OK, here is my problem: I don't have any idea what you
> people are talking about. And neither do the folks asking
> me questions. :-(

That crux of the issue here is that this problem is *hard*,
and it is difficult to generalize from one situation to
another. Kind of like asking -- "How do I write a good IDL
program?" :-)

Experience with similar data sets is very helpful -- that is,
we learn by doing (and making mistakes and re-doing).

>
> This, in particular, is opaque to me:
>
> If you need to do a Fourier transform, consider using
> least-squares estimation rather than interpolating
> and using an FFT.
>
> OK, I will, but *how*!?

This is actually quite easy. You can use REGRESS. I'll try to
write a short example that will demonstrate, among other things,
that when there is *no* missing data, least squares is exactly
equivalent to the FFT.

> Does IDL even *do* these things!? Or do I have to go learn
> Matlab?

IDL does a number of different kinds of interpolation. For the
basics you can look in my book. The chapter on interpolation
happens to be the sample chapter that is posted on my web site

http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf

Cheers, Ken

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Thu, 17 Dec 2009 23:36:06 GMT
View Forum Message <> Reply to Message

Kenneth P. Bowman writes:

> IDL does a number of different kinds of interpolation. For the
> basics you can look in my book. The chapter on interpolation

> happens to be the sample chapter that is posted on my web site
>
> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf

I have the book. I'll have a look. Thanks. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by d.poreh on Fri, 18 Dec 2009 09:14:30 GMT
View Forum Message <> Reply to Message

On 17 Dez., 15:36, David Fanning <n...@dfanning.com> wrote:
> Kenneth P. Bowman writes:
>> IDL does a number of different kinds of interpolation. For the
>> basics you can look in my book. The chapter on interpolation
>> happens to be the sample chapter that is posted on my web site
>
>> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf
>
> I have the book. I'll have a look. Thanks. :-)
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming:http://www.dfanning.com/
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

In this book: Carr, J.R. 1995. Numerical Analysis for the Geological
Sciences Prentice Hall, Inc, NJ (592 P).
There is a chapter talking about missed data.
Cheers

Subject: Re: "Correct" Data Philosophy
Posted by lecacheux.alain on Fri, 18 Dec 2009 09:35:02 GMT
View Forum Message <> Reply to Message

On 18 déc, 00:36, David Fanning <n...@dfanning.com> wrote:
> Kenneth P. Bowman writes:
>> IDL does a number of different kinds of interpolation. For the
>> basics you can look in my book. The chapter on interpolation
>> happens to be the sample chapter that is posted on my web site
>
>> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf
>
> I have the book. I'll have a look. Thanks. :-)
>
> Cheers,
>
> David
>

Regarding interpolation, a summary rule might be the following:
Interpolation is deeply related to sampling; if your function is
sampled
in a Shannon compliant way, you CAN always interpolate (the cubic
spline
being then an excellent approximate of the ideal interpolating
function).
If it is not, you CANNOT do anything.
More generally, your ability to "correct" an image depends on whether
or not you can
get a sufficient knowledge of its statistics. For example, the "bad
pixel" problem
in CCD images, means that you implicitely assume that the pixel
distribution in the image
cannot contain such an outlier: then you know how you can set a
threshold or build
an adapted filter. IDL contains most of the tools needed for that.
But if you have no idea of your data statistics, neither IDL, Matlab
nor anything else
will be able to help you...
alx.

Subject: Re: "Correct" Data Philosophy
Posted by H. Evans on Fri, 18 Dec 2009 15:31:54 GMT
View Forum Message <> Reply to Message

On Dec 17, 5:43 pm, David Fanning <n...@dfanning.com> wrote:
> Folks,

>
> Every couple of weeks I get an e-mail from someone whose
> data is "missing" and they want to replace it with the
> "correct" value. These e-mails bug me because if the
> data is "missing" how the hell would I know what the
> "correct" value is suppose to be?]

Hi, some of my data went missing yesterday. If you find it, could you send it back to me? Thanks!

Subject: Re: "Correct" Data Philosophy
Posted by Laura on Fri, 18 Dec 2009 16:44:20 GMT
View Forum Message <> Reply to Message

>> Is it similar to "interpolation" or "approximation" or "estimation"?
>
> Yeah, it's similar to all of those, I guess. But, how
> would you do it in IDL?
>
>> How about linear/bilinear/trilinear interpolation? Or minimum
>> curvature surface or thin-plate-spline? It also depends on how many
>> values are available and/or missing. There are other fitting/
>> interpolation functions too.
>
> Does IDL even *do* these things!? Or do I have to go learn
> Matlab?
>
> I guess I was hoping for a couple of examples. I really don't
> have the time or energy to open up a whole new research area
> here, although I can see that it might occupy my time quite
> fruitfully for a number of years. :-(
>

In addition to the interpolation functions in Ken's sample book chapter, there are other interpolating functions in IDL, MIN_CURV_SURF and GRID_TPS, for smooth interpolation. Basically you can get higher order continuity (continuous first order partial derivatives) and the result will be smoother than linear interpolations. MIN_CURV_SURF also has a keyword to set as a TPS interpolation.

GRID_TPS use "thin plate spline" as the interpolating function, which I used a lot in 3D modeling before moving to IDL. They can estimate the values using data samples on irregular grid (which means as long as you know the sample data locations and values, you are fine, they don't need to be on regular grids).

MIN_CURV_SURF probably uses minimum curvature flow? I don't know how

they implemented it in IDL, but it's much slower than GRID_TPS, and the results are quite similar. However, I think IDL has some limit on the number of data samples. A couple of thousands seemed to be fine, but when I tried more, the functions failed. Probably it is due to the memory limit, because basically you need to solve a NxN equation matrix.

Again, as other people mentioned in this thread, it really depends on what the application of the data is. I think estimating/interpolating data should be OK in some applications. If you try to enlarge a picture, you basically estimate the intermediate values among original pixels. Sometimes when the data is missing at some points, estimating the missing values can give people a rough idea what possible data will be there.

Laura

Subject: Re: "Correct" Data Philosophy
Posted by Paul Van Delst[1] on Fri, 18 Dec 2009 18:03:59 GMT
View Forum Message <> Reply to Message

Laura wrote:
> On Dec 17, 11:43 am, David Fanning <n...@dfanning.com> wrote:
>> Folks,
>>
>> Every couple of weeks I get an e-mail from someone whose
>> data is "missing" and they want to replace it with the
>> "correct" value. These e-mails bug me because if the
>> data is "missing" how the hell would I know what the
>> "correct" value is suppose to be?
>>
>> But, generally speaking, they want some method to
>> guess at the "correct" values by looking around the
>> neighborhood, shuffling their feet, etc. I guess we
>> have all been tempted to fudge data, if only for
>> aesthetic reasons, so maybe it is a legitimate request.
>>
>> What would you tell them to do?
>>
>
> Is it similar to "interpolation" or "approximation" or "estimation"?
>
> How about linear/bilinear/trilinear interpolation? Or minimum
> curvature surface or thin-plate-spline? It also depends on how many
> values are available and/or missing. There are other fitting/
> interpolation functions too.

And, just to emphasise the case dependence of an interpolate solution to this problem: do you need the derivatives of your data to be continuous? If so, not just any old interpolation function will do.

Ken Bowman hit it on the head: adapt your analysis and display methods to the data.

Anything else is what I would call "plotology" (is to data display as, e.g., astrology is to astronomy). If I see "real" data (i.e. from some sort of instrument or model/analysis) and it looks beautiful to behold, I'm immediately suspicious.

The "correct" philosophy to have towards data is, IMO, to not have one. The data is what it is. If one expects it to be something else, their analysis will likely trend it that way. That's called bias. Most data is already biased but one hopes it's mainly due to our measurement errors or lack of understanding about the real world rather than our massaging techniques. (bias correction techniques are themselves the subject of many meetings and conferences)

Anyway....

cheers,

paulv

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Fri, 18 Dec 2009 20:22:05 GMT
View Forum Message <> Reply to Message

Laura writes:

> GRID_TPS use "thin plate spline" as the interpolating function, which
> I used a lot in 3D modeling before moving to IDL. They can estimate
> the values using data samples on irregular grid (which means as long
> as you know the sample data locations and values, you are fine, they
> don't need to be on regular grids).

OK, I'm thinking of this problem sort of like that time I missed an easy overhead and lost to that smart-aleck young kid and came home and maybe pushed the door a little too hard with my tennis bag and there was a bit of a hole in the dry wall. "Thin Plate Spline" sounds like the wire gauze I had to use to repair the darn thing. Is it like that?

If so, how could I use it to "repair" some dropped data points in the center of my image, for example?

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by R.G.Stockwell on Sat, 19 Dec 2009 14:52:46 GMT
View Forum Message <> Reply to Message

"Kenneth P. Bowman" <k-bowman@null.edu> wrote in message
news:k-bowman-565924.17033817122009@news.tamu.edu...
> In article <MPG.2594571640f8a8219896ab@news.giganews.com>,
> David Fanning <news@dfanning.com> wrote:
>
>> OK, here is my problem: I don't have any idea what you
>> people are talking about. And neither do the folks asking
>> me questions. :-(
>
> That crux of the issue here is that this problem is *hard*,
> and it is difficult to generalize from one situation to
> another. Kind of like asking -- "How do I write a good IDL
> program?" :-)
>
> Experience with similar data sets is very helpful -- that is,
> we learn by doing (and making mistakes and re-doing).
>
>>
>> This, in particular, is opaque to me:
>>
>> If you need to do a Fourier transform, consider using
>> least-squares estimation rather than interpolating
>> and using an FFT.
>>
>> OK, I will, but *how*!?
>
> This is actually quite easy. You can use REGRESS. I'll try to
> write a short example that will demonstrate, among other things,
> that when there is *no* missing data, least squares is exactly
> equivalent to the FFT.

This is true. However, the moment you remove even one point,
all the columns off the matrix (i.e. Ax=b, where b is the data,
and x is the spectrum) are no longer orthogonal, and thus
one cannot use x = A^t b (which is the fft).

We are now stuck with using x = (A^tA)^-1 A^t b
which requires many many more calculations, and in my
experience, is most often an illposed matrix.

Even greatly reducing the number of spectral points, in order
to ensure an overdetermined system, did not lead to reliable results.

gaps are problematic to spectral analysis. It is basically a divide
by zero problem. Your starting point is the true spectrum convolved
with a spectrum of your gap funtion (i.e. 1's and 0's), and there is no
good way to deconvolve that.

A cimmon solution I have used has been to interpolate the gaps, perform a
local
spectral analysis (using the S-Transform for instance :), and
then reinserting these gaps into that local spectrum.

cheers,
bob

PS lomb scargle is not an adequate solution to the problem.
It calculates _one_ spectral component in a least squares manner,
and should not be used to calculate the full spectrum.
Numerical Recipes screwed that one up.

Subject: Re: "Correct" Data Philosophy
Posted by fburton on Sun, 20 Dec 2009 14:57:38 GMT
View Forum Message <> Reply to Message

In article <k-bowman-565924.17033817122009@news.tamu.edu>,
Kenneth P. Bowman <k-bowman@null.edu> wrote:
> IDL does a number of different kinds of interpolation. For the
> basics you can look in my book. The chapter on interpolation
> happens to be the sample chapter that is posted on my web site
>
> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf

Unfortunately, "File not found!" for that URL.

Francis

Subject: Re: "Correct" Data Philosophy
Posted by Kenneth P. Bowman on Sun, 20 Dec 2009 16:57:18 GMT

View Forum Message <> Reply to Message

In article <1261321058.431566@irys.nyx.net>,
fburton@nyx.net (Francis Burton) wrote:

> In article <k-bowman-565924.17033817122009@news.tamu.edu>,
> Kenneth P. Bowman <k-bowman@null.edu> wrote:
>> IDL does a number of different kinds of interpolation. For the
>> basics you can look in my book. The chapter on interpolation
>> happens to be the sample chapter that is posted on my web site
>>
>> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf
>
> Unfortunately, "File not found!" for that URL.
>
> Francis

Sorry, we did a software upgrade on Friday and naturally that
was one of the things the broke. I'll try to get it fixed on
Monday. (After I get my e-mail working :-( ).

Ken

Subject: Re: "Correct" Data Philosophy
Posted by Laura on Mon, 21 Dec 2009 17:19:14 GMT
View Forum Message <> Reply to Message

On Dec 18, 3:22 pm, David Fanning <n...@dfanning.com> wrote:
> Laura writes:
>> GRID_TPS use "thin plate spline" as the interpolating function, which
>> I used a lot in 3D modeling before moving to IDL. They can estimate
>> the values using data samples on irregular grid (which means as long
>> as you know the sample data locations and values, you are fine, they
>> don't need to be on regular grids).
>
> OK, I'm thinking of this problem sort of like that time I missed
> an easy overhead and lost to that smart-aleck young kid and
> came home and maybe pushed the door a little too hard with my
> tennis bag and there was a bit of a hole in the dry wall.
> "Thin Plate Spline" sounds like the wire gauze I had to
> use to repair the darn thing. Is it like that?
>
> If so, how could I use it to "repair" some dropped
> data points in the center of my image, for example?
>

Here's how I use the GRID_TPS in IDL:

In my example, I have the original data on very sparse grids with some missing values, but I want to interpolate the data at a higher resolution:

```
FUNCTION TPSInterpolation, org_data, missValue, newDimx, newDimy,
minX, minY, maxX, maxY
; org_data is the original data on a regular grid located at (or
bounded by) [minX, minY, maxX, maxY]
; missValue is the filled-in value in org_data indicating the real
value is missing there
; newDimx and newDimy are the dimensions of the resulting data,
; if you just want to fill in the value on orginal grids, I think you
can use the dimensions of the org_data

  data =fltarr(newDimx, newDimy)
  orgInd = where(org_data NE missValue, count)
  if (count EQ 0) then begin
    data = congrid(org_data, newDimx, newDimy)
    return, data
  endif

  sz=size(org_data)
  dimx = sz[1]
  dimy = sz[2]

  xSpan = maxX-minX
  ySpan = maxY-minY

  dx0=xSpan/(dimx-1)
  dy0=ySpan/(dimy-1)

  xVector=findgen(dimx)*dx0 + minX ;xlocation
  yVector=findgen(dimy)*dy0 + minY ;ylocation
  indices = array_indices(org_data, orgInd)

  xPos = xVector[indices[0,*]] ;Xp
  yPos = yVector[indices[1,*]] ;Yp
  values=org_data(orgInd) ; Values

  dx=xSpan/(newDimy-1)
  dy=ySpan/(newDimy-1)

  data = grid_tps(xPos, yPos, values, COEFFICIENTS = coef, NGRID=
[newDimx, newDimy], START=[minX, minY], DELTA=[dx,dy])
  return, data
END
```

Note: If you want to use MIN_CURVE_SURF then the call function can be set as:

data = min_curve_surf(values, xPos, yPos, GS=[dx, dy], BOUNDS = [minX, minY, maxX, maxY], NX=newDimx, NY=newDimy)

If you want to fill in the missing value in a large array, I think dividing them into blocks and working on each block separately will be a good idea.

Hope this helps.

Laura

Subject: Re: "Correct" Data Philosophy
Posted by fburton on Mon, 21 Dec 2009 19:08:59 GMT
View Forum Message <> Reply to Message

In article <k-bowman-334716.10571820122009@news.tamu.edu>, Kenneth P. Bowman <k-bowman@null.edu> wrote:
>>> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf
>>
>> Unfortunately, "File not found!" for that URL.
>
> Sorry, we did a software upgrade on Friday and naturally that
> was one of the things the broke. I'll try to get it fixed on
> Monday. (After I get my e-mail working :-( ).

No worries!

Francis

Subject: Re: "Correct" Data Philosophy
Posted by Kenneth P. Bowman on Mon, 21 Dec 2009 22:03:59 GMT
View Forum Message <> Reply to Message

We did a major software upgrade on our servers last Friday, which broke a few things, then had a two-hour power outage today just to add to the fun.

I think things are working now.

You can download the interpolation chapter from my book here

http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf

I also made a sample program that shows how to fit sines and cosines using least-squares (REGRESS in this case).

http://csrp.tamu.edu/downloads/fft_vs_least_squares.pro.zip

Most of the program is concerned with printing and plotting. The actual calculations don't take much space.

This program creates a 1-D function containing a sine term, a cosine term, and some noise. The noise serves to ensure that there is spectral power at all frequencies. You can set the amplitude of the noise to zero to get a pure sinusoid.

Part 1 computes the FFT and inverse FFT and plots the result.

Part 2 uses REGRESS to fit sines and cosines. Using the equivalent set of sines and cosines as FFT, the coefficients are identical.

Part 3 demonstrates fitting sines and cosines with REGRESS when data points are unevenly spaced or missing. This is particularly useful when you only need to estimate a few Fourier components, as it is much slower than an FFT when n is large.

In this sample program you can see that deleting two points does not cause large errors in the estimates of the magnitudes of the original sine and cosine components.

Bob Stockwell posted a comment earlier about how using regression to compute FFTs when data are missing can affect the results. This was exactly my point at the beginning of this discussion. You really need to understand the methods that you are using. If detailed spectral analysis is necessary, then regression may not be appropriate. On the other hand, how do you deal with missing data? Interpolating to fill data gaps will also affect the spectrum. It is important to experiment with data that has known properties to determine how your particular choices affect the results.

I don't think I have ever been to a thesis defense where someone didn't ask the question: "You did such-and-such to your data. How did that assumption or approximation affect your results?"

Cheers, Ken

Subject: Re: "Correct" Data Philosophy

Posted by jkj on Tue, 22 Dec 2009 17:54:51 GMT
View Forum Message <> Reply to Message

On Dec 21, 4:03 pm, "Kenneth P. Bowman" <k-bow...@null.edu> wrote:
> We did a major software upgrade on our servers last Friday,
> which broke a few things, then had a two-hour power outage
> today just to add to the fun.
>
> I think things are working now.
>
> You can download the interpolation chapter from my book here
>
> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf
>

Thanks - just a thought, that adding a cover page with details of the book would make it more useful from a marketing perspective. By using such a cover page even though the sample chapter may be passed around "out of context" the details of the book would still be immediately available and lead any interested buyers directly to your site. I see title information embedded in the PDF title but it would be nice to see details about the book as a cover page to the chapter (and then the book definitely does not get overlooked).

-Kevin

Subject: Re: "Correct" Data Philosophy
Posted by Kenneth P. Bowman on Tue, 22 Dec 2009 19:02:21 GMT
View Forum Message <> Reply to Message

In article
<50bf0227-1748-4129-a1d6-e2f244275142@k19g2000yqc.googlegroups.com>,
jkj <kevin@vexona.com> wrote:

>> We did a major software upgrade on our servers last Friday,
>> which broke a few things, then had a two-hour power outage
>> today just to add to the fun.
>>
>> I think things are working now.
>>
>> You can download the interpolation chapter from my book here
>>

>>
>
> Thanks - just a thought, that adding a cover page with details of the

> book would make it more useful from a marketing perspective. By using
> such a cover page even though the sample chapter may be passed around
> "out of context" the details of the book would still be immediately
> available and lead any interested buyers directly to your site. I see
> title information embedded in the PDF title but it would be nice to
> see details about the book as a cover page to the chapter (and then
> the book definitely does not get overlooked).
>
> -Kevin

I see your point. I was assuming that readers would get the
pdf from here

http://idl.tamu.edu/Home.html

This sits on a university server, so there are no direct links
to purchase the book. Under university rules that would constitute
"profiting from the use of university facilities". David Fanning can
tell you how rich we are getting from writing IDL books.

If you want to buy the book, Amazon has it a very reasonable price.

Cheers, Ken

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Tue, 22 Dec 2009 20:01:20 GMT
View Forum Message <> Reply to Message

Kenneth P. Bowman wites:

> David Fanning can
> tell you how rich we are getting from writing IDL books

And Tiny Tim says "God Bless!" this holiday season!

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thue. ("Perhaps thos speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Tue, 29 Dec 2009 19:43:27 GMT
View Forum Message <> Reply to Message

David Fanning writes:

> Every couple of weeks I get an e-mail from someone whose
> data is "missing" and they want to replace it with the
> "correct" value. These e-mails bug me because if the
> data is "missing" how the hell would I know what the
> "correct" value is suppose to be?
>
> But, generally speaking, they want some method to
> guess at the "correct" values by looking around the
> neighborhood, shuffling their feet, etc. I guess we
> have all been tempted to fudge data, if only for
> aesthetic reasons, so maybe it is a legitimate request.
>
> What would you tell them to do?

I've written a small article on the subject of using
thin plate splines to correct missing data in a 2D
surface and I placed it here:

http://www.dfanning.com/code_tips/gridtps.html

I have grave misgivings about doing this, but I figure
what happens in this newsgroup stays in this newsgroup.
At least I hope so. Please, please, use good judgment
if you chose to read the article. :-(

Any many thanks to all of you who offered good ideas. This
is probably not my last word on the subject. I'm learning
a lot. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Wed, 30 Dec 2009 22:14:32 GMT
View Forum Message <> Reply to Message

Kenneth P. Bowman writes:

> You can download the interpolation chapter from my book here
>
> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf
>
> I also made a sample program that shows how to fit sines and
> cosines using least-squares (REGRESS in this case).
>
> http://csrp.tamu.edu/downloads/fft_vs_least_squares.pro.zip
>
> Most of the program is concerned with printing and plotting. The
> actual calculations don't take much space.

Ken, I have been studying this example and the last two chapters in your book much of the day. I have to say, this is probably the first time in my life that I have a practical understanding of what the FFT actually does! And from your examples, it even seems obvious to me what FFT filtering is all about.

Thanks very much for providing this information. :-)

Have a Happy New Year!

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: "Correct" Data Philosophy
Posted by Kenneth P. Bowman on Thu, 31 Dec 2009 15:18:02 GMT
View Forum Message <> Reply to Message

In article <MPG.25a57eb349d570149896c4@news.giganews.com>,
David Fanning <news@dfanning.com> wrote:

> Kenneth P. Bowman writes:
>

>> You can download the interpolation chapter from my book here
>>
>> http://csrp.tamu.edu/pdf/idl/sample_chapter.pdf
>>
>> I also made a sample program that shows how to fit sines and
>> cosines using least-squares (REGRESS in this case).
>>
>> http://csrp.tamu.edu/downloads/fft_vs_least_squares.pro.zip
>>
>> Most of the program is concerned with printing and plotting. The
>> actual calculations don't take much space.
>
> Ken, I have been studying this example and the last two
> chapters in your book much of the day. I have to say, this
> is probably the first time in my life that I have a practical
> understanding of what the FFT actually does! And from your
> examples, it even seems obvious to me what FFT filtering
> is all about.
>
> Thanks very much for providing this information. :-)
>
> Have a Happy New Year!
>
> David

Hi David,

Glad that I could help. :-)

If I ever find time to work on a second edition, I am hoping to
add chapters on other methods such as matrix solutions, EOFs,
numerical solution of ODEs, and numerical integration.

I will have to deal with the difficult problem of how much
mathematical detail to include in an introductory programming book.
But it will be fun!

Cheers, Ken

Subject: Re: "Correct" Data Philosophy
Posted by David Fanning on Thu, 31 Dec 2009 16:10:00 GMT
View Forum Message <> Reply to Message

Kenneth P. Bowman writes:

> If I ever find time to work on a second edition, I am hoping to
> add chapters on other methods such as matrix solutions, EOFs,

> numerical solution of ODEs, and numerical integration.
>
> I will have to deal with the difficult problem of how much
> mathematical detail to include in an introductory programming book.
> But it will be fun!

Well, my first criticism of your book was that here I
was loping along with all this beginning IDL information
when I ran right smack into that FFT chapter. I was
bloody and battered.

But I do find I pick this book up quite often to read those
last two chapters. I never understood your FFT example,
to tell you the truth, until I compared it with the
example you provided the other day. Then, suddenly, it
all made sense to me.

I would welcome an introductory to intermediate couple
of chapters on the topics above. But, many, many examples,
please! I'm pretty dense. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")